V. V. Arestov, V. I. Berdyshev, N. I. Chernykh, T. V. Demina, N. N. Kholschevnikova,
S. V. Konyagin, Yu. N. Subbotin, S. A. Telyakovskii, I. G. Tsar'kov, V. A. Yudin

Exposition of the lectures by S. B. Stechkin on approximation theory

Translated into English from the Russian original by A. R. Alimov,
edited by V. I. Burenkov

РОССИЙСКАЯ АКАДЕМИЯ НАУК
УРАЛЬСКОЕ ОТДЕЛЕНИЕ
ИНСТИТУТ МАТЕМАТИКИ И МЕХАНИКИ
МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМ. М. В. ЛОМОНОСОВА
УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМ. А. М. ГОРЬКОГО

ИЗЛОЖЕНИЕ ЛЕКЦИЙ С. Б. СТЕЧКИНА ПО ТЕОРИИ ПРИБЛИЖЕНИЙ

ЕКАТЕРИНБУРГ, 2010

Preface

Professor Sergey Borisovich Stechkin (1920–1995) was a famous Russian mathematician working in the theory of approximation and the theory of numbers. He was also an exceptional personality and a brilliant lecturer who for many years gave deep sophisticated courses on approximation theory at the M.V. Lomonosov Moscow State University and at the Ural State University. Stechkin was a supervisor of many students who themselves became well-known experts in the theory of approximation. In 2010, several of his former post-graduate students, now distinguished professors, prepared an upgraded version of the record of Stechkin's lectures given in 1970–1971.

Exposition of Stechkin's lectures, published in Russian in 2010 by the Institute of Mathematics and Mechanics of the Ural Branch of the Russian Academy of Sciences, immediately became popular.

Without doubt, the English translation of the Lectures would be of great interest for the international community of analysts working in the theory of approximation. This opinion was enthusiastically shared by the Editorial Board of the Eurasian Mathematical Journal and it was decided to publish the translation as one of the issues of this journal. During translation some additional editing of the original text was done. In particular, the list of references was essentially enlarged, and references were given to most of the statements mentioned in the lecture course without proofs.

V. I. Burenkov, M. Otelbaev, V. A. Sadovnichy

Preface to the Russian edition

This book emerged from the lecture notes on approximation theory given by Professor S. B. Stechkin as a special course at the Department of Mechanics and Mathematics of the M. V. Lomonosov Moscow State University in 1970–1971. All the lectures were written up by his student T. V. Demina.

Quite understandably, the writeup of a student's notes did require further adaptation, and this was accomplished recently by Stechkin's followers and former students. Lectures 1–4 have been prepared for publication by Yu. N. Subbotin, Lectures 5–7 by S. A. Tel'yakovskii, Lectures 8–10 by V. I. Berdyshev, N. N. Kholschevnikova and I. G. Tsar'kov, Lectures 11–13 by S. V. Konyagin and I. G. Tsar'kov, Lectures 14, 15 and in part Lecture 16 have been prepared by V. A. Yudin, and Lectures 16–20 by V. V. Arestov.

Preliminary processing of the original lecture notes (transcription of shorthand notes, typesetting formulas, correction of obvious misprints, preparation of the figures) followed by repetitive retyping thereof during editing and coordination has been performed by A. I. Kozko, Yu. V. Malykhin, T. V. Radoslavova, N. N. Kholshchevnikova in Moscow, and by P. Yu. Glazyrina, M. V. Deikalova, A. A. Koshelev, N. A. Kuklin, K. S. Tikhanovtseva, V. V. Shevchenko in Ekaterinburg. Especially much labour has been endured by M. V. Deikalova, Yu. V. Malykhin, and V. V. Shevchenko.

The general editing of the text as a whole has been done by N. I. Chernykh with much help by S. A. Tel'yakovskii, N. N. Kholshchevnikova, and Yu. V. Malykhin.

Various variants of the special course on approximation theory were delivered by Stechkin regularly, during almost each of his many years of pedagogical work at the Moscow and Ural State Universities. A distinctive feature of all Stechkin's courses, including, *inter alia*, the following lecture notes, is the originality in choosing the material and its exposition. Students who attended Stechkin's lectures, mandatory and special courses, do remember and highly appreciate exceptional skill and artistry in delivering lectures. The brilliant teaching of Professor Stechkin always enchanted the listeners.

Unfortunately, it is not feasible in any exposition to replicate the style of presentation in which Professor Stechkin delivered his lectures. So we have not tried to emulate his style, as any attempt to do so would be futile.

Nevertheless, we hope that this exposition will be of use both for students and staff. All footnotes were made during preparation and editing of the manuscript for publication.

Contents

Lecture 1

Interpolation

1.1. Basic concepts

Definition. A *metric space* $R = \{X, \varrho\}$ is a set X and a real-valued function ϱ on $X \times X$ (or 'metric') which satisfies:

1) $\varrho(x, y) \geqslant 0, \ \ \rho(x, y) = \rho(y, x) \quad \forall\, x, y \in X;$
2) $\varrho(x, y) = 0 \Longleftrightarrow x = y;$
3) $\varrho(x, y) \leqslant \varrho(x, z) + \varrho(z, y) \quad \forall\, x, y, z \in X.$

The starting point of the approximation theory is the concept of the *best approximation*, that is, the distance of a given element $x \in X$ to a given nonempty subset M of X,

$$\textit{best approximation} \ \ E(x) = E(x, M)_R = \inf_{y \in M} \varrho(x, y) = \varrho(x, M) \geqslant 0.$$

The set of all *elements of best approximation* in M for a given element x is denoted by $Y(x)$:

$$y^* = y^*(x) \in Y(x) \Longleftrightarrow \begin{cases} 1) & y^* \in M, \\ 2) & \varrho(x, y^*) = E(x, M)_R. \end{cases}$$

(elements of the set $Y(x)$ are also called *nearest points* or *best approximants*).

The operator $x \mapsto Y(x)$ is called the *operator of best approximation* or the *metric projection* of x to M.

Given a point x and a set M, the set $Y(x)$ may be empty, consist of one or more points.

Example 1.1. Let $R = \{\mathbb{R}^2, \varrho\}$, where ϱ is the Euclidean distance on the plane $\mathbb{R}^2$, and let M be a closed circular disc. It is easily seen that the set $Y(x)$ is a singleton for any x. If M is the interior of an open circular disc, then $Y(x) = \varnothing$ for any $x \notin M$. If M is a circumference and x is its centre, then $Y(x) = M$.

Definition. A set M is called a set of *existence* (*uniqueness*) if, for each $x \in X$, the set $Y(x)$ is non-empty (empty or a singleton). A set M is called a *Chebyshev set* if it is a set of existence and a set of uniqueness, that is, if for each $x \in X, \ Y(x)$ is a singleton.

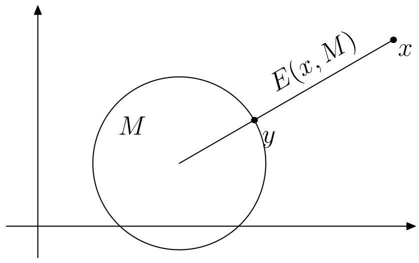

Fig. 1.1

A line in the Euclidean plane serves as an example of a Chebyshev set. The first significant results on Chebyshev sets were obtained in the theory of approximation of functions by Chebyshev, one of the founders of this theory. In 1859 he showed that (in modern terminology) in the space $C[0,1]$ the subspace P_n of all polynomials of degree $\leqslant n$ and the set $R_{n,m}$ of all rational functions

$$\frac{a_0 + a_1 x + \ldots + a_n x^n}{b_0 + b_1 x + \ldots + b_m x^m}$$

with fixed n and m are Chebyshev sets.

However, in general, the operator of the best approximation may have unpleasant peculiarities. Consider the following example.

Example 1.2. Let $X = C[a,b]$, $-\infty < a < b < +\infty$. The norm of a function $f \in C[a,b]$ is defined as $\|f\|_C = \max\limits_{x\in[a,b]} |f(x)|$, $\varrho(f,g) = \|f-g\|_C$. As M we consider the set of all constant functions on $[a,b]$. Clearly, for any $f \in C[a,b]$, there exists a unique approximant $c^* \in M$ (see Fig. 1.2),

$$c^* = \big(\max_{x\in[a,b]} f(x) + \min_{x\in[a,b]} f(x)\big)/2$$

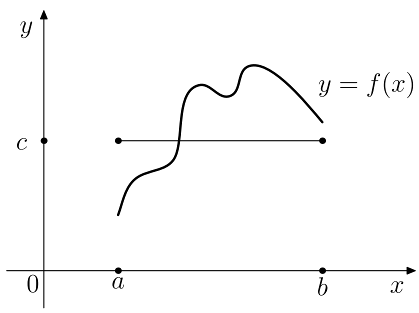

Fig. 1.2

Consequently, the best approximation operator Y (the metric projection), which assigns to each function $f \in C[a,b]$ its element of best approximation $c^* \in M$, is not linear. Indeed, taking f_1 and f_2 as shown in Fig. 1.3, we have $c^*(f_1) = c^*(f_2) = c^*(f_1 + f_2) = h/2$, and so $Y(f_1 + f_2) \neq Y(f_1) + A(f_2)$.

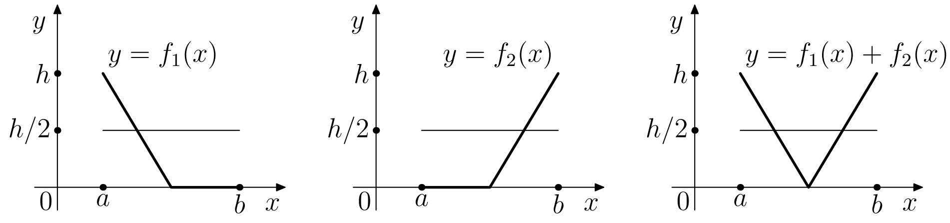

Fig. 1.3

So, in some function spaces the best approximation operator Y is not linear, and hence finding $E(x)$ and $y^*(x)$ may be quite difficult. This suggests considering also more simple methods of approximation, in particular, use is made of various linear methods.

1.2. Linear approximation problem

Let L be a subspace of $C[a, b]$, and let A be a linear operator from $C[a, b]$ to L. This defines the linear method A of approximation of elements in $C[a, b]$ by means of the subspace L. For an f the element Af is the approximating element.

Interpolation is the first classical linear method of approximation.

1.3. Lagrange interpolation

Let $f \in C[a, b]$ (for the time being, we assume that f is complex-valued).

On $[a, b]$, we consider different points x_k, $a \leqslant x_0 < x_1 < \ldots < x_n \leqslant b$. The points $\{x_k\}$ are called *interpolation nodes* (interpolation points).

Given a set of nodes $\{x_k\}$ and a collection of numbers $\{y_k\}$, $k = 0, 1, \ldots, n$, the problem is to construct a polynomial $p_n(x) = a_0 + a_1 x + \ldots + a_n x^n \in \mathcal{P}_n$ such that $p_n(x_k) = y_k$, $k = 0, 1, \ldots, n$.

The following natural questions arise:

1) Is the problem solvable?
2) How many solutions it has?

Here in order to determine the coefficients a_i, $i = 0, 1, \ldots, n$, we have a system of linear equations with nonzero Vandermonde determinant. Thus the problem has a unique solution for any x_k and y_k. It is possible to explicitly write down the solution. To do so, for any $k = 0, 1, \ldots, n$, we construct the Lagrange fundamental polynomial $l_k(x)$, corresponding to the kth node this is a polynomial of degree n such that $l_k(x_i) = \delta_{i,k}$, where $\delta_{i,k}$ is the Kronecker delta.

It is easily seen that

$$l_k(x) = \frac{(x - x_0)(x - x_1) \cdots (x - x_{k-1})(x - x_{k+1}) \cdots (x - x_n)}{(x_k - x_0)(x_k - x_1) \cdots (x_k - x_{k-1})(x_k - x_{k+1}) \cdots (x_k - x_n)}.$$

Setting $\omega(x) = \prod_{k=0}^{n}(x - x_k)$, we have

$$\omega'(x_k) = (x_k - x_0)(x_k - x_1)\cdots(x_k - x_{k-1})(x_k - x_{k+1})\cdots(x_k - x_n)$$

and so

$$l_k(x) = \frac{\omega(x)}{(x - x_k)\omega'(x_k)}.$$

Then the polynomial

$$p_n(x) = p_n\bigl(x, \{y_k\}, \{x_k\}\bigr) = \sum_{k=0}^{n} y_k\, l_k(x)$$

is the required interpolation polynomial, known as the *Lagrange interpolation polynomial.*

Consequently, for any function $f \in C[a,b]$ and any set of interpolation nodes $\{x_k\}$, $k = 0, 1, \ldots, n$, there exists a unique polynomial of degree not exceeding n,

$$p_n(x, f) = p_n(x, f, \{x_k\}) = \sum_{k=0}^{n} f(x_k) l_k(x), \tag{1.1}$$

such that

$$p_n(x_k, f) = f(x_k), \qquad k = 0, 1, \ldots, n.$$

This defines the operator $P_n : f \mapsto p_n(x, f)$ from $C[a,b]$ to $C[a,b]$. We list briefly its simplest properties.

1) Given $f \in \mathcal{P}_n$, for any set of nodes $\{x_k\}$, we have $p_n(x, f) \equiv f(x)$, that is, $P_n(f) = f$.
2) The operator P_n is linear (homogeneous and additive) and bounded:

$$P_n(c_1 f_1 + c_2 f_2) \equiv c_1 P_n(f_1) + c_2 P_n(f_2), \qquad f_i \in C[a,b], \qquad c_i \in \mathbb{C}, \qquad i = 1, 2;$$

also, for any function $f \in C[a,b]$,

$$\|p_n(\cdot, f)\|_C \leqslant L_n \|f\|_C, \quad \text{where} \quad L_n = \|P_n\|_{C \to C} < \infty.$$

Furthermore, by (1.1),

$$|p_n(x, f)| \leqslant L_n(x) \|f\|_C.$$

Here $L_n(x) = \sum_{k=0}^{n} |l_k(x)|$ and $L_n = \|L_n(x)\|_C$.

We claim that these inequalities are sharp in $C[a,b] \equiv C$.

Indeed, given a fixed $\xi \in [a,b]$, consider a function $f_\xi(x)$ such that

a) $f_\xi(x) = \operatorname{sign} l_k(\xi)$ for $x = x_k$, $k = 0, 1, \ldots, n$,
b) $|f_\xi(x)| \leqslant 1$ for $x \in [a,b]$,
c) $f_\xi(x)$ is continuous in x on $[a,b]$.

We have

$$\|f_\xi\|_C = 1, \qquad p_n(x, f_\xi) = \sum_{k=0}^{n} f_\xi(x_k) l_k(x)$$

and in particular,

$$p_n(\xi, f_\xi) = \sum_{k=0}^{n} |l_k(\xi)| = L_n(\xi)\|f_\xi\|_C.$$

Taking for ξ a point x^* at which $L_n(x)$ attains the maximum on $[a, b]$, we have

$$p_n(x^*, f_{x^*}(\cdot)) = L_n\|f_{x^*}\|_C$$

and hence

$$\|p_n(x, f_{x^*}(\cdot))\|_C = L_n\|f_{x^*}\|_C.$$

Consequently, the constant L_n is the norm of the operator $P_n\colon f \mapsto p_n(x, f)$:

$$\|P_n\|_{C\to C} = L_n.$$

Furthermore, for any fixed $x \in [a, b]$, $L_n(x)$ is the norm of the functional $P_x(f) = p_n(x, f)$ in $C[a, b]$:

$$\|P_x(f)\|_{C\to\mathbb{C}} = L_n(x),$$

because $|P_x(f)| \leqslant L_n(x)\|f\|_C$ for any $f \in C[a, b]$, and

$$|P_x(f_x(\cdot))| = L_n(x)\|f_x\|_C.$$

The constant L_n is called the *Lebesgue constant*, and $L_n(x)$ is known as the *Lebesgue function* of the linear method $p_n(x, f, \{x_k\})$ of approximation of functions f in $C[a, b]$ by Lagrange interpolation polynomials. Clearly, these concepts can be extended to other linear approximation methods.

The smaller the norm (the Lebesgue constant) of an interpolation method, the better is the method. For a fixed n, the Lebesgue constant L_n depends on the interpolation nodes $\{x_k\}$. If $[a, b] = [-1, 1]$, then it is possible to choose nodes in such a way that $L_n = \frac{2}{\pi}\ln n + O(1)$ as $n \to +\infty$; this happens if one takes, as interpolation nodes, the zeros of the Chebyshev polynomial

$$T_{n+1}(x) = \cos((n+1)\arccos x).$$

3) *The Cauchy identities.*

Property 1) and the formula for the interpolation polynomial, for $f(x) \equiv 1$, imply the identity

$$\sum_{k=0}^{n} l_k(x) \equiv 1,$$

and, for $f(x) = (x-u)^j$, $j = 1, \ldots, n$, $u \in \mathbb{C}$), imply the identities

$$(x-u)^j \equiv \sum_{k=0}^{n} (x_k - u)^j l_k(x) \qquad j = 1, 2, \ldots, n;$$

in particular, for $u = x$,

$$\sum_{k=0}^{n} (x_k - x)^j l_k(x) \equiv 0, \qquad j = 1, \ldots, n. \tag{1.2}$$

For $\{x_n\} \subset [a, b]$ these identities hold for all $x \in \mathbb{C}$.

1.4. Error of the Lagrange interpolation. The Lebesgue inequalities

Let $\{x_k\}_{k=0}^n$ be interpolation nodes, let $f \in C[a,b]$, and let $p_n(x,f)$ be the corresponding Lagrange interpolation polynomial. We can represent f in the following form

$$f(x) = p_n(x,f) + R_n(x,f),$$

where $R_n(x,f)$ is the remainder. Clearly, $R_n(x_k,f) = 0$ at the interpolation nodes, $k = 0,\dots,n$. Given any fixed $x \in [a,b]$, it is required to evaluate $R_n(x,f)$ and evaluate $\|R_n(\cdot,f)\|_{C[a,b]}$.

It turns out that in order to estimate the remainder in the Lagrange interpolation it suffices to know $L_n(x)$, L_n and $E(f,\mathcal{P}_n)_C = \inf\limits_{q\in\mathcal{P}_n} \|f-q\|_C$. More precisely, the following Lebesgue inequalities hold:

$$\begin{aligned} |R_n(x,f)| &\leqslant (L_n(x)+1)E(f,\mathcal{P}_n)_C, \\ \|R_n(\cdot,f)\|_C &\leqslant (L_n+1)E(f,\mathcal{P}_n)_C. \end{aligned} \tag{1.3}$$

To prove them, we observe that $P_n(x,f)$ is a linear operator and that $P_n(x,q) = q(x)$ for all $q \in \mathcal{P}_n$. We have

$$|R_n(x,f)| = |f(x) - p_n(x,f)| = |f(x) - q(x) - p_n(x,f-q)|$$

$$\leqslant |f(x)-q(x)| + L_n(x)\|f-q\| \leqslant (L_n(x)+1)\|f-q\|_C, \qquad q \in \mathcal{P}_n.$$

Hence, taking for q the best approximant to f in $\mathcal{P}_n$, this gives

$$|R_n(x,f)| \leqslant (L_n(x)+1)E(f,\mathcal{P}_n)_C, \qquad x \in [a,b],$$

and hence

$$\|R_n(\cdot,f)\|_C \leqslant (L_n+1)E(f,\mathcal{P}_n)_C.$$

Similar Lebesgue inequalities also hold for more general linear methods preserving elements of $\mathcal{P}_n$.

1.5. Cauchy form of the remainder in the Lagrange interpolation formula

The space $C^{(n+1)}[a,b]$ consists of all continuous functions having continuous derivatives up to the order $n+1$ inclusive.

Theorem 1.1. *Let $f \in C^{(n+1)}[a,b]$. Then, for any $x \in [a,b]$, there exists a point $\xi \in (a,b)$ such that*

$$R_n(x,f) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\,\omega(x) \tag{1.4}$$

(here $\xi = \xi(x,f,\{x_k\})$, and $n+1$ is the number of interpolation nodes).

Proof. Clearly, the formula in question holds for $x = x_k$, $k = 0, \ldots, n$ (with arbitrary ξ). We fix $x \in [a,b]$, $x \neq x_k$, and consider the following auxiliary function

$$F(t) = f(t) - p_n(t) - K\omega(t),$$

where $p_n(t) = p_n(t, f, \{x_k\})$, $K = R_n(x,f)/\omega(x)$, $\omega(x) \neq 0$. We note that $F(x_k) = 0$, $k = 0, \ldots, n$, and also, $F(x) = 0$ by the choice of K. Hence, the function $F(t)$ has zeros at $n+2$ different points. By the generalized Rolle's theorem, there exists a point $\xi \in (a,b)$ such that $F^{(n+1)}(\xi) = 0$. However, $F^{(n+1)}(\xi) = f^{(n+1)}(\xi) - K \cdot (n+1)!$, giving $K = f^{(n+1)}(\xi)/(n+1)!$. Hence, for $R_n(x,f) = K\omega(x)$,

$$R_n(x,f) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\omega(x). \qquad \square$$

Remark (geometrical). The remainder does not necessarily change sign along with $\omega(x)$ at interpolation nodes. The graphs of $f(x)$ and $p_n(x,f)$ can touch each other as shown in Fig. 1.4.

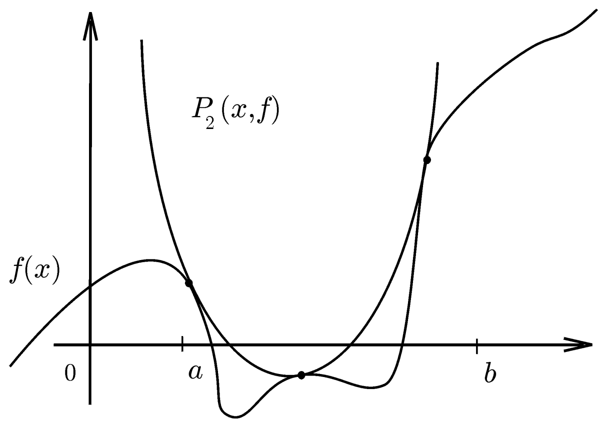

Fig. 1.4

However, there is a simple sufficient condition for the remainder to change sign at the interpolation nodes. As is seen from (1.4), *if* $f^{(n+1)}(x)$ *preserves sign, then* $R_n(x,f)$ *changes sign at the interpolation nodes* (*and only there*).

Let $M_{n+1}(f) = \max\limits_{x\in[a,b]} f^{(n+1)}(x)$.

Corollary. *We have*

$$\|R_n(\cdot, f)\|_C \leqslant \frac{M_{n+1}(f)}{(n+1)!}\, \|\omega(\cdot)\|_C. \tag{1.5}$$

The question arises: What choice of interpolation nodes minimizes $\|\omega(\cdot)\|_C$? We claim that this happens if ω is the Chebyshev polynomial[1] $\widetilde{T}_{n+1}(x, I)$ ($I = [a,b]$). Indeed, $\omega(x) = x^{n+1} + a_n x^n + \ldots + a_0$, and so

$$\inf \|\omega(\cdot)\|_{C[a,b]} = \|\widetilde{T}_{n+1}(\cdot, I)\|_{C[a,b]},$$

[1]Properties of Chebyshev polynomials are given in Lecture 2.

where $\widetilde{T}_{n+1}(\cdot, I)$ is the Chebyshev polynomial on $[a, b]$ (a monic polynomial of order $n+1$ of least deviation from zero on $[a, b]$). In particular,

$$\widetilde{T}_{n+1}(x, [-1, 1]) = 2^{-n} \cos(n+1) \arccos x.$$

1.6. Haar's interpolation theorem in $\mathbb{R}^N$

In the above, we considered the one-dimensional interpolation problem on the interval $D = [a, b] \subset \mathbb{R}^1$. Now suppose that a set $D \subset \mathbb{R}^N$, $N \geqslant 2$.

The problem arises: Is this problem sensible in a multi-dimensional case? Do there exists real-valued functions $f_0(x), f_1(x), \ldots, f_n(x)$, $x \in \overline{D} \subset \mathbb{R}^N$ (*interpolating systems on D*), whose linear combinations are capable of interpolating any family of numbers $\{y_k\}_{k=0}^n$ at any disjoint family of nodes $\{x_k\}_{k=0}^n \in D$?

It is easily verified that interpolating systems consisting of discontinuous functions exist on any set of the power of the continuum. To see this it suffices to consider a one-to-one mapping of the interval to this set and consider the resulting functions, which correspond to the interpolating system $1, x, x^2, \ldots, x^n$ considered above.

In what follows, we shall show that the interpolation problem for polynomials is solvable in the complex domain. The following theorem gives an answer to this question for $\mathbb{R}^N$ (see, e.g., [40]).

Theorem 1.2 (Haar). *Let $N > 1$. Suppose that a set $D \subset \mathbb{R}^N$ has nonempty interior. If $n \neq 0$, then on D there are no interpolation systems consisting of real-valued continuous functions.*

Proof. Consider a neighbourhood $\Delta \subset D$ of an interior point of D. Suppose that $\{x_k\}_{k=0}^n \subset \Delta$. If $\{f_k\}$, $k = 0, \ldots, n$, is an interpolating system, then the system of equations

$$\sum_{k=0}^{n} c_k f_k(x_i) = y_i, \qquad i = 0, \ldots, n,$$

is solvable for any family of numbers $\{y_k\}$. Hence $\det(f_k(x_i)) \neq 0$ for any $\{x_k\} \subset \Delta$.

By the assumption, all f_k are continuous, and hence the determinant is continuous as a function of points $\{x_i\}$ in the domain Δ. We leave fixed all the points x_k except for two points, say x_0 and x_1, and start to continuously map x_0 and x_1 one to the other (see Fig. 1.5) in such a way that all $n{+}1$ points remain different and lie in Δ. The determinant is a continuous real-valued function of the points. Swapping two rows of the determinate makes it to change sign. Hence the determinant vanishes for some intermediate set of points. However, this is impossible for continuous interpolating systems. □

Remark. Continuous interpolating systems fail to exist not only on sets with nonempty interior, but also on continua (compact connected T_2-spaces) with a branching point (see, e.g., [36, Lemma 12-4]).

As before, we continuously map the points x_0, x_1 onto to the other (see Fig. 1.6). The determinant changes sign, and so it must vanish, a contradiction.

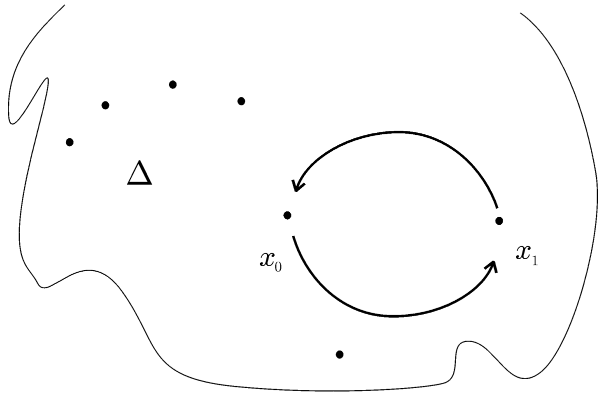

Fig. 1.5

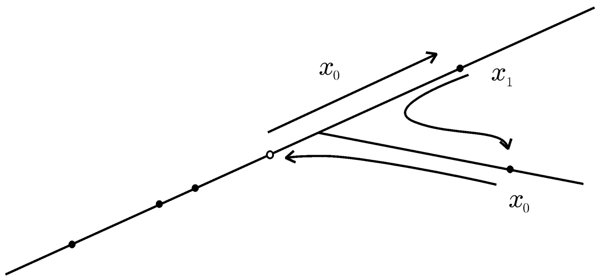

Fig. 1.6

Problem. Which sets K admit continuous interpolating systems?

For $m \geqslant 2$, there is yet no complete answer to this question. For $m = 1$ there is the well-known Mairhuber's theorem (see Lecture 12)[2]: for any $n \geqslant 1$ a compact set K must be homeomorphic to a proper subset of the unit circle in $\mathbb{R}^2$ or to the whole circle (in the last case, if and only if n is even).

[2]Also known as the Mairhuber–Curtis theorem (see e.g. [2]).

Lecture 2

Remainder of the interpolation. Chebyshev polynomials.

2.1. Estimates of the remainder

Let $\{x_k\}_{k=0}^n$ be interpolation nodes on $[a,b]$. Suppose that a function f has a continuous derivative of order $n+1$ on $[a,b]$. Let $p_n(x,f)$ be the corresponding

Lagrange interpolation polynomial for f, and let $R_n(x,f) = f(x)-p_n(x,f)$ be the remainder of the interpolation. In the previous lecture, we have estimated $\|f(\cdot)-p_n(\cdot,f)\|_C$ in terms of the best approximation $E_n(f,\mathcal{P}_n)_C$ (see Lebesgue's inequality (1.3)), and hence, since $E_n(f,\mathcal{P}_n)_C \leqslant \|f\|_C$, in terms of $\|f\|_C$. We also have a bound of $\|R_n(\cdot,f)\|_C$ in terms of the norm $\|f^{(n+1)}\|_C$ on the interval (see (1.5)). Hence,

$$\|f(\cdot) - p_n(\cdot,f)\|_C \leqslant \begin{cases} \mathcal{K}_0\|f\|_C, \\ \mathcal{K}_{n+1}\|f^{(n+1)}\|_C, \end{cases}$$

where $\mathcal{K}_0$ and $\mathcal{K}_{n+1}$ are the corresponding constants independent of f. Our purpose is to find estimates of $\|R_n(\cdot,f)\|_C$ in terms of $\|f^{(m+1)}\|_C$, that is, to obtain bounds of the form

$$\|R_n(\cdot,f)\|_C \leqslant \mathcal{K}_{m+1}\|f^{(m+1)}\|_C \tag{2.1}$$

for other orders m of the derivative, where $K_{m+1} = K_{m+1}(n)$ is independent of f. Assuming such a bound is valid and taking a function $f \in \mathcal{P}_m$ with $\|f^{(m+1)}(\cdot)\|_C = 0$, it follows that $\|R_n(\cdot,f)\|_C = 0$, forcing $f(x) \equiv p_n(x,f)$. Hence f is a polynomial of degree at most n. This gives a necessary condition on m for such a bound to hold: $m \leqslant n$. Consequently, it is impossible to evaluate $\|R_n(\cdot,f)\|_C$ in terms of derivatives of orders higher than n, since the condition $f^{(m+1)}(x) \equiv 0$, for some $m > n$, does not imply $f(x) \equiv p_n(x,f)$ (for example, if f is a polynomial of order $m > n$).

Lemma 2.1. *A necessary and sufficient condition for* (2.1) *to hold is that* $m \leqslant n$.

Proof. The Lagrange interpolation formula for f is as follows:

$$p_n(x,f) = \sum_{k=0}^{n} f(x_k)l_k(x).$$

By the Cauchy identity $\sum_{k=0}^n l_k(x) \equiv 1$, hence

$$R_n(x,f) = f(x) - \sum_{k=0}^n f(x_k) l_k(x) = \sum_{k=0}^n \{f(x) - f(x_k)\} l_k(x).$$

For $f(y)$ we write Taylor's formula of order m at a point x with the remainder in the integral form,

$$f(y) = f(x) + p(x,y) + \frac{1}{m!} \int_x^y (y-t)^m f^{(m+1)}(t)dt,$$

where $f(x) + p(x,y) = q_x(y)$ is the Taylor polynomial of f at x. In particular,

$$f(x_k) = f(x) + q_x(x_k) + \frac{1}{m!} \int_x^{x_k} (x_k - t)^m f^{(m+1)}(t)dt,$$

where

$$q_x(x_k) = \sum_{s=1}^m \frac{1}{s!} f^{(s)}(x)(x_k - x)^s$$

(note that, in general, $q_x(x_k)$ is not a polynomial in x).

By the Cauchy identity (1.2) we have, for $m \leqslant n$,

$$\sum_{k=0}^n q_x(x_k) l_k(x) = \sum_{s=1}^m \frac{1}{s!} f^{(s)}(x) \sum_{k=0}^n (x_k - x)^s l_k(x) \equiv 0,$$

and hence, substituting $f(x_k)$ into the formula for $R_n(x,f)$, this gives

$$R_n(x,f) = -\frac{1}{m!} \sum_{k=0}^n l_k(x) \int_x^{x_k} (x_k - t)^m f^{(m+1)}(t)dt = \int_a^b K_{n,m}(x,t,\{x_k\}) f^{(m+1)}(t)dt,$$

Hence, for all $m \leqslant n$ we have the bound

$$|R_n(x,f)| \leqslant \|f^{(m+1)}(\cdot)\|_C \int_a^b |K_{n,m}(x,t,\{x_k\})|dt, \qquad x \in [a,b],$$

giving the required inequality (2.1).

2.2. Chebyshev polynomials

The algebraic polynomials

$$T_n(x) = \cos(n \arccos x) \qquad (n = 0, 1, \dots), \qquad x \in [-1, 1]$$

are called the *Chebyshev polynomials* (of the first kind). To see that these functions are indeed polynomials: we take $x = \cos\theta$, $\theta \in [0, \pi]$. This gives

$$T_n(x) = \cos(n \arccos x) = \cos n\theta = \frac{e^{in\theta} + e^{-in\theta}}{2} = \frac{1}{2}\big\{(\cos\theta + i\sin\theta)^n + (\cos\theta - i\sin\theta)^n\big\}.$$

Hence, for $|x| \leqslant 1$,

$$T_n(x) = \frac{1}{2}\{(x + i\sqrt{1-x^2})^n + (x - i\sqrt{1-x^2})^n\}.$$

Here, the imaginary parts cancel, and there are no radicals in the real part (see [8, Ch. 3, § 6]).

Exercise. As a polynomial, the function $T_n(x)$ is defined for all x. Find a similar expression for $|x| > 1$.

2.3. Basic properties of Chebyshev polynomials (expressed by equalities)

2.3.1. Recurrence relation

Note that $T_0(x) \equiv 1$, $T_1(x) \equiv x$. Using the trigonometric equality

$$\cos(n+1)\theta = 2\cos\theta\cos n\theta - \cos(n-1)\theta \qquad (x = \cos\theta)$$

we have the recurrence relation

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \qquad (n = 1, 2, \dots).$$

Also,

$$T_n'(x) = n\sin(n\arccos x)\cdot\frac{1}{\sqrt{1-x^2}} = \frac{n\sin n\theta}{\sin\theta}.$$

From the recurrence relation it follows that the leading coefficient of the polynomial $T_n(x)$ for $n \geqslant 1$ is 2^{n-1}, and so

$$T_n(x) = \cos(n\arccos x) = 2^{n-1}x^n + \cdots.$$

All zeros of this polynomial $x_k = \cos\frac{2k-1}{2n}\pi$, $k = 1, 2, \dots, n$, lie in $(-1, 1)$. The extrema points on $[-1, 1]$ are $\widetilde{x}_k = \cos\frac{k\pi}{n}$, $k = 0, 1, \dots, n$. We have $T_n(\widetilde{x}_k) = (-1)^k$; also $T_n'(\widetilde{x}_k) = 0$ for $k \neq 0$ and $k \neq n$, and $T_n'(1) = n^2$, $T_n'(-1) = (-1)^{n-1}n^2$. Zeros and extrema points become more condensed near the ends of $[-1, 1]$ as n increases (see Fig. 2.2).

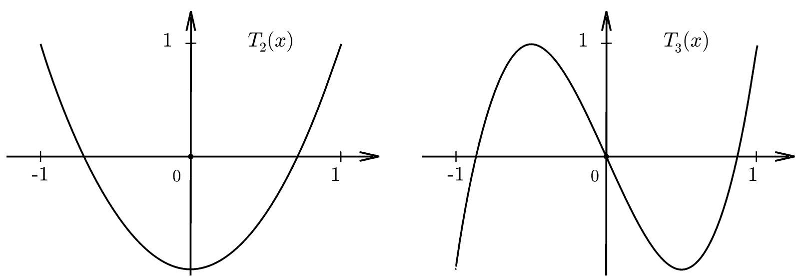

Fig. 2.1

2.3.2. Generating function

The *generating function* of a sequence $\{A_n\}$ is the function F whose Taylor coefficients are A_n.

We calculate $\sum_{n=0}^{\infty} \cos(n\theta)\, t^n$. This is the real part of the power series

$$\sum_{n=0}^{\infty} \exp(in\theta)\; t^n = \frac{1}{1 - t\exp(i\theta)} = \frac{1 - t\exp(-i\theta)}{1 - 2t\cos\theta + t^2},$$

and so

$$\sum_{n=0}^{\infty} \cos(n\theta)\; t^n = \frac{1 - t\cos\theta}{1 - 2t\cos\theta + t^2}.$$

Setting here $x = \cos\theta$, we find the generating function $F(t)$ for the sequence of Chebyshev polynomials:

$$\sum_{n=0}^{\infty} T_n(x)\, t^n = \frac{1 - tx}{1 - 2tx + t^2}.$$

Changing x into $-x$, we obtain

$$\sum_{n=0}^{\infty} T_n(-x)t^n = \sum_{n=0}^{\infty} T_n(x)(-t)^n$$

and hence, $T_n(-x) = (-1)^n T_n(x)$. Of course, this property can also be derived from the explicit formula for T_n.

Exercise. Use the representation

$$T_n(x) = \frac{1}{2}\left\{(x + \sqrt{x^2 - 1})^n + (x - \sqrt{x^2 - 1})^n\right\}, \qquad |x| > 1$$

to find the rate of growth of $|T_n(x)|$ as $n \to \infty$ at points x with $|x| > 1$.

2.3.3. Differential equation

Chebyshev polynomials T_n satisfy the differential equation (see, e.g., [30, Ch.II, § 4], [8, Ch. 3, § 6], [27, 1.1.4.2])

$$(1 - x^2)y'' - xy' + n^2 y = 0, \qquad n = 0, 1, \ldots. \tag{2.2}$$

Assuming that T_n satisfies (2.2), we shall find the coefficients of the polynomial

$$T_n(x) = \sum_{k=0}^{n} a_k x^{n-k}$$

for $n \geqslant 2$. We have $a_0 = 2^{n-1}$, $a_1 = a_3 = \cdots = 0$, since $T_n(-x) = (-1)^n T_n(x)$. Substituting this into (2.2), we obtain the recurrence relation

$$a_{2k} = -\frac{(n - 2k + 2)(n - 2k + 1)}{4k(n - k)} a_{2k-2}.$$

Hence,

$$a_{2k} = \frac{(-1)^k n(n-1)\cdots(n-2k+1)}{4^k k!(n-1)\cdots(n-k)}\, a_0 = (-1^k)\frac{n}{n-k} C_{n-k}^k 2^{n-2k-1},$$

and so

$$T_n(x) = \sum_{k=0}^{[n/2]} a_{2k} x^{n-2k}.$$

Remark. Even though a Chebyshev polynomial is majorized by 1 on $[-1,1]$, its coefficients are fairly large for large n.

2.3.4. Orthogonality

Changing the variable to $x = \cos\theta$, $0 \leqslant \theta \leqslant \pi$, it can be shown that

$$\frac{2}{\pi}\int_{-1}^{1} \frac{T_n(x)\, T_m(x)}{\sqrt{1-x^2}}\, dx = \delta_{n,m},$$

where $\delta_{n,m}$ is the Kroneker symbol ($\delta_{n,m} = 0$ if $n \neq m$, $\delta_{n,n} = 1$).

2.4. Extremal properties of Chebyshev polynomials

2.4.1. The first extremal property

The normalized Chebyshev polynomial $\widetilde{T}_n(x) = \frac{T_n(x)}{2^{n-1}}$ is a polynomial of least deviation from zero on $[-1,1]$ among all monic polynomials of degree n (see [21, § 2.2], [38, Ch. 2], [27, Theorem 1.1.8]).

Theorem 2.1. *Let $n \geqslant 1$. If $p(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_0$, then*

$$\|p\|_{C[-1,1]} \geqslant \|\widetilde{T}_n\|_{C[-1,1]} = 2^{1-n}$$

with equality if and only if $p = \widetilde{T}_n$.

Also,

$$E_{n-1}(x^n)_{C[-1,1]} := \inf_{p\in\mathcal{P}_{n-1}} \|x^n - p(x)\|_{C[-1,1]} = 2^{1-n};$$

and the best approximation $E_{n-1}(x^n)_{C[-1,1]}$ is attained only at the polynomial $p(x) = x^n - \tilde{T}_n(x) \in \mathcal{P}_{n-1}$.

Proof. The basic idea of the proof is in calculating zeros. Suppose there exists a monic polynomial $p \in \mathcal{P}_n$ such that

$$\|p(\cdot)\|_{C[-1,1]} \leqslant \|\widetilde{T}_n(\cdot)\|_{C[-1,1]}.$$

The difference $r_{n-1}(x) = \widetilde{T}_n(x) - p(x)$ vanishes on each closed interval, where $T_n(x)$ varies from ± 1 to ∓ 1, that is, on the intervals $[\widetilde{x}_k, \widetilde{x}_{k+1}]$, where $\widetilde{x}_k = \cos\frac{k\pi}{n}$, $k =$

$0, 1, \ldots, n$. Here we take into account that if $r_{n-1} = 0$ at a common end-point of some two of these intervals, then both $\widetilde{T}_n'$ and p' also vanish, and hence this is a double point of the difference $r_{n-1}(x)$, see Fig. 2.3). Hence the degree of the polynomial r_{n-1} is at most $n-1$, and the number of its zeros (counting multiplicities) is $\geqslant n$. As a result, $r_{n-1} \equiv 0$, $p(x) \equiv \widetilde{T}_n(x)$, whence all the assertions of Theorem 2.1 follow. □

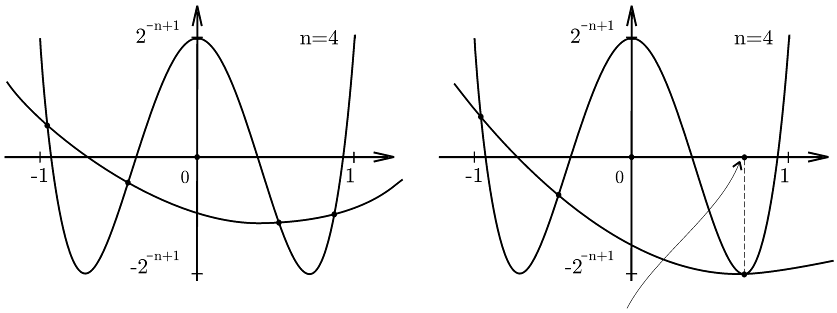

Fig. 2.3. Double zero of $r_3(x)$.

2.4.2. The second extremal property

Lemma 2.2. *Let* $p \in \mathcal{P}_n$ *and let* $\xi \in \mathbb{R}$, $|\xi| > 1$. *Suppose that* $|p(x)| \not\equiv |T_n(x)| \cdot \|p\|_{C[-1,1]}$. *Then*

$$|p(\xi)| < |T_n(\xi)| \cdot \|p(\cdot)\|_{C[-1,1]}.$$

The equality at a point ξ *outside* $[-1, 1]$ *is possible only if* $p(x) \equiv 0$ *or* $\frac{|p(x)|}{\|p\|_{C[-1,1]}} \equiv |T_n(x)|$.

Proof. Assume, to the contrary, that there exists $\xi \notin [-1, 1]$ such that $|p(\xi)| \geqslant |T_n(\xi)| \cdot \|p\|_C$. The difference $q(x) = \frac{p(x)}{\|p\|_{C[-1,1]}} - T_n(x)$ is a nontrivial polynomial of degree at most n, having at most n zeros. Assume, for definiteness, $\xi < -1$ and $\frac{p(\xi)}{\|p\|_C} \geqslant T_n(\xi) > 0$ (hence n is odd, $p(\xi) > 0$). We shall count zeros. Clearly, there is a point $\xi_0 \in [\xi, -1]$ at which $q(\xi_0) = 0$.

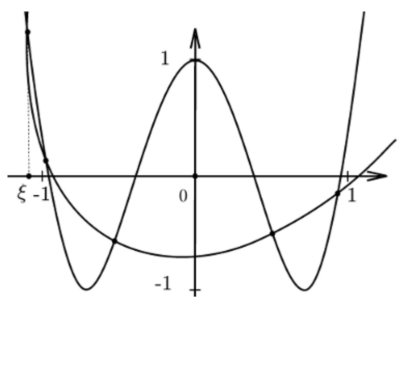

Fig. 2.4

Further, on $[-1,1]$ the graph of the polynomial $y(x) = \frac{p(x)}{\|p\|_C}$ lies within the strip $|y| \leqslant 1$. Hence, by an argument similar to that used the proof of Theorem 2.1, it follows that $q(x)$ has n zeros (counting multiplicities) on $[-1,1]$. If, in addition, $\xi_0 < -1$ (see Fig. 2.4), then $q(x) \equiv 0$ on $\mathbb{R}$. If $\xi_0 = -1$ and $q(x) \neq 0$ on $[\xi,-1)$, then a more subtle analysis is required. First, it may happen that $\xi_0 = -1$ is a second order zero, which is a double root of q. Clearly, neither of these zeros has been counted among the n considered zeros on $[-1,1]$, because of the double zeros we have counted only those which agree with some $\widetilde{x}_k = \cos\frac{k\pi}{n}$, $k = 1,2,\dots,n-1$.

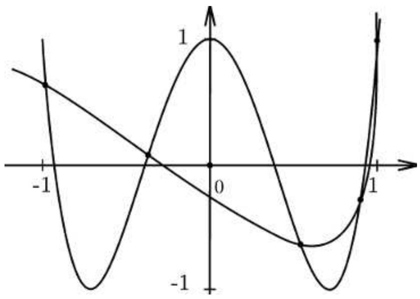

Fig. 2.5

The second case is as follows: $q'(\xi_0) \neq 0$ ($\xi_0 = -1$). In this case $q(x)$ changes sign from positive to negative at $\xi_0 = -1$, the graph of $p(x)$ in the right semi-neighbourhood of the point ξ_0 lying below the graph of T_n. Hence, in both cases, the interval $[-1,\widetilde{x}_{n-1}]$ contains at least two zeros of the polynomial $q(x)$, and the total number of zeros will not be greater than or equal to $n+1$. We again have $q(x) \equiv 0$ on $\mathbb{R}$, contradicting the assumption.

The general case (without assuming that $\xi < -1$ and n is odd) can be reduced to the one just treated by changing p and T_n to $-p$, $-T_n$ and (or) $T_n(x)$ to $T_n(-x)$. $\square$

Corollary. *If* $\|p_n\|_{C[-1,1]} \leqslant 1$, $p_n \in \mathcal{P}_n$, *and* $|\xi| > 1$, *then* $|p_n(\xi)| \leqslant |T_n(\xi)|$.

Exercise. Prove that if $p \in \mathcal{P}_n$ and $\|p(\cdot)\|_{C[-1,1]} \leqslant 1$, then $|p'(1)| \leqslant T_n'(1)$. The proof is again based on counting zeros, see Fig. 2.5.

This property explains why T_n are also called comparison polynomials.

Remark. By Lemma 2.2, if $[-a,a] \supset [-1,1]$ and $p_n \in \mathcal{P}_n$, then

$$\|p(\cdot)\|_{C[-a,a]} \leqslant \|T_n(\cdot)\|_{C[-a,a]} \cdot \|p(\cdot)\|_{C[-1,1]},$$

where the equality is possible only if $p(x) \equiv 0$ or $p(x) \equiv \pm T_n(x)\|p(\cdot)\|_{C[-1,1]}$.

Lecture 3

Chebyshev polynomials (continuation). Applications of interpolation

3.1. V. A. Markov's theorem

3.1.1. The third extremal property of Chebyshev polynomials

Consider the problem put forward by V. A. Markov. Given fixed $m \in \mathbb{N}$ and $n \geqslant m$, find, among all monic polynomials of degree m (the coefficient $a_m = 1$), the particular one of the least deviation from zero on $[-1, 1]$?

We need the following auxiliary result.

Lemma 3.1 (Lemma on zeros). *Let* $\sum_{k=0}^{N} A_k x^{\lambda_k}$, $0 \leqslant \lambda_0 < \lambda_1 < \ldots < \lambda_N$, *be a polynomial with* $N+1$ *terms, which is not identically zero. Then it has at most* N *zeros on* $(0, \infty)$.

Proof. We apply induction on N. For $N = 0$, the $A_0 x^{\lambda_0}$, $A_0 \neq 0$, has no zeros on $(0, \infty)$. For $N = 1$, any nonzero polynomial $A_0 x^{\lambda_0} + A_1 x^{\lambda_1} = x^{\lambda_0}(A_0 + A_1 x^{\lambda_1 - \lambda_0})$ has at most one zero on $(0, \infty)$. Assume that any polynomial $\sum_{k=0}^{N} A_k x^{\lambda_k} \not\equiv 0$ with $N+1$ terms has at most N zeros on $(0, \infty)$. Then the polynomial

$$G(x) = \sum_{k=0}^{N+1} A_k x^{\lambda_k} = x^{\lambda_0} \sum_{k=0}^{N+1} A_k x^{\lambda_k - \lambda_0} = x^{\lambda_0} \cdot F(x)$$

has the same number of zeros $(0, \infty)$ as $F(x) = \sum_{k=0}^{N+1} A_k x^{\lambda_k - \lambda_0}$; also, it may be assumed that $A_{N+1} \neq 0$ (for otherwise $G(x)$ is a polynomial with $N+1$ terms). But then $F'(x) = \sum_{k=1}^{N+1} B_k x^{\lambda_k - \lambda_0 - 1}$ is a nonzero polynomial with $N+1$ terms ($B_{N+1} = (\lambda_{N+1} - \lambda_0) A_{N+1} \neq 0$), which has at most N zeros, by the assumption. By Rolle's theorem, $F(x)$ also has at most $N+1$ zeros. □

Now let us go back to V. A. Markov's problem.

Let $p_n \in \mathcal{P}_n$, $p_n(x) = \sum_{k=0}^n a_k x^k$, $\mathcal{Q}_n^m = \{p_n(x) \mid p_n \in \mathcal{P}_n,\ a_m = 1\}$. It is required to find $p_n^* \in \mathcal{P}_n$ such that

$$\inf_{p_n \in \mathcal{Q}_n^m} \|p_n(\cdot)\|_{C[-1,1]} = \min_{p_n \in \mathcal{Q}_n^m} \|p_n(\cdot)\|_{C[-1,1]} = \|p_n^*(\cdot)\|_{C[-1,1]}.$$

This problem may also be considered as a best approximation problem, that is,

$$\inf_{p_n \in \mathcal{Q}_n^m} \|p_n(\cdot)\|_{C[-1,1]} = \inf_{q \in \mathcal{P}_n,\ a_m=0} \|x^m - q(x)\|_{C[-1,1]} = E(x^m, L_n^m)_{C[-1,1]},$$

where $L_n^m = \{p \in \mathcal{P}_n \mid a_m = 0\}$ is the finite-dimensional linear subspace of $C[-1,1]$, which, along with $\mathcal{P}_n$, shares the following property: the even or the odd part of any polynomial in this subspace is again a polynomial in this subspace.

Note that among polynomials with $a_m = 0$, which deliver the best approximation to the monomial x^m, there is a polynomial of the same parity as x^m.

In fact, let m be an even number, and let $p_n(x)$ be a polynomial of best approximation in this subspace. Then $q_n(x) = \frac{1}{2}\{p_n(-x) + p_n(x)\}$ is also a polynomial in this subspace, this polynomial is now even, and it approximates the monomial x^m not worse than p_n:

$$\|x^m - q_n(x)\|_C = \Big\|\frac{1}{2}\big(x^m - p_n(x)\big) + \frac{1}{2}\big((-x)^m - p_n(-x)\big)\Big\|_C \leqslant \|x^m - p_n(x)\|_{C[-1,1]}.$$

Hence, even polynomials deliver best approximation to even functions. A similar conclusion can be made for odd functions and odd polynomials. Consequently, if m and n have different parity, then a polynomial q_n^* of best approximation of the same parity as m has to be of order $n-1$, and hence $q_n^* \equiv q_{n-1}^*$.

Consider the following formula for Chebyshev polynomials:

$$T_n(x) = \cos(n \arccos x) = \sum_{k=0}^n A_k^n x^k.$$

Theorem 3.1 (Vladimir A. Markov [26]). *If m and n have the same parity, then the нуля polynomial $\frac{T_n(x)}{A_m^n}$ is a polynomial of least deviation from zero among all monic polynomials of degree n ($a_m = 1$, $0 \leqslant m \leqslant n$); the corresponding least deviation is equal to:*

$$\Big\|\frac{T_n(\cdot)}{A_m^n}\Big\|_{C[-1,1]} = \frac{1}{|A_m^n|}.$$

If m and n have opposite parity, then the polynomial $\frac{T_{n-1}(x)}{A_m^{n-1}}$ is a polynomial of least deviation from zero among all monic polynomials of degree n ($a_m = 1$, $0 \leqslant m \leqslant n-1$); the corresponding least deviation is equal to:

$$\Big\|\frac{T_{n-1}(\cdot)}{A_m^{n-1}}\Big\|_{C[-1,1]} = \frac{1}{|A_m^{n-1}|}.$$

Proof. For a proof we need to consider four cases depending on the parities of m and n. We shall prove Theorem 3.1 only in one case when m and n are even. Thus, given an arbitrary monic polynomial $p_n(x) = \sum_{k=0}^{n} a_k x^k$ with even n and m, we need to show that

$$\|p_n(\cdot)\|_{C[-1,1]} \geqslant \frac{1}{|A_m^n|}.$$

Assume, to the contrary, that there exists a polynomial p_n such that

$$\|p_n(\cdot)\|_{C[-1,1]} < \frac{1}{|A_m^n|}. \tag{3.1}$$

The polynomial p_n may fail to be even. But then the polynomial $\frac{1}{2}(p_n(x)+p_n(-x)) \in \mathcal{Q}_n^m$ is even and satisfies inequality (3.1). Hence we assume in the sequel that p_n in (3.1) is an even polynomial, which is different from $\frac{T_n(x)}{A_m^n}$. The Chebyshev polynomial $T_n(x)$ is also even. The polynomial

$$R_n(x) = \frac{T_n(x)}{A_m^n} - p_n(x) = \sum_{k=0}^{n/2} b_k x^{2k} \not\equiv 0,$$

with $b_{m/2} = 0$, has at most $n/2 = l$ nonzero coefficients. By Lemma 3.1, the polynomial $R_n(x) \not\equiv 0$ of l terms has at most $l-1$ zeros on $(0, \infty)$ (and hence on $(0, 1)$). In our case, on $[0, 1]$ $T_n(x)$ has precisely $n/2 + 1$ points $\widetilde{x}_k = \cos \frac{k\pi}{n}$, $k = 0, 1, \ldots, \frac{n}{2}$, of maximum deviations; in view of (3.1) at these points $R_n(x)$ has the same sign as $\frac{T_n(x)}{A_m^n}$. Hence, between these points there are $n/2 = l$ points at which $R_n(x) = 0$. All these l zeros of $R_n(x)$ lie in the interval $(0, 1)$. This, however, contradicts the above. Hence, instead of (3.1), the converse inequality holds, which is equality for $p_n(x) \equiv \frac{T_n(x)}{A_m^n}$.

In the remaining cases the analysis is the same. □

Remark. In a general setting, a polynomial of least deviation is not unique. For example, for $n = 2$, $m = 0$ and $0 \leqslant c \leqslant 2$, we have

$$\|1 - cx^2\|_{C[-1,1]} = 1 = \left\| \frac{T_2(x)}{A_0^2} \right\|_{C[-1,1]}.$$

Corollary (Estimates for the coefficients of polynomials). *Suppose, given a polynomial $p_n(x) = \sum_{k=0}^{n} a_k x^k$, we know its norm $\|p_n(\cdot)\|_{C[-1,1]}$. Then, if m and n have the same parity,*

$$|a_m| \leqslant |A_m^n| \cdot \|p_n(\cdot)\|_{C[-1,1]};$$

if m and n have opposite parity, then

$$|a_m| \leqslant |A_m^{n-1}| \cdot \|p_n(\cdot)\|_{C[-1,1]}.$$

Proof. Indeed, if m and n have the same parity and if $a_m \neq 0$, we have, by Theorem 3.1,

$$\left\| \frac{p_n(\cdot)}{a_m} \right\|_{C[-1,1]} \geqslant \frac{1}{|A_m^n|},$$

whence the required inequality follows. For $a_m = 0$ this is a trivial inequality. Hence, the Chebyshev polynomials have the property that their coefficients of the same parity as the order of the polynomial are the largest possible among the polynomials p_n of the same degree with $\|p_n\|_{C[-1,1]} \leqslant 1$. □

Exercise. Prove Markov's theorem for m and n of opposite parity.

Remark. Since $a_m = \frac{p_n^{(m)}(0)}{p!}$, the inequalities of the corollary can be written as follows

$$|p_n^{(m)}(0)| \leqslant m! \cdot |A_m^n| \cdot \|p_n(\cdot)\|_{C[-1,1]},$$
$$|p_n^{(m)}(0)| \leqslant |T_n^{(m)}(0)| \cdot \|p_n(\cdot)\|_{C[-1,1]} \qquad (0 \leqslant m \leqslant n)$$

in case m and n have the same parity; also,

$$|p_n^{(m)}(0)| \leqslant m! \cdot |A_m^{n-1}| \cdot \|p_n(\cdot)\|_{C[-1,1]},$$
$$|p_n^{(m)}(0)| \leqslant |T_{n-1}^{(m)}(0)| \cdot \|p_n(\cdot)\|_{C[-1,1]} \qquad (0 \leqslant m \leqslant n-1)$$

if case m and n have opposite parity.

3.2. Extremal interpolation in the class W^{n+1}

3.2.1. The fourth extremal property of Chebyshev polynomials

Let $a \leqslant x_0 < x_1 < \ldots < x_n \leqslant b$ be interpolation nodes on $[a, b]$, let $f \in \mathfrak{M} \subset C^{(n+1)}[a, b]$, and let $p_n(x, f) = \sum_{k=0}^{n} f(x_k) l_k(x)$ be the Lagrange interpolation polynomial. We already know that

$$R_n(x, f) = f(x) - p_n(x, f) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\omega(x), \tag{3.2}$$

where $\xi \in [a, b]$ and $\omega(x) = (x - x_0) \cdots (x - x_n)$. This being so, given a class of functions $\mathfrak{M}$, how should one select the nodes so as to minimize the remainder term of the interpolation formula over the whole class?

For the class $\mathfrak{M}$, consider the quantity

$$\sup_{f \in \mathfrak{M}} \|R_n(\cdot, f, \{x_k\})\|_{C[-1,1]} = F_n(\mathfrak{M}, \{x_k\}), \tag{3.3}$$

assuming $[a, b] = [-1, 1]$. The problem is how to find the nodes so as to have (3.3) as small as possible — in other words, we need to estimate

$$\inf_{\{x_k\}} F_n(\mathfrak{M}, \{x_k\}) = \Phi_n(\mathfrak{M}).$$

If this is so, we would have, for any $f \in \mathfrak{M}$, in the extremal nodes,

$$\|R_n(\cdot, f, \{x_k\})\|_{C[-1,1]} \leqslant \Phi_n(\mathfrak{M}).$$

Note that the problem of finding nodes, for a given function, so that to minimize the norm $\|f(\cdot) - p_n(\cdot, f)\|_{C[-1,1]}$ is a *different problem*, which will not be considered here.

As $\mathfrak{M}$, we consider the class

$$W^{(n+1)} = \{f \in C^{n+1}[-1, 1] : \ \|f^{(n+1)}\|_{C[-1,1]} \leqslant 1\}.$$

Let us evaluate

$$\inf_{\{x_k\}} \sup_{f\in W^{(n+1)}} \|R_n(\cdot, f, \{x_k\})\|_{C[-1,1]}.$$

We fix $x \in [-1, 1]$. By equality (3.2) we have

$$\sup_{f\in W^{(n+1)}} |R_n(x, f, \{x_k\})| \leqslant \frac{|\omega(x)|}{(n+1)!}.$$

Since there exists a function such that $f^{(n+1)}(x) \equiv \pm 1$ (for example, if f is a polynomial with the leading coefficient $\pm\frac{1}{(n+1)!}$), this inequality is equality for any $x \in [-1, 1]$:

$$\sup_{f\in W^{(n+1)}} |R_n(x, f, \{x_k\})| = \frac{|\omega(x)|}{(n+1)!}.$$

Hence,

$$\sup_{f\in W^{(n+1)}} \|R_n(\cdot, f, \{x_k\})\|_{C[-1,1]} = \frac{\|\omega(\cdot)\|_{C[-1,1]}}{(n+1)!},$$

with the supremum attended for the same function. As a result, the problem reduces to finding

$$\inf_{\{x_k\}} \|\omega(\cdot)\|_{C[-1,1]}.$$

Since $\omega(x) = \prod_{k=0}^{n}(x - x_k) = x^{n+1} + \dots$, we have

$$\inf_{\{x_k\}} \|\omega(\cdot)\|_{C[-1,1]} \geqslant \inf_{x^{n+1}+\dots} \|p_{n+1}(\cdot)\|_{C[-1,1]} = \|\widetilde{T}_{n+1}(\cdot)\|_{C[-1,1]} = 2^{-n}.$$

In fact, this formula also holds with the equality sign, because the class of $\omega(x)$ is question is the class of all monic polynomials with zeros $x_k \in [-1, 1]$, $k = 0, \dots, n$, and this class contains the normalized Chebyshev polynomial $\widetilde{T}_{n+1}$. Hence,

$$\inf_{\{x_k\}} \sup_{f\in W^{(n+1)}} \|R_n(\cdot, f, \{x_k\})\|_{C[-1,1]} = \frac{1}{(n+1)!} \cdot 2^{-n},$$

the infimum being attended at the nodes $\{x_k\}$ which are zeros of T_{n+1}.

Problem. Find

$$\inf_{\{x_k\}} \sup_{f\in W^{(r)}} \|R_n(\cdot, f, \{x_k\})\|_{C[-1,1]}$$

for all $0 \leqslant r \leqslant n + 1$.

For $r = 0$ the problem was solved asymptotically; for $r = n + 1$ the solution was just given; for the remaining r the problem is open.

3.3. Interpolation in the complex domain

Let D be a domain of the complex plane $\mathbb{C}$, and let $w = f(z)$ be a complex-valued function on D. Suppose we are given a collection of points $z_k \in D$, $k = 0, \ldots, n$.

We construct a polynomial $p_n(z, f)$ so as to have

$$p_n(z_k) = f(z_k).$$

The Lagrange interpolation formula also holds in the complex case:

$$p_n(z,f) = \sum_{k=0}^{n} f(z_k) l_k(z), \qquad l_k(z) = \frac{\omega(z)}{\omega'(z_k)(z - z_k)}, \qquad \omega(z) = \prod_{m=0}^{n}(z - z_m).$$

Let f be an analytic regular function defined on D (the first derivative f' exists everywhere on D). Suppose that D is a simply connected domain. In order to obtain an expression for the remainder term, we consider a contour C in D containing inside all points z_k. Inside the contour C, the function $f(z)\big((z-z_0)\cdots(z-z_n)\big)^{-1}$ has singularities only at the points $z_0, \ldots, z_n$, the singularities are either removable (provided $f(z_k) = 0$) or simple poles, because $z_k \neq z_l$ for $k \neq l$. Hence, by the calculus of residues,

$$\frac{1}{2\pi i} \int_C \frac{f(t)\, dt}{(t - z_0)\cdots(t - z_n)}$$
$$= \sum_{k=0}^{n} \frac{f(z_k)}{(z_k - z_0)\cdots(z_k - z_{k-1})(z_k - z_{k+1})\cdots(z_k - z_n)} = \sum_{k=0}^{n} \frac{f(z_k)}{\omega'(z_k)}.$$

Let z be a fixed point inside the contour C. We set

$$J(z) = \frac{1}{2\pi i} \int_C \frac{f(t)\, dt}{(t - z)\prod_{k=0}^{n}(t - z_k)}.$$

Hence,

$$J(z) = \frac{f(z)}{\prod_{k=0}^{n}(z - z_k)} - \sum_{k=0}^{n} \frac{f(z_k)}{(z - z_k)\omega'(z_k)}$$

and so

$$J(z) \cdot \omega(z) = f(z) - \sum_{k=0}^{n} \frac{f(z_k)\omega(z)}{(z - z_k)\omega'(z_k)}$$
$$= f(z) - p_n(z,f) = R_n(z,f) = \frac{1}{2\pi i} \int_C \frac{\omega(z) f(t)\, dt}{(t - z)\omega(t)}$$

for any z inside C. We write $f(z)$ as a Cauchy integral

$$f(z) = \frac{1}{2\pi i} \int_C \frac{\omega(t) f(t)\, dt}{(t - z)\omega(t)}.$$

Then

$$p_n(z,f) = f(z) - R_n(z,f) = \frac{1}{2\pi i} \int_C \frac{\{\omega(t) - \omega(z)\} f(t) dt}{(t - z)\omega(t)}$$

is the interpolation polynomial.

3.4. Simplest applications of the Lagrange interpolation formula

Let f be a function on $[a, b]$. A question that is often encountered in applications is to determine the values of functionals and operators on elements f of function spaces. Of course, in general we can consider only approximate methods.

3.4.1. Evaluation of definite integrals

In what follows, we consider this problem for a simple functional $\int_a^b f(x)\,dx$. Also, for the simplest functional — the value of a function at a point — we assume that we have at our disposal an exact or approximate method of evaluation.

Let $f \in C[a, b]$, let $p_n(x, f)$ be an interpolation polynomial, let $\{x_k\}$ be interpolation nodes, and let $\{f(x_k)\}$ be the values of f at the nodes.

Instead of integrating a function, we integrate its Lagrange interpolation polynomial. We have the following quadrature formula, known as the Cotes formula:

$$\int_a^b p_n(x, f)\,dx = \sum_{k=0}^{n} f(x_k) \int_a^b l_k(x)\,dx = \sum_{k=0}^{n} A_k f(x_k) \approx \int_a^b f(x)\,dx, \tag{3.4}$$

where $l_k(x) = l_k(x; \{x_i\}_0^n)$, $A_k = A_k(n)$, $\{x_k\}_{k=0}^n$ are the nodes of the quadrature formula, and $\{A_k^n\}_{k=0}^n$ are the coefficients of the quadrature formula (or the Cotes coefficients).

A peculiarity of this quadrature formulas is that it is exact for all polynomials of degree not exceeding n; that is, if f is a polynomial of degree at most n, then this formula is exact, with equality on the right.

In order that the quadrature formula (3.4) be exact for any polynomial of degree at most n, it must be exact for the functions x^p, $p = 0, 1, \ldots, n$.

Hence, the Cotes coefficients can be easily calculated:

$$\sum_{k=0}^{n} A_k x_k^p = \int_a^b x^p dx \qquad (p = 0, 1, \ldots, n).$$

This is a system with Vandermonde determinant; hence all A_k can be determined uniquely.

Similarly, Lagrange interpolation polynomials can be used for approximate computation of the values of various functionals. The remainders (or errors) of the corresponding quadrature formulas can be expressed in terms of the norms of the functionals and the remainders of the interpolation formulae.

It turns out that cubature formulas can be constructed even when no interpolation formula is available. For example, for multivariate functions on a cube, there exist no continuous interpolation systems, but cubature formulas can be constructed exact on polynomials of a given degree.

3.4.2. Evaluation of the operator $A(f)$

A similar technique applies to approximate evaluation of linear operators, provided we know how to evaluate $A(l_k)(x)$. This trick does not work in the case of unbounded operators. For example, consider the fairly simple differentiation operator, which is additive and homogeneous. The problem of restoring $f'(x)$ from the values of f at a finite number of points (with no additional information about function f or its derivatives) makes no sense, because the differentiation operator is unbounded on the class of continuous functions; hence it is not possible to obtain a guaranteed error of an approximate formula for $f'(x)$. The differentiation operator can be made continuous on the subspace $C^{(r)}[a,b] \subset C[a,b]$ of r times continuously differentiable functions, $r \geqslant 2$, if we equip $C^{(r)}[a,b]$ with a Sobolev-type norm, putting

$$\|f\| = \|f\|_{C[a,b]} + \|f^{(r)}\|_{C[a,b]}.$$

The boundedness of differentiation operators of orders $1, 2, \ldots, r-1$ easily follows, for example, from the Taylor formula with remainder in the Cauchy form and using the following corollary of the inequality of brothers A. A. and V. A. Markov for algebraic polynomials $p_n(x)$,

$$\|p_n^{(k)}\|_{C[a,b]} \leqslant \frac{2^k}{(b-a)^k} n^{2k} \|p_n\|_{C[a,b]}.$$

For periodic function and functions defined on the whole $\mathbb{R}$, the boundedness of these differential operators on spaces $C_{2\pi}^{(r)}$ and $C^{(r)}(\mathbb{R})$ equipped with Sobolev norm follows from the corresponding Kolmogorov's inequalities and will be discussed in what follows (see § 19.4 of Lecture 19 and § 20.1 of Lecture 20).

Lecture 4

Quadrature processes and interpolation with derivatives

4.1. Quadrature formulas

Given a function space B of functions f defined on $[a, b]$, with the norm $\|f\|$, consider a quadrature formula

$$\mathcal{L}(f) = \sum_{k=1}^{n} A_k f(x_k) \tag{4.1}$$

for approximate evaluation of some functional. Here and in what follows, $A_k = A_k(n, \mathcal{L})$, $\{x_k\}_1^n$ is a collection of distinct points of $[a, b]$.

A quadrature formula is a bounded linear functional, provided the value of the function at a point is a bounded linear (additive and homogeneous) functional. Clearly, $\mathcal{L}$ is a bounded linear functional when $B = C[a, b]$. As a rule, quadrature formulas are dealt with in a class of functions coinciding either with the space $C[a, b]$ or with its subspace B equipped with a different metric; for example, if $B = C^{(r)}[a, b]$ with the Sobolev norm $\|f\| = \|f\|_{C[a,b]} + \|f^{(r)}\|_{C[a,b]}$.

Thus we will consider quadrature formulas in spaces in which the value of a function at a point is a bounded linear functional.

1. Any quadrature formula, as a linear functional, has the norm

$$\|\mathcal{L}\| = \sup_{\|f\| \leqslant 1} |\mathcal{L}(f)|.$$

This is its first characteristics. In $C[a, b]$, we have the estimate

$$\|\mathcal{L}\|_C \leqslant \sum_{k=1}^{n} |A_k|,$$

the equality holding whenever $f \in C[a, b]$ is such that $f(x_k) = \operatorname{sign} A_k$. Such a function always exists (see Fig. 4.1). Hence,

$$\|\mathcal{L}\|_C = \sum_{k=1}^{n} |A_k|.$$

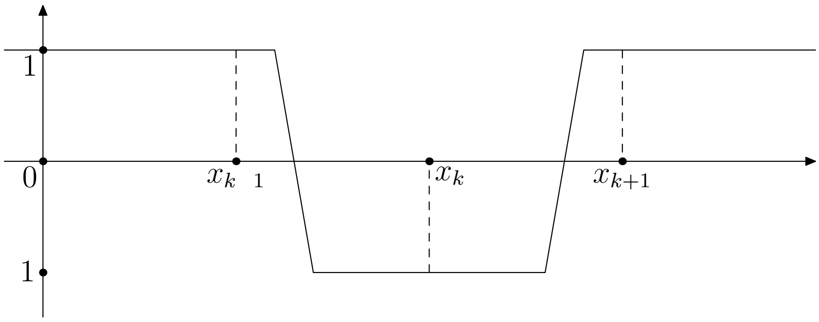

Fig. 4.1

An error of magnitude ε in the evaluation of $f(x_k)$ incurs an error of magnitude at most $\varepsilon\sum_{k=1}^{n}|A_k|$ in the quadrature formula.

2. The second characteristics of a quadrature formula is the *region of exactness*. A quadrature formula is a method of evaluating some functional, for example, the functional

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

In general, $M(f) \neq \mathcal{L}(f)$. However, for any quadrature formula and the corresponding functional M there exists a set $\mathcal{Q}$ such that

$$M(f) = \mathcal{L}(f) \qquad \forall f \in \mathcal{Q}.$$

This set $\mathcal{Q}$ is called the *region of exactness* of a quadrature formula for $M(f)$.

For any quadrature formula, the set of exactness is nonempty, because it always contains the zero function); in some cases $\mathcal{Q}$ may coincide with the whole space.

Definition. A quadrature formula is said to *have precision* m if $\mathcal{P}_m \subset \mathcal{Q}$ and $\mathcal{P}_{m+1} \not\subset \mathcal{Q}$ for some m.

In particular, a quadrature formula has precision $m = 0$ for $M(f)$ if, for any constant function c, we have $M(c) = \mathcal{L}(c) = c\cdot\sum_{k=1}^{n} A_k$ and $M(f) \neq \sum_{k=1}^{n} A_k x_k$ for $f(x) \equiv x$. If $M(f) = \int_a^b f(x)\,dx$, we have $M(c) = (b-a)c$, and so a necessary and sufficient condition that a quadrature formula (4.1) have precision $m = 0$ is that

$$\sum_{k=1}^{n} A_k = b - a \quad \text{and} \quad \sum_{k=1}^{n} A_k x_k \neq \frac{1}{2}(b^2 - a^2).$$

The first of this conditions is necessary and sufficient that formula (4.1) for the functional $\int_a^b f(x)\,dx$ have precision $m \geqslant 0$. In the general case, any quadrature formula has precision at least m, provided it is exact on $1, x, x^2, \ldots, x^m$.

4.2. Quadrature processes and their convergence

A quadrature process is a sequence of quadrature formulas $\mathcal{L}_n$, $n \in \mathbb{N}$, for a given linear functional $M(f)$ on $C[a,b]$ (see [31, Ch. VI]).

Problem. Under what conditions we have, for any function $f \in C[a,b]$,

$$\mathcal{L}_n(f) \to M(f) \qquad (n \to \infty)?$$

This is the weak convergence of functionals. Hence, the sequence $\{\mathcal{L}_n\}$ must be weakly convergent.

4.2.1. Sufficient conditions for convergence of a quadrature process

Theorem 4.1. *A sufficient condition for a quadrature process to converge is that*:

1) *the sequence of the norms of quadrature formulas* $\{\mathcal{L}_n\}$ *is bounded*;
2) $m(n) \to \infty$ *as* $n \to \infty$*; here,* $m(n)$ *is the precision of the quadrature formula* $\mathcal{L}_n$.

Proof. There is no loss of generality in assuming that $m(n)$ increases with n. Hence $\mathcal{L}_n(p) = M(p)$ for any polynomial $p \in \mathcal{P}_m$ for $m(n) \geqslant m$; consequently, $L_n(p) \to M(p)$ as $n \to \infty$. Therefore, for any m, the quadrature process converges on $\mathcal{P}_m$ and hence on $\cup_m \mathcal{P}_m$. Further, by the Weierstrass theorem, $\overline{\bigcup_m \mathcal{P}_m} = C[a,b]$, and hence the quadrature process converges on a dense set, the norms $\|\mathcal{L}_n\|_C$ being uniformly bounded. As a result, the quadrature process converges on $C[a,b]$. □

Remark. The Cotes quadrature process (3.4) satisfies condition 2) of Theorem 4.1, because $m(n) = n$ (see § 3.4.1). Since the quadrature formulas are exact for constant functions on $[a,b]$, we have $\sum_{k=1}^n A_k = b - a$ in (3.4). If, in addition, all the Cotes coefficients $A_k = A_k(n)$ are nonnegative, then the norms of the quadrature process,

$$\sum_{k=1}^{n} |A_k| = \sum_{k=1}^{n} A_k = b - a,$$

are bounded and hence the Cotes quadrature process satisfies the hypotheses of Theorem 4.1.

4.2.2. Some quadrature processes and their remainder terms

The Cotes quadrature formula was already used in (3.4) to evaluate the integral $M(f) = \int_a^b f(x)\,dx$:

$$\mathcal{L}_n(f) = \int_a^b p_n(x,f)\,dx = \sum_{k=0}^{n} A_k f(x_k), \qquad A_k = A_k(n);$$

here $p_n(x,f)$ is the Lagrange interpolation polynomial on the mesh $\{x_k\}_{k=0}^n \subset [a,b]$, $A_k = \int_a^b l_k(x)\,dx$. This quadrature formula has precision at least n, and also

$$\|\mathcal{L}_n(f)\| = \sum_{k=1}^{n} |A_k|.$$

Given an $f \in C^{(n+1)}[a,b]$, suppose that $\|f^{(n+1)}\|_{C[a,b]} = M_{n+1}$. Then

$$\Big|\int_a^b f(x)\,dx - \mathcal{L}_n(f)\Big| = \Big|\int_a^b (f(x) - p_n(x,f))\,dx\Big|$$
$$= \Big|\int_a^b \frac{f^{(n+1)}(\xi)}{(n+1)!}\omega(x)\,dx\Big| \leqslant \frac{M_{n+1}}{(n+1)!}\int_a^b |\omega(x)|\,dx.$$

Similar bounds can also be written in terms of M_r for $0 \leqslant r \leqslant n+1$, using estimates (2.1) for $|f(x) - p_n(x,f)|$ via $\|f^{(r)}\|$.

4.2.3. Gauss-type quadrature formulas

We already have at our disposal a quadrature formula $\sum_{k=0}^n A_k f(x_k)$ to approximately evaluate the integral $\int_a^b f(x)\,dx$, which is exact for any $f \in \mathcal{P}_n$. The formula is constructed at the nodes $\{x_k\}_{k=0}^n$ and has $n+1$ coefficients $\{A_k\}_{k=0}^n$; hence it is determined by $2(n+1)$ parameters.

Theorem 4.2 (see, e.g., [33]). *Let L_{n+1} be an $(n+1)$-dimensional subspace of $C[a,b]$. Then there exists an $(n+1)$-point quadrature formula which is exact for any $f \in L_{n+1}$.*

Proof. Let $\varphi_0(x), \ldots, \varphi_n(x)$ be a system of linearly independent functions in L_{n+1}. Then there exist points x_i, $i = 0, \ldots, n$, such that

$$\det |\varphi_j(x_i)| \neq 0.$$

This can be proved by induction on n, starting with $\varphi_0(x) \not\equiv 0$ on $[a,b]$. We take the points $\{x_i\}_{i=0}^n$ as the nodes of the quadrature formula. Then we have the system of linear equations for A_k,

$$\sum_{k=0}^{n} A_k \varphi_j(x_k) = \int_a^b \varphi_j(x)\,dx, \qquad j = 0, \ldots, n,$$

which uniquely determines A_k (the determinant of the system is nonzero). The obtained formula is exact for any $f \in L_{n+1}$, because f is linearly expressible via φ_k, $k = 0, \ldots, n$, and since the formula is exact for φ_k. □

Problem. Let L_m be an m-dimensional subspace of $C[a,b]$, and let $n+1 < m \leqslant 2(n+1)$. When there exists an $(n+1)$-point quadrature formula that is exact on the whole L_m?

The answer is unknown. For $L_m = \mathcal{P}_{m-1}$ the problem was solved by Gauss.

Theorem 4.3 (Gauss; see, e.g., [21, § 2.4]). *There exist nodes* $\{x_0, x_1, \ldots, x_n\}$ *and coefficients* $A_0, A_1, \ldots, A_n$ *such that the quadrature formula is exact for any polynomial* $p \in \mathcal{P}_{2n+1}$*, that is,*

$$\sum_{k=0}^{n} A_k p(x_k) = \int_a^b p(x)\,dx.$$

Proof. 1) Suppose that we have a quadrature formula which is exact for any $p \in \mathcal{P}_{2n+1}$. Consider the polynomial $\omega(x) = (x-x_0)\cdots(x-x_n) \in \mathcal{P}_{n+1}$ and take an arbitrary $q \in \mathcal{P}_n$. We claim that

$$\int_a^b \omega(x)q(x)\,dx = 0, \qquad q \in \mathcal{P}_n;$$

in other words, we claim that $\omega(x)$ is orthogonal to any polynomial of degree at most n.

Indeed, $p(x) = \omega(x)q(x)$ is a polynomial of degree at most $2n+1$. By the assumption, the quadrature formula is exact for $p(x)$; that is,

$$\int_a^b \omega(x)q(x)\,dx = \sum_{k=0}^{n} A_k p(x_k) = 0.$$

Here the sum is zero, because $p(x_k) = \omega(x_k)q(x_k) = 0 \cdot q(x_k) = 0$. Hence, the polynomial $\omega(x)$ is orthogonal to any polynomial of degree at most n.

Thus, if there exists a Gauss quadrature formula, then its nodes are roots of a polynomial that is orthogonal to any q of $\mathcal{P}_n$. Such polynomials can be obtained by orthogonalizing the system of functions $\{x^k\}_{k=0}^{n+1}$ on $[a, b]$ (for example, by using the Gram–Schmidt process with respect to the inner product $(f, g) = \int_a^b f(x)g(x)dx$). This gives Legendre polynomials $\{P_k\}$. The Legendre polynomial P_{n+1} has the required properties: this is a monic polynomial, is orthogonal to the subspace $\mathcal{P}_n$, its zeros are simple and lie in $[a, b]$. The last assertion follows from the following simple observation: assuming that the polynomial P_{n+1} has at most n changes of sign on $[a, b]$, the polynomial $q \in \mathcal{P}_n$ vanishes at these points, giving $\left|\int_a^b q(x)P_{n+1}(x)\,dx\right| = \int_a^b |q(x)|\,|P_{n+1}(x)|\,dx > 0$.

2) We take the zeros $\{x_k\}_{k=0}^n$ of the Legendre polynomial P_{n+1} as the required nodes of the sought-for quadrature formula; the polynomial P_{n+1} is orthogonal to any polynomial of degree at most n. Consider the Cotes quadrature formula $\sum_{k=0}^n A_k f(x_k)$ with nodes at $\{x_k\}_{k=0}^n$; this formula is exact for any $q \in \mathcal{P}_n$. We claim that this is a Gauss-type formula; that is, this formula is exact for polynomials of degree at most $2n+1$.

Let p be any polynomial in $\mathcal{P}_{2n+1}$. By the construction, the polynomial $\omega(x) = \prod_{k=0}^n (x - x_k) = P_{n+1}(x)$ lies in $\mathcal{P}_{n+1}$ and coincides with the Legendre polynomial. We write

$$p(x) = q(x) \cdot \omega(x) + r(x),$$

where $q, r \in \mathcal{P}_n$. Note that $p(x_k) = r(x_k)$. Since the polynomial $\omega = P_{n+1}$ is orthogonal to q and since the formula is exact for r, we have

$$\int_a^b p(x)\,dx = \int_a^b q(x)\omega(x)dx + \int_a^b r(x)\,dx = \int_a^b r(x)\,dx$$
$$= \sum_{k=0}^{n} A_k r(x_k) = \sum_{k=0}^{n} A_k p(x_k),$$

and so

$$\int_a^b p(x)\,dx = \sum_{k=0}^{n} A_k p(x_k), \qquad p \in \mathcal{P}_{2n+1}.$$

Hence, the obtained formula is a Gauss-type formula. □

Remark. A Gauss quadrature formula exists for any integer $n \geqslant 0$ (see [33, Theorem 5.9]).

Theorem 4.4. *All Cotes coefficients in a Gauss quadrature formula are nonnegative.*

Proof. Indeed, for any polynomial $p \in \mathcal{P}_{2n+1}$,

$$\sum_{k=0}^{n} A_k p(x_k) = \int_a^b p(x)\,dx; \tag{4.2}$$

in particular, this formula holds for the square l_k^2 of any Lagrange fundamental polynomial l_k with nodes at $x_0, x_1, \dots, x_n$ (the degree of the squared polynomial is at most $2n$). We have $l_m^2(x_k) = \delta_{km}$, and hence, substituting $p = l_m^2$ in (4.2), it follows that

$$A_k = \int_a^b l_k^2(x)dx > 0, \qquad k = 0, 1, \dots, n,$$

the result required. □

Note that since the quadrature formula is also exact on the constant functions, we have

$$\sum_{k=0}^{n} |A_k| = \sum_{k=0}^{n} A_k = b - a.$$

Hence, the hypotheses of Theorem 4.1 on convergence of a quadrature are satisfied (the norms are bounded and the precision $m = 2n+1$ increases to ∞ as $n \to \infty$). We have the following result.

Proposition 4.1. *For any $f \in C[a, b]$, the Gauss quadrature process $\sum_{k=0}^{n} A_k(n) f(x_k^n)$ converges to $\int_a^b f(x)\,dx$.*

4.3. Interpolation with derivatives

The general Birkhoff's interpolation problem is as follows (see, e.g., [27], [25]). Let $f \in C^{(m)}[a, b]$. Suppose we are given points $x_1, x_2, \dots, x_k$, and the table of parameters

x_1	$0 \leqslant r_0^{(1)} < r_1^{(1)} < \dots < r_{s_1}^{(1)} \leqslant m$	$s_1 + 1$
x_2	$0 \leqslant r_0^{(2)} < r_1^{(2)} < \dots < r_{s_2}^{(2)} \leqslant m$	$s_2 + 1$
.	$\dots$	.
.	$\dots$	.
.	$\dots$	.
x_k	$0 \leqslant r_0^{(k)} < r_1^{(k)} < \dots < r_{s_k}^{(k)} \leqslant m$	$s_k + 1$

In total, there are $N = s_1 + s_2 + \ldots + s_k + k$ parameters $r_i^{(j)}$. It is required to build a polynomial $p \in \mathcal{P}_{N-1}$ such that, for $s = 0, 1, \ldots, s_i$ and $i = 1, 2, \ldots, k$,

$$p^{(r_s^{(i)})}(x_i) = f^{(r_s^{(i)})}(x_i).$$

Which conditions should be placed on numbers $r_s^{(i)}$ so that the problem in question be solvable for any $x_1, \ldots, x_k \in [a, b]$? This is an open problem. In certain cases the problem is solvable. However, even in some simple cases the problem fails to have a solution.

Exercise. Construct a table such that the problem would not always be solvable (hint: take 3–4 points and derivatives of orders not exceeding 2).

4.3.1. Interpolation with multiple nodes

The Birkhof's problem is solvable in a particular case of interpolation with multiple nodes (see [23]). Given nonnegative integers $s_1, \ldots, s_k$ it is required to construct a polynomial such that

$$p^{(r)}(x_i) = f^{(r)}(x_i), \qquad 0 \leqslant r \leqslant s_i \qquad (i = 1, \ldots, k).$$

Here, $N = s_1 + s_2 + \cdots + s_k + k$, p is a polynomial of degree $N - 1$, and a function f is continuously differentiable as many times as necessary.

Theorem 4.5. *An interpolation problem with multiple nodes is always uniquely solvable.*

Proof. The problem being linear, it suffices to show that the corresponding homogeneous problem always has a unique (zero) solution in the class of polynomials of degree $N - 1$. In other words, if $a \leqslant x_1 < x_2 < \ldots < x_k \leqslant b$, then we need to show that if p is a polynomial such that

$$p^{(r)}(x_i) = 0, \qquad 0 \leqslant r \leqslant s_i, \qquad i = 1, \ldots, k,$$

then p identically vanishes.

In the case in question any point x_i, $i = 1, \ldots, k$, is a zero of multiplicity $s_i + 1$ of the polynomial p. Hence $p(x)$, which has degree $N - 1$, must divide the polynomial $\prod_{i=1}^{k}(x - x_i)^{s_i+1}$. This, however, is possible only if $p \equiv 0$. □

4.3.2. Hermite interpolation problem

Suppose in an interpolation with multiple nodes we have $s_1 = s_2 = \cdots = s_n = 1$ and

$$p(x_k) = f(x_k), \qquad p'(x_k) = f'(x_k) \qquad (k = 1, \ldots, n).$$

Hence $N = 2n$ and so the polynomial p must be of degree $2n - 1$. We can write

$$p(x) = \sum_{k=1}^{n}\{f(x_k)A_k(x) + f'(x_k)B_k(x)\},$$

where $A_k(x)$ and $B_k(x)$ are fundamental polynomials of Hermite interpolation; that is,

$$\begin{cases} A_k(x_i) = \delta_{i,k}, \\ A'_k(x_i) = 0, \end{cases} \qquad \begin{cases} B_k(x_i) = 0, \\ B'_k(x_i) = \delta_{i,k}, \end{cases} \tag{4.3}$$

and A_k, B_k are polynomials of degree at most $2n-1$. Let us find them explicitly. Let $l_k(x)$ be the fundamental Lagrange polynomials corresponding to the same nodes $x_1 < x_2 < \ldots < x_n$; i.e., $l_k \in \mathcal{P}_{n-1}$, $l_k(x_i) = \delta_{i,k}$ and

$$l_k(x) = \frac{\omega(x)}{(x-x_k)\omega'(x_k)},$$

where $\omega(x) = \prod_{k=1}^{n}(x-x_k)$. Clearly, the polynomials $A_k(x) = \left\{1-(x-x_k)2l'_k(x_k)\right\} l_k^2(x)$, $B_k(x) = (x-x_k)\, l_k^2(x)$ of degree at most $2n-1$ satisfy equalities (4.3).

4.3.3. Remainder of the interpolation problem with multiple nodes

Suppose, for an interpolation problem with multiple nodes, we are given the table of parameters $x_1, \ldots, x_k$, $s_1, \ldots, s_k$, where $s_1 + s_2 + \ldots + s_k + k = N$. The degree of the polynomial is $N-1$. Let $H(x) = H(x, f)$ be a unique polynomial of multiple interpolation for a function f. The following result for the remainder term of this interpolation formula can be proved with the help of Rolle's theorem.

Theorem 4.6. *Let $f \in C^{(N)}[a,b]$. Then, for any $x \in [a,b]$, there exists a point $\xi \in (a,b)$ such that*

$$f(x) - H(x,f) = \frac{f^{(N)}(\xi)}{N!}\Omega(x),$$

where $\Omega(x) = \prod_{j=1}^{k}(x-x_j)^{s_j+1}$ is a polynomial of degree N.

Proof. For $x = x_i$ the formula holds, because $\Omega(x_i) = 0$, $i = 1, \ldots, k$. Hence we may assume that $x \neq x_i$, $x \in [a,b]$, $i = 1, \ldots, k$. For the following auxiliary function

$$\varphi(z) = f(z) - H(z) - \frac{f(x) - H(x)}{\Omega(x)}\,\Omega(z),$$

we have $\varphi(x) = 0$, and further, since $\Omega^{(r)}(x_i) = 0$ for $r = 0, 1, \ldots, s_i$, it follows that

$$\varphi^{(r)}(x_i) = 0 \qquad (r = 0, 1, \ldots, s_i; \quad i = 1, \ldots, k).$$

Hence φ has precisely $s_1 + \ldots + s_k + k + 1 = N + 1$ zeros (counting multiplicity), and so by Rolle's theorem, there exists a point $\xi \in (a,b)$ such that $\varphi^{(N)}(\xi) = 0$, giving

$$f^{(N)}(\xi) - \frac{f(x) - H(x)}{\Omega(x)} N! = 0.$$

□

Corollary. *Let $f \in C^{(N)}[a,b]$ and $x \in [a,b]$. Then*

$$|f(x) - H(f,x)| \leqslant \frac{|\Omega(x)|}{N!}\,\|f^{(N)}\|_{C[a,b]}.$$

Lecture 5

Fourier series. Fejér sums

5.1. Some elementary results concerning Fourier series

We shall be concerned with periodic functions of period $\omega = 2\pi$. Let $H_{2\pi}$ be some class of 2π-periodic functions defined on the whole real axis. To every measurable function $f \in L_{2\pi} = L[0, 2\pi)$ that is Lebesgue integrable over the primitive period we can assign its Fourier series

$$f(x) \sim \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k \cos kx + b_k \sin kx) = \sum_{k=0}^{\infty} A_k(x), \tag{5.1}$$

where

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos kx\, dx, \qquad b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin kx\, dx$$

are the Fourier coefficients of the function f. This is only a formal correspondence, because for $f \in L_{2\pi}$ the series may fail to converge to f almost everywhere (Kolmogorov's example) or in the norm of the space $L_{2\pi}$.

Consider the sequence of partial sums of the Fourier series

$$s_n(f) = s_n(f, x) = \sum_{k=0}^{n} A_k(x) \in \mathcal{T}_n.$$

The linear operator S_n acts on the space $L_{2\pi}$ and maps any function f into an element of the subspace $\mathcal{T}_n$ of trigonometric polynomials of order n:

$$S_n :\ L_{2\pi} \longrightarrow \mathcal{T}_n, \qquad S_n f = s_n(f).$$

Let $G, H \subset L_{2\pi}$ be normed linear spaces with norms $\|\cdot\|_G$ and $\|\cdot\|_H$. As usual, let

$$\|S_n\|_{H\to G} = \sup_{\|f\|_H \leqslant 1} \|s_n(f)\|_G$$

be the norm of S_n as an operator from H to G. In case $G = H$, we write $\|S_n\|_{H\to H} = \|S_n\|_H$ or simply $\|S_n\|$, if no confusion can arise.

According to Bessel's inequality, $\|s_n(f)\|_{L^2} \leqslant \|f\|_{L^2}$ for all $f \in L^2_{2\pi}$, and since $s_n(f) = f$ for $f \in \mathcal{T}_n$, we have

$$\|S_n\|_{L^2} = 1.$$

Theorem 5.1. *If the subset of trigonometric polynomials is dense in H and if the Lebesgue constants $\|S_n\|_H$ are uniformly bounded, then*

$$s_n(f) \xrightarrow{H} f \qquad (n \to \infty) \qquad \forall\, f \in H,$$

i.e.,

$$\|f - s_n(f)\|_H \to 0 \qquad (n \to \infty) \qquad \forall\, f \in H.$$

This fact follows from the pointwise convergence theorem for a sequence of linear operators in a complete normed linear space X, if:

1) the norms of the operators are uniformly bounded;

2) the sequence of the operators converges on a dense subset Γ of X, then it converges on X.

Indeed, as Γ we may take the set $\mathcal{T} = \cup_n \mathcal{T}_n$.

Remark. The Lebesgue constants are bounded in any space $L^p_{2\pi}$ for $1 < p < \infty$,

$$\|S_n\|_{L^p} \leqslant A_p.$$

This result is stated without proof.

Let L_n be the Lebesgue constant in the space $C_{2\pi}$ of continuous 2π-periodic functions, $\|S_n\|_{C_{2\pi}} = L_n$.

Theorem 5.2. *As $n \to \infty$, we have, in order of magnitude,*

$$L_n \asymp \ln n;$$

that is, there exists constants a and A, $0 < a \leqslant A < \infty$, and a natural number n_0 such that

$$\forall\, n \geqslant n_0 \qquad a \leqslant \frac{L_n}{\ln n} \leqslant A.$$

In this case, we say that L_n has the order of magnitude $\ln n$. The order relation '$\asymp$' (the Hardy symbol) is symmetric, transitive and reflexive.

Proof. 1) We first claim that $L_n = O(\ln n)$; we need to show that there exists A such that $L_n \leqslant A \ln n$ for any $n \geqslant n_0$. We apply Dirichlet's formula for partial sums of Fourier series,

$$s_n(f, x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x+t) \mathcal{D}_n(t)\, dt,$$

where

$$\mathcal{D}_n(t) = \frac{1}{2} + \sum_{k=1}^{n} \cos kt = \frac{\sin\left(n + \frac{1}{2}\right) t}{2 \sin \frac{t}{2}}$$

is the Dirichlet kernel. The graph of $D_n(t)$ is depicted in Fig. 5.1. If $\|f\|_C \leqslant 1$ (i.e., $|f(x)| \leqslant 1$ for all x), we have

$$|s_n(f, x)| \leqslant \frac{1}{\pi} \int_{-\pi}^{\pi} |\mathcal{D}_n(t)|\, dt.$$

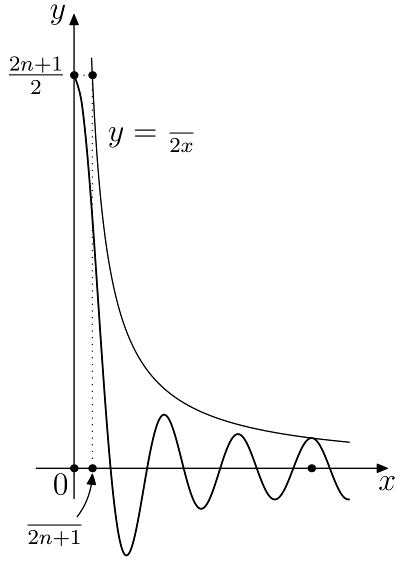

Fig. 5.1.

Since for $\mathcal{D}_n(t)$ we have, for some $C_1 > 0$ and any $n \geqslant 1$,

$$|\mathcal{D}_n(t)| \leqslant \frac{1}{2\left|\sin\frac{t}{2}\right|} \leqslant \frac{C_1}{|t|} \qquad \forall\ t,\ |t| \leqslant \pi,$$

$$|\mathcal{D}_n(t)| \leqslant n + \frac{1}{2} \leqslant C_1 n \qquad \forall\ t,$$

it follows that

$$|s_n(f,x)| \leqslant \frac{1}{\pi}\int_{-\pi}^{\pi} |\mathcal{D}_n(t)|\,dt = \frac{2}{\pi}\int_0^{\pi} |\mathcal{D}_n(t)|\,dt$$

$$\leqslant \frac{2}{\pi}C_1\left\{\int_0^{\frac{1}{n}} n\,dt + \int_{\frac{1}{n}}^{\pi}\frac{dt}{t}\right\} \leqslant C_2\{1+\ln n\} \leqslant A\ln n \quad \forall\ n \geqslant 2,$$

where A is some absolute constant.

2) Now we claim that $L_n \geqslant c\ln n$ for some $c > 0$; that is,

$$\|S_n\|_C = \sup_{\|f\|_C \leqslant 1} \|s_n(f)\|_C \geqslant c\ln n,$$

or, what is the same,

$$\forall\ n \in \mathbb{N} \quad \exists\ f_n \in C, \quad \|f_n\|_C \leqslant 1 :\ \|s_n(f_n)\|_C \geqslant c\ln n.$$

It suffices to show that there exists a point $x_0 \in [0, 2\pi)$ and functions $f_n \in C_{2\pi}$ such that

$$\|f_n\|_C \leqslant 1, \qquad |s_n(f_n, x_0)| \geqslant c\ln n.$$

We first show that the polynomials

$$d_n(x) = \sum_{k=1}^{n} \frac{\sin kx}{k}$$

are uniformly bounded: there exists B such that $|d_n(x)| \leqslant B$ for all n and x.

Given $x > 0$, consider a natural number $m = m(x) \asymp \frac{1}{x}$; in other words, we take $0 < x \leqslant \pi$ so that the inequality

$$\frac{A_1}{x} \leqslant m(x) \leqslant \frac{A_2}{x}$$

holds with some numbers $0 < A_1 < A_2 < \infty$ independent of x. Hence, for $n > m$,

$$|d_n(x)| = \left|\sum_{k=1}^{m} \frac{\sin kx}{k} + \sum_{k=m+1}^{n} \frac{\sin kx}{k}\right| \equiv |s_1 + s_2| \leqslant |s_1| + |s_2|.$$

By the choice of m,

$$|s_1| = \left|\sum_{k=1}^{m} \frac{\sin kx}{k}\right| \leqslant \sum_{k=1}^{m} \frac{kx}{k} = mx \leqslant A_2.$$

To estimate s_2, we recall Abel's inequality (see, e.g., [50, § 2.301]):

$$\text{if } \Big|\sum_{1}^{p} a_k\Big| \leqslant A \text{ for all } p, b_k \geqslant 0, \quad b_k \downarrow, \quad \text{then} \quad \Big|\sum_{k=m}^{n} a_k b_k\Big| \leqslant 2Ab_m.$$

Also, we use the formula (see, e.g., [45, Ch. 5, § 7])

$$\widetilde{\mathcal{D}}_n(x) = \frac{\sin \frac{nx}{2} \cdot \sin \frac{(n+1)x}{2}}{\sin \frac{x}{2}}$$

for the conjugate Dirichlet kernel $\widetilde{\mathcal{D}}_n = \sum_{k=1}^{n} \sin kx$. As a result,

$$\Big|\sum_{k=1}^{p} \sin kx\Big| \leqslant \frac{C_1}{|x|}, \qquad |s_2| \leqslant \frac{C_1}{|x|} \cdot \frac{1}{m} \leqslant \frac{C_1}{A_1}.$$

Hence, $|s_2|$ are uniformly bounded. Further, if $n \leqslant m(x)$ (and so $n \leqslant \frac{A_2}{x}$), we have $s_2 = 0$ (here, as usual, it assumed that $\sum_{k=m}^{n} \alpha_k = 0$ for $n < m$) and now $d_n(x)$ can be estimated similarly to s_1:

$$|d_n(x)| \leqslant nx < A_2.$$

Hence, for all n and x,

$$|d_n(x)| \leqslant B,$$

where B is some absolute constant.

Given a fixed natural $n \geqslant 2$, the Fejér polynomials are defined as follows:

$$A_n(x) = \frac{\cos x}{n-1} + \frac{\cos 2x}{n-2} + \cdots + \frac{\cos(n-1)x}{1},$$

$$B_n(x) = \frac{\cos(n+1)x}{1} + \frac{\cos(n+2)x}{2} + \cdots + \frac{\cos(2n-1)x}{n-1}.$$

Consider the function

$$f_n(x) = A_n(x) - B_n(x).$$

We have

$$s_n(f_n) = A_n(x), \qquad A_n(0) = \sum_{k=1}^{n-1} \frac{1}{k} \asymp \ln n.$$

Further,

$$\begin{aligned} f_n(x) &= \sum_{k=1}^{n-1} \left\{ \frac{\cos kx}{n-k} - \frac{\cos(n+n-k)x}{n-k} \right\} \\ &= 2 \sin nx \sum_{k=1}^{n-1} \frac{\sin(n-k)x}{n-k} = 2 \sin nx d_{n-1}(x), \end{aligned}$$

giving $|f_n(x)| \leqslant 2B$. Hence, $f_n(x)$ is a uniformly bounded sequence of functions such that $s_n(f_n)(0) \asymp \ln n$. Hence, if $f_n^*(x) = \frac{f_n(x)}{2B}$, we have $\|f_n^*\| \leqslant 1$ and

$$L_n \geqslant \|s_n(f_n^*)\|_C \geqslant c \ln n.$$

Consequently, the Lebesgue constants L_n have the order of magnitude $\ln n$. □

Thus we have showed that for any $n \in \mathbb{N}$ there exists a 2π-periodic function $f_n \in C_{2\pi}$, $\|f_n\|_C \leqslant 1$, such that $\|s_n(f_n)\|_C \geqslant a \ln n$. We ask the question: Is it possible that the function f be independent of n? In other words, is it true that

$$\exists\, f \in C_{2\pi} \qquad \exists\, a > 0 \qquad \forall\, n > 1 \qquad \|s_n(f)\|_C \geqslant a \ln n \ ?$$

This assertion fails to hold. Namely, for any function $f \in C[0, 2\pi]$ we have $\|s_n(f)\|_C = o(\ln n)$ as $n \to \infty$. In its turn, this relation cannot be improved over the whole class $C[0, 2\pi]$. This is a corollary of the following result, which we state without proof.

Theorem 5.3 (D. E. Men'shov). *For every function φ such that $\varphi(n) = o(\ln n)$ as $n \to \infty$, there exists a continuous 2π-periodic function $f = f_\varphi$ such that, for all sufficiently large n,*

$$\|s_n(f)\|_C \geqslant \varphi(n).$$

A result similar to Theorem 5.2 also holds in the space $L_{2\pi}$:

$$\|S_n\|_L \asymp \ln n, \qquad n \to \infty.$$

Consequently, in $L^p_{2\pi}$ the Lebesgue constants are uniformly bounded and, for any $p \in (1, \infty)$, have the order of magnitude $\ln n$ both in $L_{2\pi}$ and in $C_{2\pi}$.

The norms of the operator S_n in $L^p_{2\pi}$ are depicted in Fig. 5.2 as functions of p.

In general, among all spaces $L^p_{2\pi}$, $1 \leqslant p \leqslant \infty$, the space $C_{2\pi}$ is the worst and the space $L^2_{2\pi}$ is the best from the standpoint of approximation by Fourier sums; also, the properties of Fourier sums in the space $L_{2\pi}$ resemble those in $C_{2\pi}$, their properties are slightly better in the former space; the spaces $L^p_{2\pi}$, $1 < p < \infty$, are all similar to $L^2_{2\pi}$. In $L_{2\pi}$ and in $C_{2\pi}$ the norms of the operators S_n are unbounded, hence there

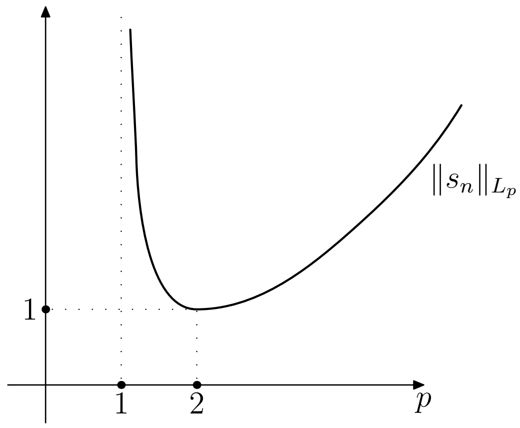

Fig. 5.2.

exist functions whose Fourier series fail to converge in $L_{2\pi}$ and in $C_{2\pi}$; moreover, their partial sums are unbounded. In contrast, for any function $f \in L^p_{2\pi}$, $1 < p < \infty$, its Fourier series (5.1) converges in $L^p_{2\pi}$ to f; that is, $\|f - S_n(f)\|_{L^p} \to 0$ as $n \to \infty$. Taking this into account, for a function $f \in L^p_{2\pi}$, $1 < p < \infty$, the equality sign '=' is used in (5.1) instead of the correspondence sign '($\sim$)'.

It is worth pointing out the following celebrated result by Carleson (1966, for $p = 2$) and Hunt (1967, for $1 < p < \infty$): if $f \in L^p_{2\pi}$, $1 < p < \infty$, then $s_n(f, x) \to f(x)$ as $n \to \infty$ a.e. (see, e.g., [15]).

5.2. Fejér sums

Given a 2π-periodic function $f \in L_{2\pi}$, the polynomial

$$\sigma_n(f, x) = \frac{1}{n+1} \sum_{k=0}^{n} s_k(f, x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x+t) K_n(t)\, dt, \tag{5.2}$$

is called the Fejér sum; here $K_n(t)$ is the Fejér kernel, which is the arithmetic mean of the first n Dirichlet kernels.

The following properties Fejér sums can be easily established.

1. *Main property.* It follows by (5.2) that if $f(x) \geqslant 0$ for all x, then $\sigma_n(f, x) \geqslant 0$ for all x and n, because the Fejér kernels are nonnegative:

$$K_n(t) = \frac{1}{n+1} \sum_{k=0}^{n} \mathcal{D}_k(t) = \frac{\sin^2 (n+1)\frac{t}{2}}{2(n+1)\sin^2 \frac{t}{2}} \geqslant 0.$$

The operators with this property are called *positive.*

Let $K^+ = \{f(x) \geqslant 0\}$ be the cone of all nonnegative functions in $C_{2\pi}$ (a set K is a cone if $\lambda f \in K$ whenever $f \in K$, $\lambda > 0$). Interior points of K^+ are strictly positive functions. Speaking of the cone K^+ in $L^p_{2\pi}$, we shall assume that $K^+ \subset L^p$ and that the topology on K^+ is induced by the topology of $L^p_{2\pi}$. In $L^p_{2\pi}$, $1 \leqslant p < \infty$, the cone K^+ has no interior points. In contrast to these spaces, the cone K^+ in $C_{2\pi}$ has nonempty interior.

2. Fejér sums map the cone of positive functions into itself:

$$\sigma_n(K^+) \subset K^+.$$

3. Fejér sums preserve constant functions:

$$\sigma_n(A, x) = A.$$

4. The norm of a Fejér sum in $C_{2\pi}$ is 1:

$$\|\sigma_n\|_C = 1.$$

Indeed, if $\|f(x)\|_C \leqslant 1$, we have

$$\left|\frac{1}{\pi}\int_{-\pi}^{\pi} f(x+t)K_n(t)\,dt\right| \leqslant \frac{1}{\pi}\int_{-\pi}^{\pi} K_n(t)\,dt = 1,$$

inasmuch as $K_n(t) = \frac{1}{2} + \sum_{k=1}^{n}\bigl(1 - \frac{k}{n+1}\bigr)\cos kt$. On the other hand, $\sigma_n(1,x) \equiv 1$, and hence

$$\|\sigma_n\|_C = \sup_{\|f\|_C \leqslant 1} \|\sigma_n(f,x)\|_C = 1.$$

A similar result holds in the $L^p_{2\pi}$ spaces:

$$\|\sigma_n\|_{L^p} = 1 \qquad (1 \leqslant p < \infty).$$

The proof runs along similar lines.

The following result may be found in advanced calculus books.

Theorem 5.4 (Fejér, 1904; see, e.g., [30, Ch. 8]). *For any continuous function f, the Fejér sums converge uniformly to f,*

$$\|\sigma_n(f) - f\|_C \to 0, \qquad n \to \infty.$$

Lecture 6

Approximation of continuous functions by Fourier sums. De la Vallée Poussin sums

6.1. Approximation by Fourier and Fejér sums in $C_{2\pi}$

Let f be a 2π-periodic continuous function, and let $s_n(f)$ be the nth partial sum of the Fourier series of f.

1) According to Theorem 5.2, the norms $\|S_n\|_C$ have the order of magnitude $\ln n$ as $n \to \infty$.

Remark. Actually, a more precise relation holds:

$$\|S_n\|_C = \frac{4}{\pi^2} \ln n + c + O\left(\frac{1}{n}\right), \qquad n \to \infty;$$

here c is some absolute constant.

2) The Lebesgue inequality

$$\|f - s_n(f)\|_C \leqslant (\|S_n\|_C + 1) E_n(f)_C \tag{6.1}$$

is an important property of approximation by Fourier sums; here $E_n(f)_C$ is the best approximation of a function $f \in C_{2\pi}$ by trigonometric polynomials of order n.

We mention some other approximative properties of Fourier sums.

3a) *Approximative test for uniform convergence of Fourier series.* We know that the Fourier series does not always uniformly converge to a continuous function. What are sufficient conditions for uniform convergence? From (6.1) it follows that if

$$E_n(f)_C \ln n \to 0 \qquad (n \to \infty),$$

that is, if

$$E_n(f)_C = o\left(\frac{1}{\ln n}\right) \qquad (n \to \infty),$$

then

$$\|f - s_n(f)\|_C \to 0 \qquad (n \to \infty).$$

The same can be said in general. If $E_n(f)_C = \varphi(n)$, then we know the rate of approximation of f by its Fourier sums:

$$\|f - s_n(f)\|_C = O\big(\varphi(n) \ln n\big) \qquad (n \to \infty).$$

Consequently, if a function is 'bad' from the standpoint of the rate of decay of its best approximations, then the Fourier partial sums are either poor approximations to f or even fail to converge; in contrast, for 'good' functions they provide good approximations. Comparing the rate of approximation of two functions, we see that, in general, the Fourier sums are better approximants to that for which best approximations decrease more rapidly.

3b) For any function $\psi(n) > 0$ with any rate of decay, there is a function $f \in C_{2\pi}$, which is not a trigonometric polynomial, such that

$$\|f - s_n(f)\|_C = O(\psi(n)).$$

In order to prove this, consider an arbitrary sequence $\varepsilon_n \downarrow 0$, $\varepsilon_n = O(\psi(n))$, and set $a_n = \varepsilon_n - \varepsilon_{n+1}$. Hence, $\sum_{k=n}^{\infty} a_k = \varepsilon_n$. Let $f(x) = \sum_{k=1}^{\infty} a_k \cos kx$. We have

$$\|f - s_n(f)\| \leqslant \sum_{k=n+1}^{\infty} a_k = \varepsilon_{n+1},$$

the result required.

3c) The trigonometric system is dense in $C_{2\pi}$. Hence, for any function $f \in C_{2\pi}$, we have $E_n(f)_C \to 0$ as $n \to \infty$. Therefore,

$$\|s_n(f)\|_C \leqslant \|f\|_C + \big(\|S_n\|_C + 1\big) E_n(f)_C = O(1) + o(\ln n) = o(\ln n) \qquad (n \to \infty).$$

Consequently, for any continuous function the partial Fourier sums grow slower than $\ln n$, even though the supremum of the norms of the partial Fourier sums over the whole class of continuous functions has the order of magnitude $\ln n$.

Examine next the same questions for Fejér sums.

1) We have $\|\sigma_n\|_C = 1$, because the operator is positive.

2) Any function $f \in C_{2\pi}$ can be uniformly approximated by the Fejér sums (see, e.g., [35, Ch. 5, § 5,3], [30, Ch. X, § 5], [27, § 3.1.2]); that is,

$$\|f - \sigma_n(f)\|_C \to 0 \qquad (n \to \infty).$$

3) The Fejér sums, unlike the partial Fourier sums, cannot approximate continuous functions very rapidly. The following result holds.

Theorem 6.1. *For any nonconstant function $f \in C_{2\pi}$ there exists a number $c = c(f) > 0$ such that*

$$\|f - \sigma_n(f)\|_C \geqslant n^{-1} c(f) \quad \text{for all } n;$$

that is, no nonconstant function $C_{2\pi}$ can be approximated by Fejér sums with order better than n^{-1}.

From this standpoint, the Fejér sums are worse than the Fourier sums: they poorly approximate good functions.

Proof. Let $f(x) \not\equiv \text{const}$ and let

$$f(x) \sim \frac{a_0}{2} + \sum_{k=1}^{\infty} A_k(x).$$

Then there exists $k_0 > 0$ such that $A_{k_0}(x) \not\equiv 0$. Hence either $a_{k_0}(f) \neq 0$ or $b_{k_0}(f) \neq 0$. Suppose, for definiteness, $a_{k_0} \neq 0$. We have

$$\|f - \sigma_n(f)\|_C = \max_x |f(x) - \sigma_n(f, x)|$$
$$\geqslant \frac{1}{2\pi} \int_0^{2\pi} |f(x) - \sigma_n(f,x)|\, dx \geqslant \frac{1}{2\pi} \left| \int_0^{2\pi} \{f(x) - \sigma_n(f,x)\} \cos k_0 x\, dx \right| \equiv J.$$

But $\sigma_n(f,x) = \sum_{k=0}^{n} \bigl(1 - \frac{k}{n+1}\bigr) A_k$, and hence, for $n \geqslant k_0$,

$$J = \frac{1}{2} \left| \frac{k_0}{n+1} a_{k_0} \right| \geqslant \frac{1}{2(n+1)} |a_{k_0}| \geqslant \frac{c(f)}{n}.$$

For $n < k_0$, the sum σ_n does not contain the term $\cos k_0 x$, and hence $J = \frac{|a_{k_0}(f)|}{2}$. □

Theorem 6.2. *There exists a function $f \in C_{2\pi}$ such that*

$$\|f - \sigma_n(f)\|_C \asymp \frac{1}{n}, \qquad n \to \infty.$$

Proof. Consider $f_1(x) = \sin x$. Hence $f_1(x) - \sigma_n(x, f_1) = \frac{1}{n+1} \sin x$, and so

$$\|f_1 - \sigma_n(f_1)\|_C = \frac{1}{n+1}.$$

So there exist continuous functions for which the order of approximation by Fejér sums is precisely n^{-1}. □

By Theorems 6.1 and 6.2 it follows that no continuous function, except for constant functions, can be approximated by the nth Fejér means with rate exceeding cn^{-1}; this bound cannot be improved.

An approximation method with such a property is called *saturated*. The best possible rate of approximation of continuous functions with a saturated method is called the *saturation order* of this method (in $C_{2\pi}$).

Consequently, the saturation order of Fejér sums is n^{-1} (see, e.g., [7, § 3.6], [24, Ch. 7, § 4]).

At the same time, from 3b) it follows that the rate of approximation of continuous functions by the Fourier sums can be arbitrarily high. Hence, the method of partial Fourier sums is not saturated.

For a saturated method, the class of functions for which the rate of approximation by this method coincides with the saturation order of this method is called the *saturation class* of the method (see [8, Ch. 11, § 2]). A nonsaturated method has empty saturation class. The saturation class of the Fejér method will be described later.

6.2. De la Vallée Poussin sums

The de la Vallée Poussin sums work well both for 'good' and 'bad' functions.

Let $f \in C[0, 2\pi]$, $0 \leqslant m \leqslant n$. The polynomials

$$\sigma_{n,m}(f) = \frac{1}{n-m+1} \sum_{k=m}^{n} s_k(f)$$

are called de la Vallée Poussin sums. In particular, $\sigma_{n,0}(f) = \sigma_n(f)$ are Fejér sums and $\sigma_{n,n}(f) = s_n(f)$ are Fourier sums.

Approximative properties of de la Vallée Poussin sums for $m \asymp n$ and $n - m \asymp n$ (that is, when $an \leqslant m \leqslant An$, where $a > 0$, $A < 1$) are of special interest.

The de la Vallée Poussin sums can be represented as follows:

$$\sigma_{n,m}(f) = \sum_{k=0}^{n} \lambda_k^{(n,m)} A_k(x) \qquad (A_k(x) = A_k(x, f) = a_k(f) \cos kx + b_k(f) \sin kx);$$

here $\lambda_k^{(n,m)} = 1$ for $k \leqslant m$ and $\lambda_k^{(n,m)} = \frac{n-k+1}{n-m+1}$ for $m \leqslant k \leqslant n$.

The de la Vallée Poussin sums and Fejér sums are related as follows (see [30, Ch. VIII]):

$$\sigma_{n,m}(f) = \frac{1}{n-m+1} \sum_{k=m}^{n} s_k(f)$$

$$= \frac{1}{n-m+1} \left\{ \sum_{k=0}^{n} s_k(f) - \sum_{k=0}^{m-1} s_k(f) \right\} = \frac{n+1}{n-m+1} \sigma_n(f) - \frac{m}{n-m+1} \sigma_{m-1}(f).$$

6.3. Properties of de la Vallée Poussin sums

6.3.1. Norm estimate

Since $\|\sigma_n\|_C = 1$, we have

$$\|\sigma_{n,m}\|_C \leqslant \frac{n+1}{n-m+1} + \frac{m}{n-m+1} = \frac{n+m+1}{n-m+1}.$$

The range

$$an \leqslant m \leqslant An, \qquad 0 < a < A < 1, \tag{6.2}$$

is called the basic range of m. For $m \leqslant An$, $A < 1$, and in particular, for m from the basic range (6.2),

$$\|\sigma_{n,m}\|_C \leqslant \frac{n+m+1}{n-m+1} \leqslant \frac{n+An+1}{n-An+1} < \frac{1+A}{1-A}.$$

Consequently, in this case the de la Vallée Poussin sums do not degenerate into Fourier sums; in other words, if $m \leqslant An$, $A < 1$, then the de la Vallée Poussin sums are uniformly bounded in norm.

Remark. Actually, more can be said:

$$\|\sigma_{n,m}\|_C = \frac{4}{\pi^2}\ln\frac{n+m+1}{n-m+1} + O(1);$$

this is Nikol'skii's theorem (the proof will be given in Lecture 17).

6.3.2. Regularity

An approximation method $\sigma(f,n)$ is called *regular* if, for any continuous function f,

$$\|f - \sigma(f,n)\|_C \to 0 \qquad (n\to\infty).$$

The question arises: When does this hold for de la Vallée Poussin sums? A necessary condition for regularity is that the norms be uniformly bounded and that convergence be secured on a dense subset. We have

$$\sigma_{n,m}(\cos kx) = \lambda_k^{(n,m)}\cos kx \to \cos kx \qquad (n\to\infty),$$

because $\lambda_k^{(n,m)} \to 1$ as $n\to\infty$, when k and m are fixed. The same can be said for sines. As a result, the de la Vallée Poussin sums are uniformly convergent on sines and cosines, and the norms are uniformly bounded, provided there is no degeneration into Fourier sums—that is, when $m \leqslant An$, $A<1$.

Hence, for $m \leqslant An$, $A<1$, the de la Vallée Poussin approximation method is regular.

Remark. This condition is also necessary. Indeed, if the de la Vallée Poussin method σ_{n,m_n} is regular for some m_n, $m_n<n$, then $m_n<An$ starting with some n, where $A<1$.

6.3.3. Invariance property

The de la Vallée Poussin approximation method $\sigma_{n,m}$ leaves fixed those function for which a portion of spectrum, starting with some number exceeding m, is zero — these are the functions such that $a_k=b_k=0$ for all $k>m$. Hence, the subspace $\mathcal{T}_m$ of trigonometric polynomials of degree at most m is an invariant subspace for the method $\sigma_{n,m}$; that is, $\sigma_{n,m}(t)=t$ for $t\in\mathcal{T}_m$.

6.3.4. Lebesgue's inequality for de la Vallée Poussin sums in $C_{2\pi}$

If, for a general linear operator $P: f\to p(f)$, we have $p(t)=t$ for all $t\in\mathcal{T}_m$, then, by Lebesgue's inequality,

$$\|f-p(f)\| \leqslant (\|P\|+1)E_m(f).$$

For de la Vallée Poussin sums, the Lebesgue inequality is as follows

$$\|f-\sigma_{n,m}(f)\| \leqslant (\|\sigma_{n,m}\|+1)E_m(f) \leqslant \left(\frac{n+m+1}{n-m+1}+1\right)E_m(f) = \frac{2(n+1)}{n-m+1}E_m(f).$$

Within the basic range (6.2) of m, we have $n-m+1 \geqslant (1-A)(n+1)$, and so

$$\|f-\sigma_{n,m}(f)\| \leqslant \frac{2(n+1)}{n-m+1}E_m(f) \leqslant \frac{2(n+1)}{(1-A)(n+1)}E_m(f) \leqslant \frac{2}{1-A}E_{[an]}(f).$$

For a function f such that $E_{[an]}(f) \leqslant RE_n(f)$, where R is independent of n, we have

$$\|f-\sigma_{n,m}(f)\| \leqslant \frac{2}{1-A}E_{[an]}(f) \leqslant \frac{2R}{1-A}E_n(f),$$

provided that there is no degeneration into Fourier and Fejér sums.

Hence, under the above conditions on m and f, we have precisely the order of best approximation.

Problem. Prove that the de la Vallée Poussin method is not saturated, provided that $an \leqslant m \leqslant An$ and $0 < a < A < 1$.

For the Fejér method there are no invariant subspaces except for the subspace of constant functions, and so in this cases Lebesgue's inequality can only be written with $E_0(f)$:

$$\|f-\sigma_n(f)\| \leqslant 2E_0(f).$$

However, the following result is valid.

Theorem 6.3 (S. B. Stechkin). *The following estimate holds:*

$$\|f-\sigma_n(f)\| \leqslant \frac{c}{n+1}\sum_{k=0}^{n}E_k(f).$$

From this theorem, which we state without proof, it follows that if f is a continuous and nonconstant function and if $\sum_{k=0}^{\infty}E_k(f) < \infty$, then

$$\|f-\sigma_n(f)\| \asymp n^{-1}, \qquad n \to \infty,$$

since in this case $\|f-\sigma_n(f)\| \geqslant c(f)\,n^{-1}$. This is so, for example, if $E_n(f) = O(n^{-\gamma})$, $\gamma > 1$.

We thus have a sufficient condition ensuring that a function lies in the saturation class for the Fejér method (we recall, that in this case the saturation class is composed of the functions that can be approximated by Fejér sums with the rate n^{-1}).

A necessary and sufficient condition that a function lie in the saturation class for the Fejér method can be expressed in terms of conjugate functions.

We recall that the class $\operatorname{Lip}\alpha$ of Hölder-continuous functions consists of functions f which are such that, for any x' and x'',

$$|f(x')-f(x'')| \leqslant M|x'-x''|^{\alpha},$$

where $0 < \alpha \leqslant 1$ and M is some constant.

Theorem 6.4 (G. Alexits, M. Zamanski, see [7, Ch. 3, § 3.6]). *A necessary and sufficient condition ensuring that a function f lies in the saturation class for the Fejér method is that $\widetilde{f} \in \operatorname{Lip}1$, where $\widetilde{f}$ is the conjugate function to f.*

We state this result without proof. The concept of the conjugate function $\widetilde{f}$ is easier to introduce in the case when $f(x)$ is a boundary value of function $u(z)$ that is harmonic inside the unit disc and which is continuous in the closed disc, $z = re^{ix}$, $0 \leqslant r < 1$. In this case the conjugate function $\widetilde{f}(x)$ is the boundary value of the function $v(z)$, $z = re^{ix}$, $r \to 1$, which is defined by the condition $u(z) + iv(z) = f(z)$; here $f(z)$ is analytic inside the unit disc.

In the general case, the conjugate function to an integrable 2π-periodic function f is defined as follows:

$$\widetilde{f}(x) = \lim_{\varepsilon \to +0} \frac{1}{2\pi} \int_{\varepsilon}^{\pi} \big(f(x-t) - f(x+t)\big) \cot \frac{t}{2}\, dt.$$

The conjugate function $\widetilde{f}$ to a summable function f may fail to be summable. If (5.1) is the Fourier series for f and if $\widehat{f} \in L_{2\pi}$, then the Fourier series for $\widetilde{f}$ is as follows:

$$\widetilde{f}(x) \sim \sum_{k=1}^{\infty} (-b_k \cos kx + a_k \sin kx).$$

Lecture 7

Linear summation methods of Fourier series in $C_{2\pi}$

7.1. Definition of linear summation methods

Methods of approximation by Fourier sums s_n, Fejér sums σ_n, and de la Vallée Poussin sums $\sigma_{n,m}$ are particular cases of linear summation methods of Fourier series.

Let $f \in L(-\pi, \pi)$. We extend f to a function on the real line which is periodic with period 2π, assuming $f(x + 2\pi) = f(x)$ for all x; the resulting function $f \in L_{2\pi}$. With any such a function we associate its Fourier series $f(x) \sim \sum_{m=0}^{\infty} A_m(x)$.

Definition. Suppose we are given a series of elements A_m of a Banach space

$$\sum_{m=0}^{\infty} A_m \tag{7.1}$$

(no assumptions regarding its convergence is made). Also suppose we are given an infinite numerical matrix

$$T = (\lambda_m^{(n)}) \qquad (m = 0, 1, \dots ; \quad n = 0, 1, \dots).$$

By using this matrix and series (7.1), we construct the sequence of series

$$\tau_n = \sum_{m=0}^{\infty} \lambda_m^{(n)} A_m.$$

If all these series are convergent, then we say that for series (7.1) a (linear) summation method T is defined, which transforms series (7.1) in the sequence $\{\tau_n\}$:

$$A \overset{(T)}{\longmapsto} \{\tau_n\}.$$

For the Fourier sums s_n, the matrix T is as follows:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ 1 & 1 & 0 & 0 & \cdots \\ 1 & 1 & 1 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix};$$

and for the Fejér sums σ_n, the matrix T has the form

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ 1 & 1/2 & 0 & 0 & \cdots \\ 1 & 2/3 & 1/3 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}.$$

It is also easy to write the matrix T for the de la Vallée Poussin sums $\sigma_{n,m}$. In all three cases, we have $\lambda_m^{(n)} = 0$ starting from some m, and so τ_n are finite sums. Matrices of this type (and the corresponding summation methods) are called *row-finite*. In these cases, convergence conditions for the series τ_n are satisfied.

We shall study row-finite summation methods of Fourier series

$$f \sim \sum_{m=0}^{\infty} A_m(x) = \frac{a_0}{2} + \sum_{m=1}^{\infty} (a_m \cos mx + b_m \sin mx)$$

with

$$T = (\lambda_m^{(n)}) \qquad (n = 0, 1 \ldots; \quad m = 0, 1, \ldots, M(n)).$$

In this case, for each function f we associate the sequence of trigonometric polynomials

$$\tau_n(f, x) = \sum_{m=0}^{M(n)} \lambda_m^{(n)} A_m(x).$$

Approximation theory is concerned, *inter alia*, with the deviation in approximating a function f by polynomials $\tau_n(f, x)$; in particular, as $n \to \infty$.

Consider the following analogue of the Dirichlet integral for an arbitrary summation method

$$\tau_n(f, x) = \frac{1}{\pi} \int_{-\pi}^{\pi} K_n(t) f(x + t)\, dt;$$

here

$$K_n(t) = \frac{\lambda_0^{(n)}}{2} + \sum_{m=1}^{M(n)} \lambda_m^{(n)} \cos mt$$

is the corresponding kernel of the summation method. This sequence of kernels determines the summation method. In $C_{2\pi}$ we consider the linear operator

$$\mathfrak{T}_n : \ f(x) \mapsto \tau_n(f, x),$$

Regarding $\mathfrak{T}_n$ as an operator from $C_{2\pi}$ to $C_{2\pi}$, we claim that

$$\|\mathfrak{T}_n\|_C = \sup_{\|f\|_C \leqslant 1} \max_x |\tau_n(f, x)| = \frac{1}{\pi} \int_{-\pi}^{\pi} |K_n(t)|\, dt.$$

Indeed, the inequality

$$\|\mathfrak{T}_n\|_C = \sup_{\|f\|_C \leqslant 1} \max_x \left| \frac{1}{\pi} \int_{-\pi}^{\pi} K_n(t) f(x + t)\, dt \right| \leqslant \frac{1}{\pi} \int_{-\pi}^{\pi} |K_n(t)|\, dt$$

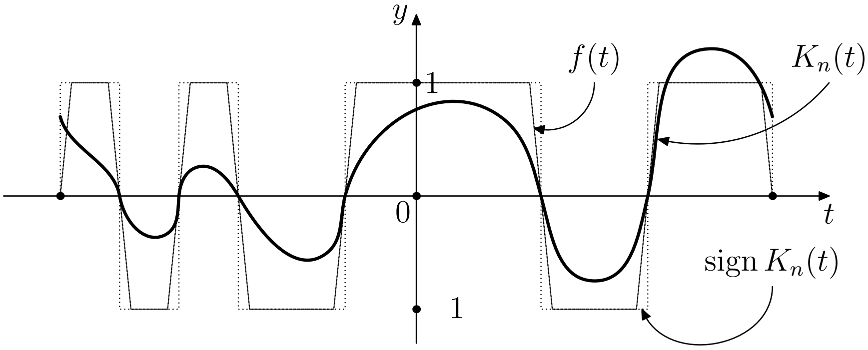

Fig. 7.1

is straightforward. To prove the equality, it suffices to take for f continuous functions which are $L_{2\pi}$-close to $\operatorname{sign} K_n(t)$ (it presents no difficulties to construct such an f, since K_n is a trigonometric polynomial; see Fig. 7.1).

The question arises: When a linear summation method of Fourier series is regular? In other words, under what conditions the following holds

$$\forall\ f \in C \qquad \|\tau_n(f, x) - f(x)\|_C \to 0 \qquad (n \to \infty)?$$

If this is so, the summation method is called *regular* or *Fourier-regular*.

The following result is a criterion for a linear summation method of Fourier series to be regular.

Theorem 7.1. *Let T be a row-finite matrix. A necessary and sufficient condition ensuring that a linear summation method of Fourier series generated by the matrix T is regular, is that*

1) *the estimate*

$$\|\mathfrak{T}_n\|_C \leqslant M$$

holds for some M and all n;

2) $\mathfrak{T}_n(\cos kx) \xrightarrow{C} \cos kx$, $\mathfrak{T}_n(\sin kx) \xrightarrow{C} \sin kx$ *as $n \to \infty$ uniformly in x for any k.*

Indeed, this is a criterion for convergence of linear operators: the norms are uniformly bounded and convergence is secured on a dense subset (in this case, on the set of all trigonometric polynomials).

Condition 2) can be rewritten as follows:

$$\mathfrak{T}_n(\cos mx) = \lambda_m^{(n)} \cos mx \to \cos mx \qquad (n \to \infty),$$
$$\mathfrak{T}_n(\sin mx) = \lambda_m^{(n)} \sin mx \to \sin mx \qquad (n \to \infty);$$

consequently, condition 2) is satisfied if and only if $\lambda_m^{(n)} \to 1$ as $n \to \infty$ for any fixed m. Condition 2) is always an easily verifiable condition.

As regards Condition 1), we have

$$\|\mathfrak{T}_n\|_C = \frac{1}{\pi}\|K_n\|_L = \frac{1}{\pi}\int_{-\pi}^{\pi} |K_n(t)|\, dt,$$

and so, we need to be able to estimate $\|K_n\|_L$.

Assume that the coefficients $\lambda_m^{(n)}$ are generated by a function φ; that is, $\lambda_m^{(n)} = \varphi(m/n)$ (see Fig. 7.2).

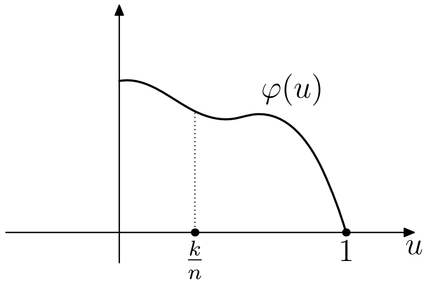

Fig. 7.2

We have, for $M(n) = n$,

$$K_n(t) = \frac{\varphi(0)}{2} + \sum_{m=1}^{n} \varphi\Big(\frac{m}{n}\Big) \cos mt = n\Big\{\frac{\varphi(0)}{2} \cdot \frac{1}{n} + \sum_{m=1}^{n} \varphi\Big(\frac{m}{n}\Big) \cos\Big(n \cdot \frac{mt}{n}\Big) \cdot \frac{1}{n}\Big\}.$$

The term in brackets is the Riemann integral sum for the integral

$$\int_0^1 \varphi(u) \cos(n \cdot ut)\, du;$$

it can be shown that if φ is sufficiently smooth, then the corresponding quadrature formula converges. Changing variable to $nt = y$, this gives

$$\|\mathfrak{T}_n\|_C \approx \frac{2}{\pi} \int_0^{\pi} \left| \int_0^1 \varphi(u) \cos(u \cdot nt)\; du \right| n\, dt = \frac{2}{\pi} \int_0^{n\pi} \left| \int_0^1 \varphi(u) \cos uy\; du \right|\; dy.$$

If the integral

$$\frac{2}{\pi} \int_0^{\infty} \left| \int_0^1 \varphi(u) \cos uy\; du \right|\; dy$$

converges, then the norms $\|\mathfrak{T}_n\|_C$ are bounded. It can be shown that if the integral diverges, then the method is irregular.

Remark. The formula obtained is a fairly precise approximate formula for $\|\mathfrak{T}_n\|_C$.

The regular methods are by no means the only interesting methods — the method of Fourier sums in not regular. At the lack of regularity, the growth of the norms $\|\mathfrak{T}_n\|_C$ should be investigated.

7.2. Approximation by linear summation methods of Fourier series of classes of functions

Let K be a compact class of functions in $C_{2\pi}$ or a class that can be compactified by suitable normalization (for example, the class of functions with $|f'(x)| \leqslant 1$ is not compact, but it becomes compact with the normalization $f(0) = 0$ and taking closure).

Given a row-finite summation method of Fourier series $\mathfrak{T}_n : f(x) \to \tau_n(x, f)$, consider

$$\sup_{f\in K} \|f - \tau_n(f)\|_C.$$

Let $K = W^r$ be the class of 2π-periodic functions with continuous rth derivative, $|f^{(r)}(x)| \leqslant 1$, and let $\tau_n = s_n$.

We state the following result without proof.

Theorem 7.2 (A. N. Kolmogorov, see e.g. [17, § 27]). *For any r,*

$$\sup_{f\in W^r} \|f - s_n(f)\|_C = n^{-r}\Big\{\frac{4}{\pi^2}\ln n + O(1)\Big\}, \qquad n \to \infty.$$

For a compact class K we may consider the classes $\operatorname{Lip}\alpha$ ($0 < \alpha \leqslant 1$), $H[\omega]$, W^r, or $A(q)$. Here, the class $H[\omega]$ is defined, for a given modulus of continuity ω, as consisting of all functions f such that $\omega(f, \delta) \leqslant M\omega(\delta)$ with some absolute constant M; $A(q)$ is the class of 2π-periodic functions f that are analytic in the strip of width $2q$ parallel to the real axis and which are such that $|f(x \pm iq)| \leqslant 1$.

Estimating the error of approximation of a class K by a linear method τ_n is somewhat simpler if K is a class of integral transforms, which consists of the functions representable by the formula

$$f(x) = \frac{a_0}{2} + \frac{1}{\pi}\int_{-\pi}^{\pi} K(t)\varphi(x+t)\,dt \qquad (\varphi \perp 1, \text{ i.e., } a_0(\varphi) = 0)$$

with some kernel $K(t)$.

The classes W^r, $r > 0$, are good from this point of view because they consist of functions representable as integral transforms. In this case, taking the Favard kernel $\mathfrak{D}_r(t) = \sum_{k=1}^{\infty} k^{-r}\cos\big(kt + \frac{r\pi}{2}\big)$ for $K(t)$, we have

$$f(x) = \frac{a_0}{2} + \frac{1}{\pi}\int_{-\pi}^{\pi} \mathfrak{D}_r(t) f^{(r)}(x+t)\,dt.$$

To represent the whole class W^r (including noninteger r), considering an arbitrary continuous function φ instead of the derivative $f^{(r)}$, we obtain the class

$$W^r = \Big\{f(x) = \frac{a_0}{2} + \frac{1}{\pi}\int_{-\pi}^{\pi} \mathfrak{D}_r(t)\varphi(x+t)\,dt : \ |\varphi(t)| \leqslant 1, \quad \varphi \perp 1\Big\}.$$

A similar representation (with suitable kernel $K(t)$) also holds for functions of the class $A(q)$ and for some other classes.

For such classes, we have

$$f(x) = \frac{a_0}{2} + \frac{1}{\pi}\int_{-\pi}^{\pi} K(t)\varphi(x+t)\,dt,$$

and so

$$\tau_n(f,x) = \frac{a_0\lambda_0^{(n)}}{2} + \frac{1}{\pi}\int_{-\pi}^{\pi} \tau_n(K,t)\varphi(x+t)\,dt.$$

If $\lambda_0^{(n)} = 1$, we have

$$f(x) - \tau_n(f,x) = \frac{1}{\pi}\int_{-\pi}^{\pi} \{K(t) - \tau_n(K,t)\}\varphi(x+t)\,dt.$$

Since $\varphi \perp 1$, i.e., $\frac{1}{\pi}\int_{-\pi}^{\pi}\varphi(x-t)\,dt = 0$, we can subtract from the kernel of this convolution any constant function, giving

$$f(x) - \tau_n(f,x) = \frac{1}{\pi}\int_{-\pi}^{\pi} \{K(t) - \tau_n(K,t) - c\}\varphi(x+t)\,dt.$$

Hence,

$$\begin{aligned}\sup \|f(x) - \tau_n(f,x)\|_C &\leqslant \inf_c \sup_{|\varphi|\leqslant 1} \Big|\frac{1}{\pi}\int_{-\pi}^{\pi} \{K(t) - \tau_n(K,t) - c\}\varphi(x+t)\,dt\Big| \\ &\leqslant \inf_c \Big\{\frac{1}{\pi}\int_{-\pi}^{\pi} |K(t) - \tau_n(K,t) - c|\,dt\Big\} = E_0(K - \tau_n(K))_{L_{2\pi}}.\end{aligned}$$

In fact, the inequality here becomes an equality for a wide class of kernels K.

For the classes $\operatorname{Lip}\alpha$ and $H[\omega]$, the problem is more difficult, because these classes are not representable with the help of integral transforms.

This problem has been examined for a broad class of summation methods τ_n and classes K (not only for the ones representable with the help of integral transforms).

7.3. Interpolation processes

Suppose we are given a matrix of nodes $(x_k^{(n)})$, $k = 0, 1, \dots, n$, $n \in \mathbb{N}$, on the fundamental interval $[a, b]$. For any function $f \in C[a, b]$ and any n it is possible to construct the Lagrange polynomial $p_n(f, x, (x_k^{(n)}))$ interpolating f at the nodes $x_k^{(n)}$. This defines the linear operator $P_n\colon f \mapsto p_n(f, x, (x_k^{(n)}))$. In this case we say that an interpolation process is defined.

Consider the problem: Does there exist a regular interpolation process? In other words, does there exists a matrix of nodes such that

$$\forall\, f \in C \qquad \|f - p_n(f)\|_C \to 0 \qquad (n \to \infty).$$

Theorem 7.3 (G. Faber, see e.g. [47, § 8.1.2]). *For any matrix, the process in question is irregular*:

$$\forall\ (x_k^{(n)}) \qquad \|P_n\|_C \to \infty \qquad (n \to \infty).$$

We shall prove a stronger result.

Theorem 7.4. *For any matrix of modes* $(x_k^{(n)})_{k=0}^n$ *and* $n \in \mathbb{N}$,

$$\|P_n\|_C \geqslant C \ln n,$$

where $P_n\colon f \mapsto p_n(f, x, (x_k^n)_{k=0}^n)$ *and* $C > 0$ *is independent of* n.

Lemma 7.1 (on trigonometric polynomials). *For any* n *points* θ_k, $0 \leqslant \theta_1 < \theta_2 < \ldots < \theta_n \leqslant \pi$, *there exists an even polynomial* $t_{n-1}(\theta) = a_0/2 + \sum_{k=1}^{n-1} a_k \cos k\theta$, *such that*

$$|t_{n-1}(\theta_k)| \leqslant 1, \quad k = 1, \ldots, n, \quad \textit{and} \quad \|t_{n-1}\|_C \geqslant a \ln n,$$

where a *is some constant.*

Proof. We proceed to build such a polynomial. For the Fejér polynomials

$$A_n(\theta) = \frac{\cos\theta}{n-1} + \ldots + \frac{\cos(n-1)\theta}{1},$$
$$B_n(\theta) = \frac{\cos(n+1)\theta}{1} + \ldots + \frac{\cos\theta(2n-1)}{n-1},$$

it was proved (see the proof of Theorem 5.2) that, for any n,

$$\|A_n(\theta) - B_n(\theta)\|_C \leqslant M$$

and $|A_n(0)| \asymp \ln n$ as $n \to \infty$.

Given fixed $\theta_1, \ldots, \theta_n$, consider the Lagrange fundamental polynomials $C_k(\theta)$ of order $n-1$ for the trigonometric interpolation process:

$$C_k(\theta) = \frac{\prod\limits_{i \neq k} (\cos\theta - \cos\theta_i)}{\prod\limits_{i \neq k} (\cos\theta_k - \cos\theta_i)}.$$

Hence $C_k(\theta_i) = \delta_{k,i}$.

We set

$$u(\theta) = A_n(2\theta) - \sum_{k=1}^n \{B_n(\theta_k + \theta) + B_n(\theta_k - \theta)\} C_k(\theta).$$

This is a trigonometric polynomial of order at most $3n$.

It is easily verified that $a_0(u) = \pi^{-1} \int_{-\pi}^{\pi} u(\theta)\, d\theta = 0$. Hence there exists a point α such that $u(\alpha) = 0$. We fix such α, and construct an even trigonometric polynomial of order at most $n-1$,

$$t_{n-1}(\theta) = \{A_n(\theta + \alpha) + A_n(\theta - \alpha)\} - \sum_{k=1}^n \{B_n(\theta_k + \alpha) + B_n(\theta_k - \alpha)\} C_k(\theta).$$

We have

$$t_{n-1}(\theta_k) = \{A_n(\theta_k + \alpha) + A_n(\theta_k - \alpha)\} - \{B_n(\theta_k + \alpha) + B_n(\theta_k - \alpha)\}.$$

Hence, $|t_{n-1}(\theta_k)| \leqslant 2\|A_n - B_n\|_C \leqslant 2M$, and so

$$t_{n-1}(\alpha) = u(\alpha) + A_n(0) = A_n(0) \asymp \ln n \qquad (n \to \infty),$$

that is, for t_{n-1} we have, for some positive a,

$$\|t_{n-1}\|_C \geqslant a \ln n.$$

It remains to divide $t_n(\theta)$ by $2M$, in order to have $|t_{n-1}(\theta_k)| \leqslant 1$. □

Prove of Theorem 7.4. We estimate the norm of the operator P_{n-1}.

Since the theorem is concerned with interpolation by an algebraic polynomial with nodes $\{x_k\}$ on $[a,b]$, we 'transplant' the even trigonometric polynomial $\tau_n(\theta)$ built in the lemma to the inteval $[a,b]$ by changing variable to $x = \frac{a+b}{2} + \frac{b-a}{2}\cos\theta$. Then corresponding to the points $\{x_k\}$ on $[a,b]$ there are points $\{y_k\}$, $-1 \leqslant y_k \leqslant 1$,

$$x_k = \frac{a+b}{2} + \frac{b-a}{2} y_k$$

and points θ_k, $0 \leqslant \theta_k \leqslant \pi$, at which $y_k = \cos\theta_k$.

As a result, we obtain an algebraic polynomial $p^*_{n-1}(x)$ such that

$$|p^*_{n-1}(x_k)| \leqslant 1 \qquad (k = 1, 2, \ldots, n), \qquad a \leqslant x_n < x_{n-1} < \ldots < x_1 \leqslant b,$$
$$\|p^*_{n-1}\|_{C[a,b]} \geqslant a \ln n.$$

Since there exists a continuous function f_n on $[a,b]$ such that $f_n(x_k) = p_{n-1}(x_k)$, $\|f_n\|_{C[a,b]} \leqslant 1$, we have

$$\|P_{n-1}\|_{C\to C} = \sup_{\|f\|_C \leqslant 1} \|p_{n-1}(x, f, \{x_k\}_{k=0}^n)\|_C \geqslant \|p^*_{n-1}\|_C.$$

□

Remark. A Gaussian quadrature process converges for any $f \in C[a,b]$:

$$\int_a^b f(x)\,dx - \sum_{k=1}^n A_k f(x_k) \to 0 \qquad (n \to \infty).$$

However, if at the same nodes we construct the interpolation process, then it will not converge for some function f. Nevertheless,

$$\int_a^b p_{n-1}(f, x)\,dx \to \int_a^b f(x)\,dx \qquad (n \to \infty).$$

There are no convergent interpolation processes, but there are convergent quadrature processes (say, a Gaussian process).

Lecture 8

Best approximation in normed linear spaces

8.1. Preliminaries from the theory of normed linear spaces

A *normed linear space* $X = (L, \|\cdot\|)$ is a linear space L over $\mathbb{R}$ or $\mathbb{C}$ equipped with a real-valued function (the *norm*) $\|\cdot\|: L \to [0, \infty)$ satisfying the following conditions (the axioms of a normed linear space):

1) $\|\lambda x\| = |\lambda| \cdot \|x\|$, $\lambda \in \mathbb{R}$, $x \in L$;

2) $\|x\| = 0 \Rightarrow x = \theta$ ($\theta = \theta_X \equiv 0$ is the zero of the space X);

3) $\|x + y\| \leqslant \|x\| + \|y\|, \quad x, y \in L$.

If the function $\|\cdot\|$ satisfies only the axioms 1) and 3) and there may exist a nonzero element x for which $\|x\| = 0$ then $\|\cdot\|$ is called a *seminorm*.

Given an arbitrary linear set (system) L, we recall the following linear (algebraic) concepts.

8.1.1. Linear dependence and independence

The concept of linear independence is important for linear n-dimensional spaces over $\mathbb{R}$ or $\mathbb{C}$. A finite set of elements $x_1, \ldots, x_n$ of L is called *linearly dependent* if there exist numbers $\{c_k\}_{k=1}^n$, not all zero, such that

$$\sum_{k=1}^{n} c_k x_k = 0;$$

otherwise the system is *linearly independent*.

The maximal number of linearly independent elements of a space L, if finite, is called its *dimension*. A space L has infinite dimension if, for any natural number n, there exists a linearly independent system $M \subset L$ of cardinality n ($M^{\#} = n$).

A finite set M is said to be *linearly independent* if, for any natural number n not exceeding the cardinality of M and any family $x_1, \ldots, x_n$ of distinct elements of M, these n elements are linearly independent. A set is said to be *linearly independent*, provided that any its finite nonempty subset is linearly independent.

8.1.2. Algebraic basis

Let $L_1 \subset L$ be a linear subsystem of L. A subset M of L_1 is called an *algebraic basis* (or *Hamel basis*) for L_1 if, firstly, it consists of linearly independent elements, and secondly, if, for any $x \in L_1$, $x \neq 0$, there exists a number n and distinct elements $x_1, \dots, x_n \in M$ such that $x = \sum_{k=1}^{n} c_k x_k$, where not all c_k are zero. From this definition it follows that if, in addition, $x = \sum_{k=1}^{m} d_k y_k$ for some family of elements $y_1, \dots, y_m \in M$, then the set of those $\{x_k\}$ and $\{y_k\}$ with nonzero c_k and d_k can be obtained from each other by permutation, the coefficients of equal elements being equal.

Any linear system L always has an algebraic basis.

Indeed, let

$$x_1, \dots, x_n, \dots, x_{\omega_j}, \dots$$

be a well ordering of L. An element x is said to be expressible in terms of a subset $A \subset L$ if x is a linear combination of a finite subset of A. We remove x_1 and all the elements expressible in terms of x_1; if the ensuing set is nonempty, we denote its first element by $\widetilde{x}_2$ and remove from it all the elements that are linearly expressible in terms of x_1 and $\widetilde{x}_2$. Continuing this reduction process inductively (in general case, by transfinite induction, if L is not a finite or countable set), we obtain the system $x_1, \widetilde{x}_2, \widetilde{x}_3, \dots$ (finite, countable or transfinite), which is clearly a basis for the linear system L.

This being so, any linear set has an algebraic basis, and any element of L is uniquely expressible (up to zero coefficients) in terms of a finite number of basis elements.

8.1.3. Bases in normed linear spaces

Now let $X = (L, \|\cdot\|)$ be a normed linear space. We already know that any linear space always has an algebraic basis. For normed linear spaces, however, we shall need another type of a basis.

A subset M of a normed linear space X is called a basis for X if, for any element x of X, there is a unique system of numbers $\{c_k\}$ such that

$$x = \sum_{k=1}^{\infty} c_k x_k, \qquad x_k \in M;$$

the equality means that $\lim_{n\to\infty} \left\| x - \sum_{k=1}^{n} c_k x_k \right\| = 0$.

Here a basis need not be countable; nevertheless, for any x there is at most countable family of elements $x_k \in M$ such that $x = \sum_{k=1}^{\infty} c_k x_k$.

It is known that any basis for an infinite-dimensional separable normed linear space (a space having a countable dense subset) is always countable. In this case, $x = \sum_{k=1}^{\infty} c_k x_k$; i.e., any x is representable by the series involving all elements of the basis, with some c_k possibly zero. A basis $x_1, x_2, \dots$ is said to be unconditional if the sum of the series (which is the limit of partial sums $\sum_{k=1}^{n} c_k x_k$) is independent of any rearrangement of terms. If the last property is not assumed to hold, a basis is called a *Schauder basis*.

Does every separable Banach space have a basis?[1]

[1] This problem was solved in the negative by Per Enflo [11] in 1972; see also [10, Ch. 5].

Ciesielski proved that any classical separable Banach space (for example, C, L^p ($p > 1$), $C^{(r)}$) admits a basis.

8.1.4. Convexity

A subset M of L is called *convex* if it contains the closed interval $[x, y]$ that joins any two point x, y of M. The closed interval between a pair of points x, y is, by definition, $tx + (1-t)y, \quad t \in [0, 1]$. A set M is *nonconvex* whenever there are two points of M such that the interval between them is not contained in M.

There is a marked difference between the classes of convex sets in finite- and infinite-dimensional spaces. For example, every infinite-dimensional Banach space is the union of two disjoint dense convex subsets.

8.1.5. Convex hull

Let L be a linear space and let $M \subset L$. Consider all possible subsets V of L that contain M. A *convex hull* (written $\operatorname{conv} M$) of M is defined as follows

$$\operatorname{conv} M = \bigcap_{V \supset M} V.$$

Clearly, $M \subset \operatorname{conv} M$. The convex hull always exists, because $V = L \supset M$.

Exercise. Prove that a set M is convex if and only M coincides with its convex hull.

Let us visualize the general form of the convex hull of a set M.

Given any finite subset $M_n = \{x_1, \ldots, x_n\}$ of M, consider its convex hull $\operatorname{conv} M_n$. This is a simplex of dimension $n-1$, provided that $x_1, \ldots, x_n$ are linearly independent.

Let us prove that

$$\bigcup_{M_n \subset M} \operatorname{conv} M_n = \operatorname{conv} M.$$

Indeed, suppose that $x, y \in \bigcup_{M_n \subset M} \operatorname{conv} M_n$. Then x lies in some simplex $\operatorname{conv} M_n$, and y belongs to some simplex $\operatorname{conv} M_m$. Hence both x and y lie in the simplex $\operatorname{conv} M'$, where $M' = M_n \bigcup M_m$. We have $[x, y] \subset \operatorname{conv} M' \subset \bigcup_{M_n \subset M} \operatorname{conv} M_n$, and hence $\bigcup \operatorname{conv} M_n$ is convex. Clearly, this set is contained in any convex set V that contains M, and is the intersection of such sets.

Corollary. *The smallest convex set containing M is given by*

$$\bigcup_{M_n \subset M} \operatorname{conv} M_n.$$

Remark. If $x_1, \ldots, x_n \subset M$, then $\operatorname{conv} M_n$ coincides with the set

$$\Big\{x = \sum_{k=1}^{n} c_k x_k : \quad c_k \geqslant 0, \quad \sum_{k=1}^{n} c_k = 1\Big\}.$$

8.2. Characteristics of normed linear spaces

8.2.1. Separability

An important massiveness features of a space is the smallest cardinality of a dense subset of the space. A space containing a countable dense subset is called *separable.* A nonempty space is called *nonseparable* if it fails to contain a countable dense subset. Separability is the first feature to be tested for a space.

8.2.2. Completeness

A space is *complete* if every Cauchy sequence (fundamental sequence) converges to an element of this space. A complete normed linear space is called a *Banach space* (or a B-space).

8.2.3. Reflexivity

Given a Banach space X, by X^* we denote its dual space, that the space of all continuous linear functionals on X; the second dual of X is denoted by X^{**}.

The canonical embedding of X into X^{**} is defined, for any $x \in X$, by

$$F_x(f) = f(x), \qquad f \in X^*,$$

so that

$$\forall\, x \in X \qquad x \mapsto F_x \in X^{**}.$$

A space is *reflexive* if the canonical embedding of X into X^{**} is onto: $X \equiv X^{**}$; in this case any functional $F \in X^{**}$ coincides with some functional F_x on X^*.

If a space is reflexive, then X and X^{**} are structured in the same way (they are isometrically isomorphic). The converse is not true: there is a separable Banach space which is linearly isometric to its second dual but is not reflexive (see, for example, [28, § 1.11]).

In a reflexive space bounded sequence contains a weakly convergent subsequence.

8.2.4. Structure of the unit ball

Strict convexity. A space is called *strictly convex* (or *rotund*) if its unit ball is strictly convex, that is, the boundary of the unit ball contains no open line segment.

Exercise. A (Euclidean) circle is a strictly convex set, while a square is not. Clearly, if a space is not strictly convex, then there exists a hyperplane that touches the unit ball at more than one point.

Extreme points. Let $M \subset X$, $x \in M$.

A point $x \in M$ is called *nonextreme* point of M, if $x \in (a, b)$ for some points $a, b \in M$. Otherwise x is called an *extreme* point.

Let O_1 be the unit ball, and S_1 be the unit sphere, its boundary.

The following result is valid (Straszewicz's theorem, see e.g. [49, Theorem 2.6.21]).

Theorem 8.1. *The unit ball O_1 of a finite-dimensional Banach space is the closed convex hull of its extreme points.*

We state this fact without proof. However, Straszewicz's theorem may fail to hold in the general case: there are Banach spaces with no extreme points on the unit ball.

Exercise. Prove that the unit ball O_1 of $L[0,1]$ has no extreme points. Prove that the unit ball O_1 of $C[0,1]$ has exactly two extreme points.

Smoothness. A space is *smooth* if at each point of the unit sphere S_1 there is a unique supporting functional (i.e., a unique supporting hyperplane of the unit ball). The definition of a supporting hyperplane will be given later. Otherwise, the space is nonsmooth.

8.2.5. Compact sets

A compact set is a set in which every sequence has a subsequence which is convergent to a point of the set. Compact sets are always closed.

8.3. Fundamental spaces

The space $C = C(Q, X)$ is the space of continuous functions from a compact set Q into a Banach spaces X. Classical cases are $X = \mathbb{R}$ or $X = \mathbb{C}$.

1. The space $C(Q, X)$ is complete.

2. For some Q and X the space $C(Q, X)$ is separable, while for some is not. If Q is a closed interval and $X = \mathbb{R}$, then $C(Q, X)$ is separable.

3. If Q is infinite, then C is nonreflexive; in particular, $C[0,1]$ is nonreflexive.

4. For any Q and X the space $C(Q, X)$ is neither strictly convex nor smooth.

The space L^p, $1 \leqslant p < \infty$.

Suppose we are given a space Q equipped with measure μ (a countably additive nonnegative function of measurable subsets of Q). Let f be a function from Q to $\mathbb{R}$ or $\mathbb{C}$ with finite integral $\int_Q |f(x)|\, d\mu$, let L_μ be the class of all such functions, and let $\int_Q |f(x)|\, d\mu = \wr f \wr$ be the seminorm. In order to obtain a Banach space, we quotient out the space L_μ,

$$L_\mu / \{f : \wr f \wr = 0\},$$

thereby identifying the functions that differ only on nullsets. This gives us a complete space, which we henceforth denote by L_μ.

Whether L_μ is separable or not depends both on Q and on μ: if μ is the Lebesgue measure on $Q \subset \mathbb{R}^n$, then the space $L_\mu = L^1_\mu(Q) = L(Q)$ is separable.

The space L^p_μ is defined similarly; here one should take the seminorm

$$\wr f \wr = \left(\int_Q |f(x)|^p \, d\mu \right)^{1/p}.$$

We shall use the same symbol to denote both a function f and the corresponding equivalence class, and so $\wr f \wr = \|f\|_{L^p_\mu}$ is a norm.

For $p = 2$ the space L^2_μ is a Hilbert space with the inner product defined by $(f, g) = \int_Q fg \, d\mu$ over the field of real numbers, and by $(f, g) = \int_Q f \, \overline{g} \, d\mu$ over the field of complex numbers. The space L^2_μ is complete with respect to the norm $\|f\|_{L^2} = \sqrt{(f, f)}$.

We recall that in the spaces L^p, L^q, for $\frac{1}{p} + \frac{1}{q} = 1$, $1 < p, \ q < \infty$, the Hölder inequality

$$\left| \int_a^b f(x)g(x) \, dx \right| \leqslant \|f\|_{L^p[a,b]} \|g\|_{L^q[a,b]}$$

for $f \in L^p[a, b]$, $g \in L^q[a, b]$ is, for real-valued f and g, an equality if and only if $f(x)g(x) \geqslant 0$ a.e. and if $|f(x)|^p$ is a.e. proportional to $|g(x)|^q$.

Exercise. Examine this problem for functions from $L^p_\mu(Q)$ and $L^q_\mu(Q)$.

Lecture 9

General linear approximation problems

9.1. Convexity of L^p-spaces

We continue to consider L^p-spaces, $1 \leqslant p < \infty$. For $p = 2$, the space L^2 is a Hilbert space. Throughout this lecture we assume that μ is the Lebesgue measure.

We state the following result without proof.

Theorem 9.1. *A necessary and sufficient condition that a complete normed space H be a Hilbert space is that all its finite-dimensional subspaces be Euclidean* (*including the original space, if it is finite-dimensional*).

Corollary. *In a Hilbert space the planar Euclidean geometry holds.*

Example 9.1. The well-known property of a parallelogram is that the sums of the squares of the diagonals is equal to twice the sum of the squares of the sides; i.e.,

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

(the parallelogram law). In courses of functional analysis it is proved that this property characterizes the Hilbert spaces.

For $p \neq 2$, $1 < p < \infty$, the space L^p is not a Hilbert space (of course, if its dimension exceeds 1), but is always strictly convex and reflexive.

The norm of any Banach spaces satisfies the triangle inequality

$$\|x + y\| \leqslant \|x\| + \|y\|.$$

In L^p, this transformes into

$$\left\{\int_Q |x + y|^p \, dt\right\}^{1/p} \leqslant \left\{\int_Q |x|^p \, dt\right\}^{1/p} + \left\{\int_Q |y|^p \, dt\right\}^{1/p},$$

which is the conventional Minkowski's inequality. The inequality here becomes equality (in the real case) if and only if x and y are positively proportional; i.e., when $\alpha x = \beta y$

for some $\alpha, \beta \geqslant 0$, $\alpha^2 + \beta^2 > 0$ (this is equivalent to saying that x and y do not lie on one ray emanating from the origin).

Spaces with this property are called *strictly normed*; this property is equivalent to the strict convexity.

So, the space L^p is strictly convex for any $1 < p < \infty$. The space L^1 is not strictly convex. To substantiate this claim, we need to find two elements x and y of $L^1[0,1]$ with satisfy Minkowski's inequality with equality and such that x and y are not positively proportional. It suffices to consider

$$x(t) = \begin{cases} 1, & t \in [0, 1/2) \\ 0, & t \in [1/2, 1] \end{cases}, \qquad y(t) = \begin{cases} 0, & t \in [0, 1/2) \\ 1, & t \in [1/2, 1] \end{cases}.$$

Figure 9.1 gives another pair of such x and y.

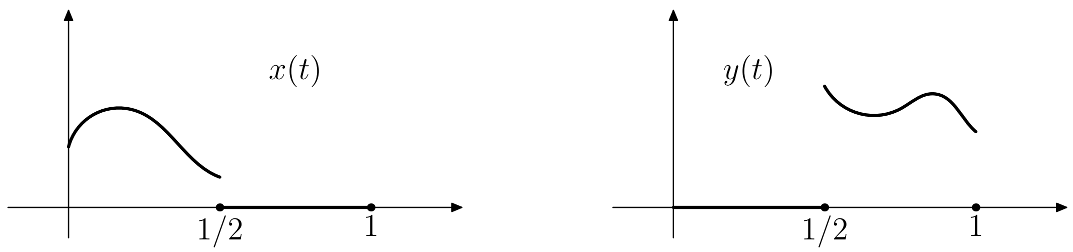

Fig. 9.1

This example can be easily extended to all L_μ-spaces.

9.2. Uniform convexity

Given the unit ball O_1 of a Banach space, we draw a hyperplane at a distance $h < 1$ from the origin, obtaining thereby the slice l of the ball O_1 (see Fig. 9.2). Let $d(l)$ be the diameter of the slice. Consider $\lim\limits_{h \to 1} d(l)$. In a Euclidean space, we have $d(l) \to 0$ as $h \to 1$.

A space is called *uniformly convex* if the diameter of the slice $d(l)$ tends to zero as $h \to 1$ uniformly over all hyperplanes (over all slices). Clearly, a uniformly convex space is always strictly convex.

The following result is valid (see, for example [29], [34]). We state it without proof.

Theorem 9.2 (Milman–Pettis[1].). *Any uniformly convex space is reflexive.*

[1]Translator's note: Stechkin attributes this result to J. A. Clarkson, which does not seem to be the case. It was indeed Clarskon [4] who introduced the notion of uniform convexity and proved that the L^p-spaces ($1 < p < \infty$) are uniformly convex. But it was only Milman [29] and Pettis [34], who proved the result stated. See historical comments on p. 208 of [9].

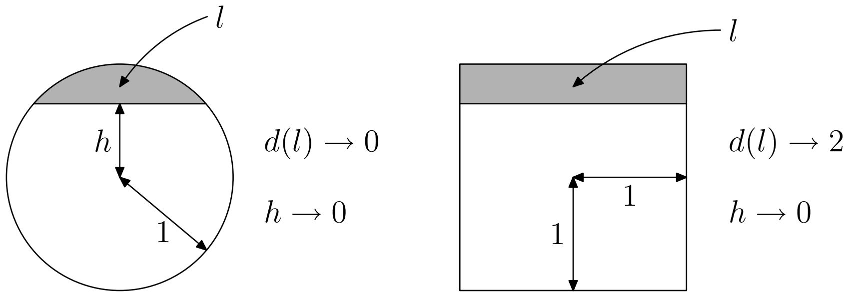

Fig. 9.2

That the space L^p, $1 < p < \infty$, is uniformly convexity is clear. Hence any space L^p, $1 < p < \infty$, is reflexive. The space L^1 is nonreflexive except for degenerate cases when the measure is concentrated at a finite number of points.

9.3. General linear approximation problems

Statement of the problem. Let X be a Banach space and let L be a proper (closed) subspace of X ($L \neq X$).

Consider the problem of best approximation of an element $x \in X$ by elements y of the subspace L:

$$\inf_{y \in L} \|x - y\|_X = E(x, L)_X.$$

9.3.1. Uniqueness problem

For any $x \in X$, we consider the set $Y(x)$ (possibly empty)

$$Y(x) = \{y^* \in L : \|x - y^*\|_X = E(x, L)_X\};$$

this is the set (or a polytope) of best approximations from L to x (the metric projection). We thus have the mapping

$$x \mapsto Y(x) \subset L.$$

Consider the following problem. Under which conditions on X we have $\operatorname{card} Y(x) \leqslant 1$ for any element x of X and any subspace L of X? In other words, in which X any element x of X has at most one element of best approximation from an arbitrary subspace L of X?

Definition. If, for any subspace L and any element $x \in X$, there is at most one element of best approximation from L, we shall say that X has the *uniqueness property* (U).

Theorem 9.3. *A Banach space has the uniqueness property (U) if and only if its strictly convex.*

Proof. Sufficiency. Given a strictly convex space X, assume that the property (U) fails. Then there exist $L \subset X$, $x \in X$ and $y_1, y_2 \in L$ such that

$$\{y_1, y_2\} \subset Y_L(x), \qquad y_1 \neq y_2.$$

We have

$$\inf_{y \in L} \|x - y\| = \|x - y_1\| = \|x - y_2\| = E(x, L)_X > 0.$$

Consider the point $y = (y_1 + y_2)/2$ and find the distance from x to y:

$$\|x - y\| = \left\| \frac{1}{2}(x - y_1) + \frac{1}{2}(x - y_2) \right\|.$$

The point $x - y$ is the midpoint of the interval $[x - y_1, x - y_2]$, whose ends lie on the sphere S of radius $\rho = E(x, L)_X$. The space being strictly convex, the point $x - y$ lies strictly inside the ball O_ρ, and so $\|x - (x - y)\| < \rho$. We thus have

$$E(x, L)_X \leqslant \|x - y\| = \left\| \frac{1}{2}(x - y_1) + \frac{1}{2}(x - y_2) \right\| < \|x - y_1\| = \|x - y_2\| = E(x, L)_X,$$

a contradiction.

Necessity. Given a space X with the property (U), we need to prove that X is strictly convex. Assuming the contrary, we can find a hyperplane L, lying at a distance of 1 from the origin θ_X, and which is such that it touches the unit ball in at least two points s_1, s_2. We translate the hyperplane L by the vector x, so that L transforms into the hyperplane $L_1 \ni \theta_X$, and $s_1 \mapsto y_1$, $s_2 \mapsto y_2$.

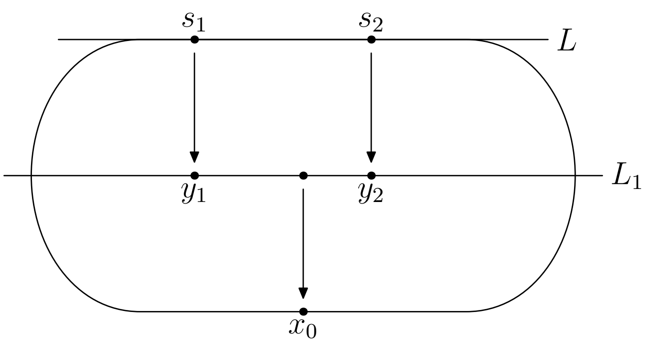

Fig. 9.3

Consider the point x_0 (see Fig. 9.3). It has at least two elements of best approximation y_1 and y_2 from L_1. This gives a contradiction, and the theorem follows. □

Remark. We showed in fact that $[y_1, y_2] \subset Y(x)$ whenever $y_1, y_2 \in Y(x)$; i.e.,

$$E(x, L) = \|x - y_1\| = \|x - y_2\|.$$

Indeed, any element y from $[y_1, y_2]$ can be written in the form $y = ty_1 + (1 - t)y_2$ for some $t \in [0, 1]$. Then

$$E(x, L) \leqslant \|x - y\| = \|t(x - y_1) + (1 - t)(x - y_2)\| \leqslant t\|x - y_1\| + (1 - t)\|x - y_2\| = E(x, L).$$

Corollary. *The set of best approximation is always convex* (*and hence, for a strictly convex space, it is either empty or consists of one element*).

Among the classical spaces only C and L^1 are not strictly convex, hence they fail to have the uniqueness property. The spaces L^p, $1 < p < \infty$, are strictly convex and hence have the uniqueness property.

Generally, all the subspaces of a space X split into two classes: those satisfying and those failing to satisfy the uniqueness property.

9.3.2. Existence problem

If, for any $x \in X$, there is at least one element of best approximation to x from any subspace L of X, then X will be said to have the existence property (E).

A hyperplane L_1 is said to touch the unit sphere $S_1 = \{z \in X : \|z\| = 1\}$ if there is an element $y \in S_1$ such that $\inf_{x \in L_1} \|x - y\| = 0$. We note that touching the sphere does not imply that a point of tangency must exist.

Theorem 9.4 (James; see, e.g., [6, p. 63], [28]). *A Banach space is reflexive if and only if any supporting hyperplane of the unit sphere has* (*at least one*) *point of tangency or, what is the same, any hyperplane* $L_x = \{y : f(y) = f(x)\}$ *has a point of tangency with the sphere* $S_E = \{z : \|z\| = E(x, L)\}$ *for any* $f \in X^* \setminus \{0\}$, *where* L *is the subspace* $\{y : f(y) = 0\}$.

Hence, a space X is reflexive if and only if, for any $f \in X^* \setminus \{0\}$, there is a point $x \in X$ such that $\|x\| = 1$ and $|f(x)| = \|f\|$ (i.e., the functional attains its norm at the element x).

Example 9.2. In $C[0, 1]$, consider the functional $f(x) = \int_0^1 \operatorname{sign} \sin 2\pi t \cdot x(t)\, dt$. The norm of this functional

$$\|f\| = \int_0^1 |\operatorname{sign} \sin 2\pi t|\, dt = 1$$

is not attained on C, because the function $x(t) = \operatorname{sign} \sin 2\pi t$ does not lie in $C[0, 1]$. In this example there is no point of tangency of the hyperplane $f(x) = 1$ with the unit sphere S_1 in $C[0, 1]$ (it is easily seen that if $x(t) \in C[0, 1]$, $\|x\|_C = 1$, then $|f(x)| < 1$).

Theorem 9.5. *A Banach space has the property* (E) *if and only if it is reflexive.*

Proof. 1) Suppose that X is nonreflexive. Then, by James's theorem, there is a hyperplane $\{f(x) = 1\}$, $\|f\| = 1$, which has no points of tangency with the unit sphere. Consider the subspace $L = \{y : f(y) = 0\}$. Then, any element x with $f(x) \neq 0$ has no elements of best approximation in L (if it were such a one, a translation would produce a point of tangency with the hyperplane L_x.

2) Suppose that X is reflexive. We need to prove that it satisfies the property (E). First of all, we observe that if $\{x_n\} \in L$ and $x_n \xrightarrow{w} x$ (x_n converges weakly to x), then $x \in L$ and

$$\|x\| \leqslant d = \varliminf_{n \to \infty} \|x_n\|.$$

Now let L be an arbitrary subspace and x be an arbitrary element of X, $x \notin L$. Let $O_{d+\varepsilon_n} = O_{d+\varepsilon_n}(x)$ be the closed ball, with centre x and radius $d + \varepsilon_n$, where $d = E(x, L)$ and $\varepsilon_n \downarrow 0$. The sets $K_n = O_{d+\varepsilon_n} \bigcap L$ constitute a nested family $\{K_n\}$ of nonempty bounded closed sets. In a reflexive space, such a nested sequence has a nonempty intersection.

Indeed, let $x_n \in K_n$, $\|x - x_n\| \to d$. The sequence $\{x - x_n\}$ is weakly compact, and so it has a subsequence $\{x - x_{n_k}\}$ which converges weakly to some element $x - x_0$, $x_0 \in \bigcap K_n$. Hence, $x_0 \in L$ and $\|x - x_0\| \leqslant d$. Since $d = E(x, L)$, the inequality cannot be strict, so $\|x - x_0\| = d$, therefore x_0 is an element of best approximation. □

Corollary. *Any finite-dimensional subspace is a set of existence.*

Remark. Any boundedly compact set (a set whose intersection with each closed ball $O_d(x)$ is compact) is a set of existence.

Example 9.3. (of a finite parameter family which fails to be boundedly compact). In $C[0, 1]$ consider the set of rational functions of the form $R_1 = \frac{a}{b+ct} \in C[0, 1]$. The family R_1 depends on three parameters a, b, c. This set is noncompact in $C[0, 1]$: the sequence $\{1/(1 + ct)\}$ converges to 0 on $(0, 1]$ as $c \to 0$, and at the point $t = 0$ it is equal to 1.

The space L^p, $1 < p < \infty$, is both reflexive and strictly convex. Hence in L^p, $1 < p < \infty$, any subspace has both the (U)- and (E)-properties.

Such subspaces of (simultaneous) existence and uniqueness are called *Chebyshev subspaces.*

Lecture 10

Criterion for best approximation in L^p. Stability

10.1. Criterion for an element of best approximation in L^p

Let H be a Hilbert space. In courses of functional analysis it is proved that y^* is an element of best approximation to an element x in a subspace M of H if and only if

$$(x - y^*, y) = 0 \qquad \forall\, y \in M.$$

If $H = L^2(Q)$, this condition can be rewritten as follows:

$$\int_Q (x - y^*) y \, dt = 0 \qquad \forall\, y \in M.$$

This formula is the special case of a more general theorem, which gives a necessary and sufficient condition for an element of best approximation in L^p, $p > 1$.

Theorem 10.1 ([47, 2.8.25], [42, Theorem 1.11]). *Let M be a subspace of $L^p(Q)$, $1 < p < \infty$, and let $x \in L^p(Q)$. A necessary and sufficient condition that y^* be an element of best approximation in M to x in $L^p(Q)$ is that*

$$\int_Q |x - y^*|^{p-1} \operatorname{sign}(x - y^*) y \, dt = 0 \qquad \forall\, y \in M. \tag{10.1}$$

Proof. Necessity. Suppose that condition (10.1) is not satisfied. Then there exists a point $y \in M$ such that

$$\int_Q |x - y^*|^{p-1} \operatorname{sign}(x - y^*) y \, dt \neq 0.$$

We claim that in this case y^* is not an element of best approximation. Let

$$\Phi(\alpha) = \|x - y^* - \alpha y\|^p \equiv \|x - y^* - \alpha y\|^p_{L^p(Q)} = \int_Q |x - y^* - \alpha y|^p \, dt.$$

Since $p > 1$, this is a differentiable function of α, and hence, by the theorem on differentiation with respect to a parameter under the integral sign,

$$\Phi'(\alpha) = -p \int_Q |x - y^* - \alpha y|^{p-1} \operatorname{sign}(x - y^* - \alpha y) y \, dt.$$

For $\alpha = 0$, we have $\Phi'(\alpha)|_{\alpha=0} \neq 0$, and hence $\alpha = 0$ is not a minimum. Hence for some α the deviation $\|x - y^* - \alpha y\|$ can be made smaller than $\|x - y^*\|$, and so y^* is not an element of best approximation, a contradiction.

Sufficiency. By (10.1) and using Hölder's inequality, we have, for each $y \in M$,

$$\int_Q |x - y^*|^p \, dt = \int_Q |x - y^*|^{p-1}(x - y^*) \operatorname{sign}(x - y^*) \, dt$$

$$= \int_Q |x - y^*|^{p-1}(x - y) \operatorname{sign}(x - y^*) \, dt \leqslant \int_Q |x - y^*|^{p-1}|x - y| \, dt$$

$$\leqslant \left\{ \int_Q |x - y^*|^p \, dt \right\}^{1/q} \cdot \left\{ \int_Q |x - y|^p \, dt \right\}^{1/p}, \qquad \frac{1}{p} + \frac{1}{q} = 1.$$

It can be assumed that $\int_Q |x - y^*|^p \, dt \neq 0$ (for otherwise y^* is an element of best approximation, as required). We have

$$\left\{ \int_Q |x - y^*|^p \, dt \right\}^{1/p} \leqslant \left\{ \int_Q |x - y|^p \, dt \right\}^{1/p};$$

i.e., for each $y \in M$, $\|x - y^*\| \leqslant \|x - y\|$, and so $\|x - y^*\| = E(x, M)$, as claimed.

Remark. For $p = 1$ condition (10.1) assumes the form

$$\int_Q \operatorname{sign}(x - y^*) y \, dt = 0, \qquad \forall\, y \in M; \tag{10.2}$$

this is a sufficient condition for y^* to be an element of best approximation to x (the proof is the same). Moreover, if *a priori* is known that $x(t) - y^*(t) \neq 0$ a.e. on Q, then condition (10.2) is also necessary (because in this case, for $\alpha = 0$, the derivative $\frac{d}{d\alpha}|x - y^* - \alpha y|\big|_{\alpha=0} = -y \operatorname{sign}(x - y^*)$ exists a.e.).

However, in the general case this is only a sufficient condition. A detailed account is given in [42, Ch. I, § 1.5] (see also [22]).

10.2. Best approximation in L^p

Example 10.1. Let us consider the special case in which $M = \{c\}$ is the one-dimensional subspace of constant functions.

Let $p = 2$ and $Q = [a, b]$. In this case condition (10.1) can be written

$$\int_a^b \{x(t) - c^*\} \, dt = 0;$$

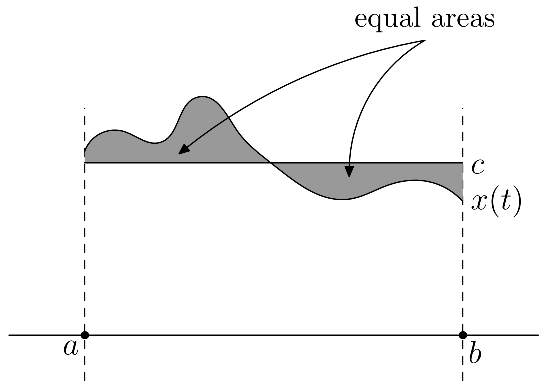

Fig. 10.1

i.e., the areas above and below the line $x(t) = c^*$ are equal.

For $p = 1$, condition (10.1) becomes

$$\int_a^b \operatorname{sign}\{x(t) - c^*\}\, dt = 0 \tag{10.3}$$

or, in the case when $L^1[a, b]$ is a space with measure μ,

$$\mu(E_+) - \mu(E_-) = 0, \tag{10.4}$$

where E_+ (E_-) is the set on which the difference $x(t) - c^*$ is positive (negative, respectively), and μ is a measure.

An example can be easily constructed to illustrate the lack of unicity of best approximation in L^1 (see Fig. 10.2), where $[a, b] = [0, 1]$, $x(t) = 1$ if $0 \leqslant t < 1/2$, $x(t) = 0$ if $1/2 \leqslant t \leqslant 1$.

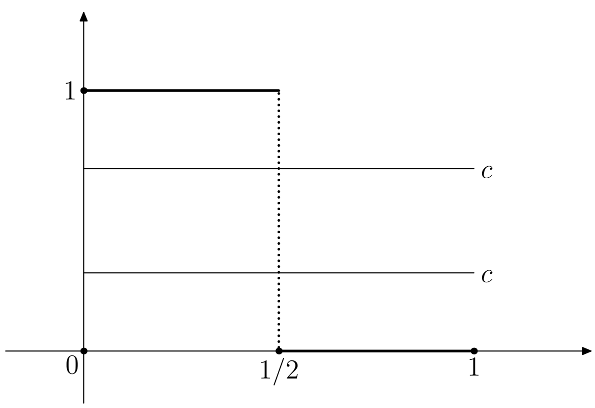

Fig. 10.2 ($c^* \in [0, 1]$)

In this case, any constant function $c^* \in [0, 1]$ is an element of best approximation.

As we have already pointed out, for $p = 1$ condition (10.1) is only sufficient for an element of best approximation. Examples can be constructed to show that condition (10.4) may fail for a constant function of best approximation in L^1 (see Fig. 10.3 for the Lebesgue measure), where $[a, b] = [0, 1]$, $x(t) = 1 - 4t$ if $0 \leqslant t < 1/8$, $x(t) = 1/2$ if $1/8 \leqslant t < 3/4$ and $x(t) = 2(1 - t)$ if $3/4 \leqslant t \leqslant 1$ and $c^* = 1/2$.

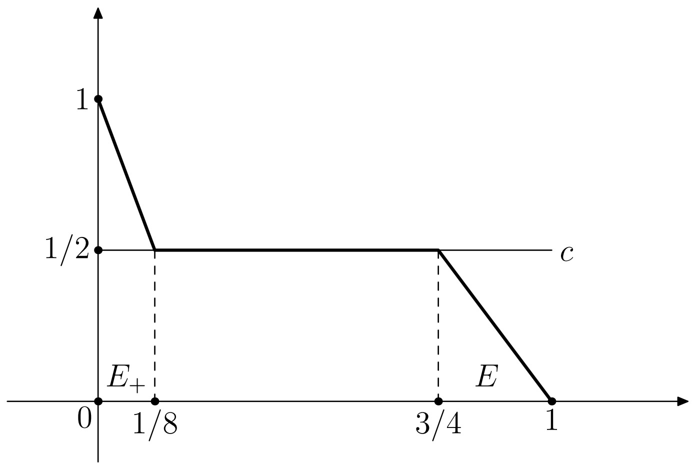

Fig. 10.3 (here $c^* = 1/2$)

If $\mu\{t \in [a, b] : x(t) - y^*(t) = 0\} = 0$, then it is legitimate to differentiate under the integral sign, and so violation of condition (10.4) means that y^* is not an element of best approximation from M to x.

Let X be a Banach space and M be a subspace of X. The best approximation problem consists in finding, for any $x \in X$, a point $y^* \in M$ such that

$$\|x - y^*\|_X \leqslant \|x - y^* - h\| \qquad \forall\, h \in M.$$

Hence the functional

$$\Phi(h) = \|x - y^* - h\|$$

must have a minimum at $h = \theta$. If $\Phi(h)$ is differentiable, then a necessary condition for a minimum is that the differential of $\Phi(h)$ should vanish at $h = \theta$. Or else, suppose that h is fixed. Then the functional

$$F_h(t) = \|x - y^* - th\|, \qquad t \in (-1, 1),$$

must have a minimum at $t = 0$ for each $h \in M$; i.e., $\mathcal{D}\|x - y^* - h\||_{h=\theta} = 0$, and so condition (10.1) simply means that the differential vanishes.

The functional Φ is convex, and so if Φ is differentiable, then the condition that the differential of $\Phi(h) = \|x - y^* - h\|$ be zero on the subspace M at $h = \theta$ is necessary and sufficient for y^* to be an element of best approximation.

10.3. Stability

Consider the problem of continuous dependence of a solution of the best approximation problem on initial conditions.

Let X be a Banach space and let $M \subset X$ be a subset of existence, $y(x)$ be an element of best approximation in M to $x \in X$, and $E(x, M)_X$ be the best approximation of x. Then $E(x, M)_X = \Phi(x)$ is a functional of x.

Since

$$E(x, M) - E(x', M) = \|x - y(x)\| - \|x' - y(x')\| \leqslant \|x - y(x')\| - \|x' - y(x')\| \leqslant \|x - x'\|,$$

it follows that *the best approximation* $E(x, M)$ *continuously* (*and even uniformly continuously*) *depends on* x.

Exercise. Prove the above bound without assuming that M is a set of existence.

Now assume that $y(x)$ and $y(x')$ are unique elements of best approximation in M to x and x', respectively. Suppose that x and x' are close to each other. Does this imply that $y(x)$ and $y(x')$ are close? In general this is not so, because the function $y(x)$ may not continuously depend on x. But there is one important case in which $y(x)$ is always continuous.

Suppose that M is *boundedly compact* (its intersection with any closed ball is compact). A boundedly compact set is always a set of existence: for any x the set of best approximations $Y(x)$ in M to x is nonempty. We are interested in continuity of the mapping

$$x \mapsto Y(x) \subset M, \qquad x \in X.$$

Let Y_ε be the ε-neighbourhood of $Y(x)$ in M; i.e., $Y_\varepsilon = \{y \in M \,:\, \rho(y, Y(x)) < \varepsilon\}$. We fix some element $x \in X$. Let us see how Y_ε is related to Y.

We clearly have $Y = \bigcap_{\varepsilon>0} Y_\varepsilon$. Let $d = E(x, M)$. Consider the set

$$Z(\varepsilon) = Z(\varepsilon, x) = \{z \in M : \ \|x - z\| \leqslant d + \varepsilon\}$$

(see Fig. 10.4). We have the following result for boundedly compact sets M.

Proposition 10.1. *For each* $\varepsilon > 0$ *there exists* $\varepsilon_1 > 0$ *such that* $Z(\varepsilon_1) \subset Y_\varepsilon$.

Proof. To prove this result, we observe that $\{Z(\varepsilon_1)\}_{\varepsilon_1}$ is a nested family of compact subsets such that $\bigcap_{\varepsilon_1>0} Z(\varepsilon_1) = Y(x)$. Then, by the property of compact sets, for any neighbourhood Y_ε of the set $Y(x)$, all $Z(\varepsilon_1)$ lie in Y_ε for all $\varepsilon_1 \leqslant \varepsilon_0$, as claimed. ☐

This result fails in general without assuming that the sets be boundedly compact.

In particular, if there is a unique element of best approximation to x in a boundedly compact set M, then all 'good' points $z \in M$ (i.e., points at which $\|x - z\|$ is almost identical to $E(x, M)$) lie in some small neighbourhood of the best approximant to x.

Theorem 10.2 (on stability). *Let* X *be a Banach space and let* M *be a boundedly compact subset of* X. *Suppose that* $x \in X$ *has a unique element of best approximation* $y^* \in M$. *Then, if* $\{x_n\}$ *is a sequence converging to* x *and if* $\{y_n\}$ *is a sequence in* M *such that* $\|x_n - y_n\| \to \|x - y^*\|$, *then* $y_n \to y^*$.

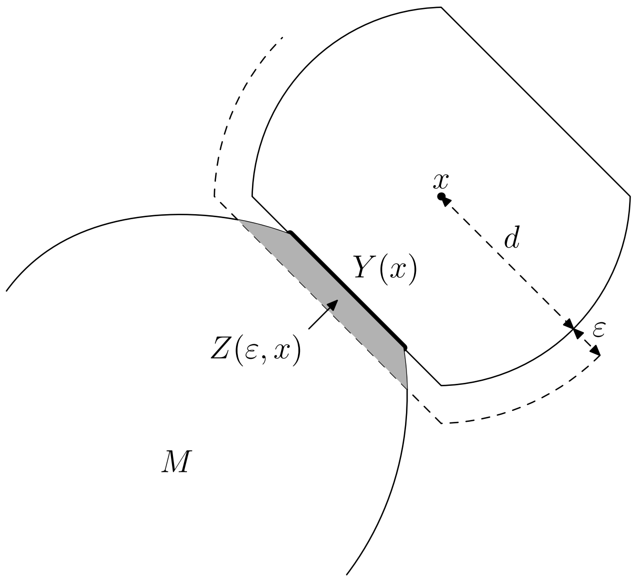

Fig. 10.4

Proof. Indeed, for any $\delta > 0$ and all sufficiently large n, we have

$$\|x - y_n\| \leqslant \|x - x_n\| + \|x_n - y_n\| \leqslant \|x - y^*\| + \delta,$$

i.e., $y_n \in Z(\delta)$. By Proposition 10.1, for any $\varepsilon > 0$ there exists a sufficiently small δ such that $Z(\delta) \subset Y_\varepsilon$. So $\|y_n - y^*\| \leqslant \varepsilon$ for all sufficiently large n, and so $y_n \to y^*$. □

Since $E(x, M)$ is continuous, we have the following result.

Corollary (see, e.g., [42, p. 390]). *Let M be a boundedly compact subset of X such that any $x \in X$ has a unique element of best approximation $y(x)$ in M (i.e., M is a boundedly compact Chebyshev set). Then $y(x)$ is a continuous function of x on X. Moreover, the function $y(x)$ is uniformly continuous on every compact subset of X.*

In the space $C = C[0, 1]$ we approximate by functions in the set

$$\{x \in C : \quad \|x'\| \leqslant 1, \quad x(0) = 0\}.$$

Is the function $y(x)$ continuous? In which Banach spaces the metric projection $x \mapsto y(x)$ onto any subspace M is uniformly continuous?

Let H be a Hilbert space, M be a subspace of H, $x \in X$. An element $y(x)$ is an element of best approximation to x if and only if

$$(x - y(x), y(x)) = 0 \qquad \forall\, y \in M.$$

Hence

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

and so

$$\|y(x)\| \leqslant \|x\|.$$

The metric projection onto a subspace is linear in H, and hence

$$\|y(x) - y(x')\| = \|y(x - x')\| \leqslant \|x - x'\|;$$

i.e., the metric projection onto a subspace is uniformly continuous in a Hilbert space and so is a bounded linear operator.

Remark. In general (and as a rule, if a space is not a Hilbert space), the metric projection fails to be linear. For example, we approximate the functions

$$f_1(t) = \begin{cases} 1 - 2t, & 0 \leqslant t < 1/2, \\ 0, & 1/2 \leqslant t \leqslant 1, \end{cases} \quad f_2(t) = f_1(1-t), \quad (f_1 + f_2)(t) = 2|t - 1/2|$$

by constant functions in $C[-1,1]$ (see Fig. 10.5) .

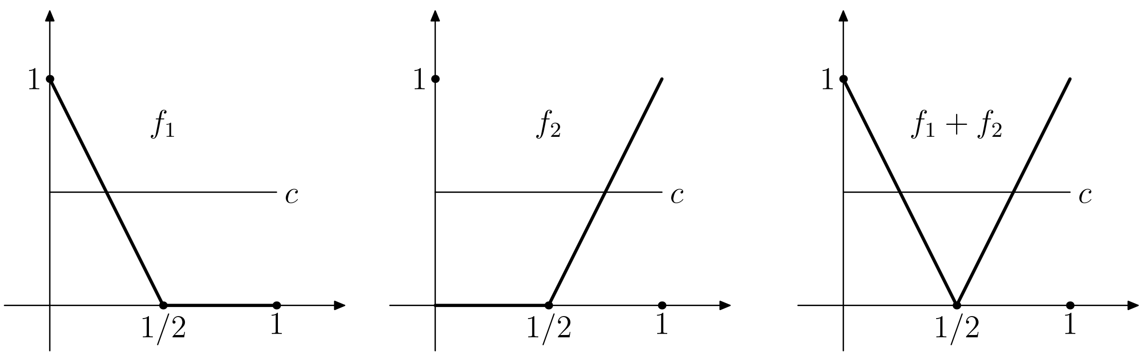

Fig. 10.5

We have in this setting: $\frac{1}{2} = c^*(f_1) = c^*(f_2) = c^*(f_1 + f_2) \neq c^*(f_1) + c^*(f_2) = 1$.

The metric projection onto subspaces is linear only in Hilbert spaces (and in some degenerate cases).

Theorem 10.3. *Let M be a subspace of a uniformly convex space X. Then M is a subspace of uniqueness and the metric projection $y(x)$ onto M depends uniformly continuously on x on any closed bounded set.*

Proof. Since a uniformly convex space is strictly convex, the first assertion follows by Theorem 9.3. Applying the triangle inequality, we have, for arbitrary points x and x' that have nearest in points M,

$$\|x - y(x')\| \leqslant \|x' - y(x')\| + \|x - x'\| \leqslant \|x' - y(x)\| + \|x - x'\| \leqslant$$

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

i.e., if x' is close to x, then $\|y(x') - x\|$ is close to $\|x - y(x)\|$ uniformly in x and x'. Hence, $y(x') \in Z(2\|x - x'\|, x)$. The space X being uniformly convex, the distance between $y(x)$ and $y(x')$ decreases uniformly with decreasing $\|x - x'\|$, provided that the norms are uniformly bounded. □

The spaces L^p, $1 < p < \infty$, are uniformly convex, and hence in L^p, $1 < p < \infty$, the metric projection $y(x)$ onto a subspace is uniformly continuous on any closed bounded subset.

In a Hilbert space we have already pointed out that the metric projection onto a subspace is uniformly continuous on the whole space.

The following example shows that in $C[a,b]$ the metric projection onto a subspace may fail to be uniformly continuous (and may even be discontinuous [43]).

Example 10.2. Consider approximation by functions $a+bx = p(x)$ in $C[0,1]$. For any $\varepsilon > 0$, we built functions $f, \widetilde{f} \in C[0,1]$ such that $\|f - \widetilde{f}\|_C \leqslant \varepsilon$, but $\|p^*(f) - p^*(\widetilde{f})\| > 1$. This means that the metric projection $p(f)$ is not uniformly continuous. An example to illustrate this situation is constructed with the help of the so-called 'lightning' function (see Fig. 10.6). Here, $\widetilde{f}(\varepsilon) = 1+\varepsilon$, $\widetilde{f}(-\varepsilon) = 1-\varepsilon$, $f(-1) = \widetilde{f}(-1) = 0$, $f(1) = \widetilde{f}(1) = 0$, $f(0) = \widetilde{f}(0) = -1$, $f(\varepsilon) = f(-\varepsilon) = 1$, $\widetilde{f}(-\varepsilon) = 1-\varepsilon$, $\widetilde{f}(\varepsilon) = 1+\varepsilon$ are the vertices of the broken lines given by the graphs of $f(x)$ and $\widetilde{f}(x)$.

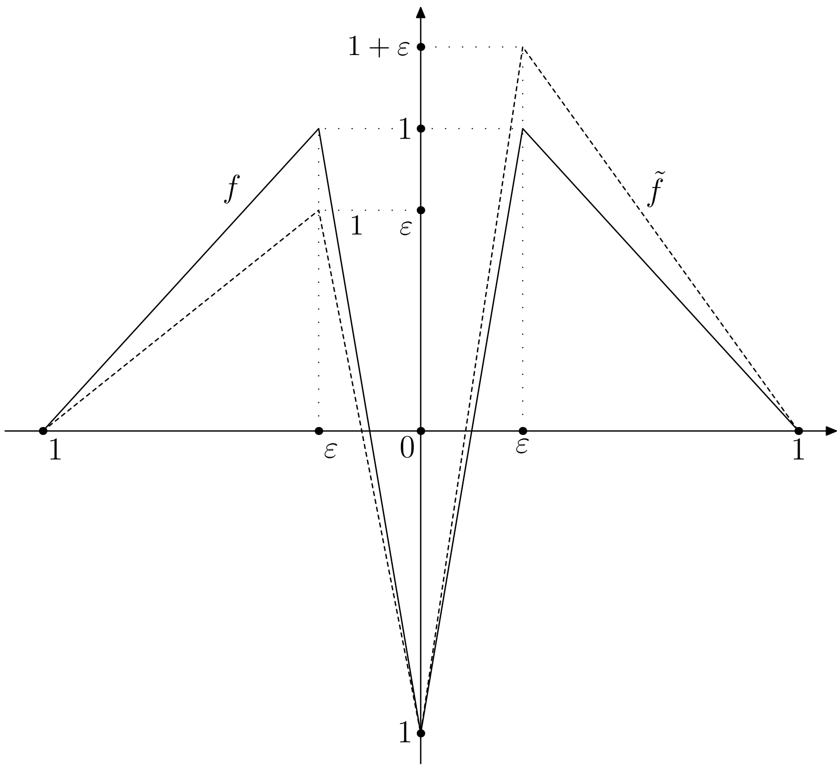

Fig. 10.6

From the Chebyshev alternation theorem (which will be proved later) it follows that $p^*(f) \equiv 0$ is an element of best approximation to f and $p^*(\widetilde{f}) = x$ is an element of best approximation to $\widetilde{f}$. In this example we have Chebyshev alternants of the required length.

Example 10.3. In $C[0,1]$ we approximate by rational functions of the form

$$M = \left\{ \frac{a}{b + cx} \right\}$$

Let us show that in this case the best approximation operator is discontinuous.

Indeed, using the classical Chebyshev's Equioscillation Theorem (see e.g. [5], [45, Ch. 7], [37]), for each $r > 0$ one can construct a continuous function f_r for which $(1 + rx)^{-1}$ is the rational function of least deviation (as seen from the Chebyshev alternants at three points) and which is such that $f_r(x) \rightrightarrows f$. We have: the rational function of best approximation to f is identically zero, while $(1+rx)^{-1} \not\rightrightarrows 0$ as $r \to \infty$. The function $R(x)$, which is the limit for the this fraction, is discontinuous on $[0,1]$: we have $R(0) = 1$, $R(x) = 0$, $0 < x \leqslant 1$.

Lecture 11

Approximative compactness. Approximation in C

11.1. Continuity of the metric projection

Let X be a metric space, $M \subset X$, and let $Y(x)$ be the set of points of best approximation (metric projection) to a point $x \in X$ in the set M. If M is a Chebyshev set (i.e., for any x its element of best approximation from M exists and is unique) and if, in addition, M is boundedly compact, then metric projection

$$x \mapsto Y(x) = \{y(x)\}$$

is continuous (see the corollary on p. 84).

Let X be a Banach space and $M \subset X$. A subset M of X is called *approximatively compact* if, for each $x \in X$, any minimizing sequence $\{y_n\}$ of elements in M (i.e., such that $\|x - y_n\| \to E(x, M)_X$ as $n \to \infty$) contains a subsequence convergent to an element of M.

If a set M is approximatively compact and if the metric projection $Y(x)$ consists of one point, then any minimizing sequence $\{y_n\}$ converges to this point. Any approximatively compact set is always closed.

Example 1. The unit sphere $S_1 = \{x \colon \|x\| = 1\}$ of an infinite-dimensional Hilbert space H is not approximatively compact, while the set $M = S_1 \bigcup \{0\}$ is approximatively compact.

Let us establish some properties of approximatively compact sets.

1. If M is approximatively compact, then $Y(x) \neq \varnothing$ for all $x \in X$.
2. For any $x \in X$ the metric projection $Y(x)$ of x onto an approximatively compact set M is always compact, because any sequence from $Y(x)$ is minimizing and, hence contains a convergent subsequence.

Theorem 11.1 (I. Singer [41], [42, p. 390]). *Let M be an approximatively compact set. Suppose that a point x_0 has a unique element of best approximation $y(x_0)$. Then the metric projection $Y(x)$ is continuous at x_0 (i.e., $y_n \to y(x_0)$ as $n \to \infty$) for any sequence $\{x_n\}$ converging to x_0 and any point y_n nearest to x_n).*

Proof. Indeed, suppose that a sequence $\{x_n\}$ converges to x_0. Consider an arbitrary $y_n \in Y(x_n)$. Then $\{y_n\}$ is a minimizing sequence for x_0, because we have

$$\|x_0 - y_n\| = \|x_0 - x_n + x_n - y_n\| \leqslant \|x_0 - x_n\| + \|x_n - y_n\|$$

$$= \|x_0 - x_n\| + E(x_n, M) \to E(x_0, M)$$

as $n \to \infty$. Hence $y_n \to y(x_0)$ as $n \to \infty$. □

Corollary. *If M is an approximatively compact Chebyshev set, then the metric projection $y(x)$ onto M is continuous at any point of X.*

Remark. It is worth pointing out one more important case when the metric projection $Y(x)$ is continuous. Let X be a Banach space and M be a hyperplane. There is no loss of generality in assuming that $0 \in M$. Suppose we are given a point $x_0 \in X \setminus M$. In this case, any element $x \in X$ is uniquely representable as

$$x = y + \alpha x_0, \qquad y \in M, \qquad \alpha \in \mathbb{R}.$$

It follows that $Y(x) = y + \alpha Y(x_0)$. Hence if some $x_0 \notin M$ has a unique element of best approximation $Y(x_0)$ in M, then any $x \in X$ has a unique element of best approximation $Y(x)$ in M, which depends continuously on x. Here the continuity follows since the metric projection in this case is linear:

$$Y(\alpha x_1 + \beta x_2) = \alpha Y(x_1) + \beta Y(x_2), \qquad \|Y(x)\| \leqslant 2\|x\|.$$

There are examples of trivial Chebyshev sets which are approximatively compact. Consider, for example, the whole space or a singleton.

Remark. There is a nonreflexive nonseparable infinite-dimensional Banach space which fails to have a nontrivial Chebyshev subspace (A. L. Garkavi [13], see also [43, § 3.2]).

Problem. Prove or disprove that every separable Banach space of dimension exceeding 1 contains nontrivial Chebyshev subspaces?

In $C[0,1]$ there are nontrivial Chebyshev subspaces (for example, the subspace of constant functions). The space $L[0,1]$ also contains nontrivial Chebyshev subspaces.

Example 2. Consider the space $L(0,1)$ and equip it with the norm $\|f\| = \int_0^1 |f(t)|\,dt$. Let

$$M = \big\{\varphi \in L(0,1) \colon \varphi(t) = 0 \quad \forall\, t \in [0, 1/2]\big\}.$$

In Fig. 11.1, φ^* is a unique element of best approximation to $f \in L(0,1)$.

Problem. Prove that any finite-dimensional normed linear space contains a nontrivial Chebyshev hypersubspace.

Remark. Actually much more can be said. Any n-dimensional normed linear space contains a Chebyshev subspace of any dimension not exceeding n (V. A. Zalgaller [48]).

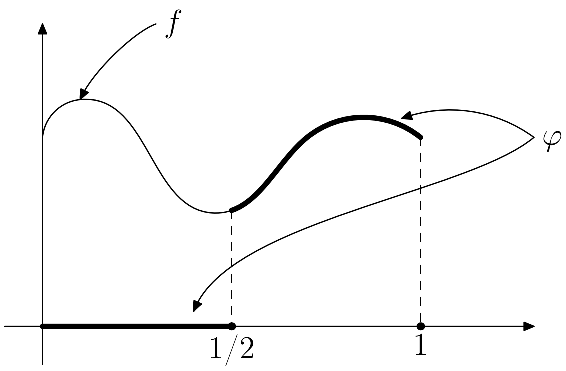

Fig. 11.1

11.2. Approximation in the space $C_{2\pi}$

Finding an element of best approximation may present some challenge. This therefore suggests the problem of evaluating the best approximation $E(f, M)$ (the distance of f to M). Usually $E(f, M)$ can be bounded from above $E(f, M) \leqslant \|f - \varphi\|$ with some appropriate function $\varphi \in M$.

We give some examples by considering the subset $M = \mathcal{T}_n$ consisting of trigonometric polynomials of degree at most n in the space $C_{2\pi}$ of continuous 2π-periodic functions.

1. Let $f \in C_{2\pi}$ and let $\sum\limits_{k=0}^{\infty} A_k(x)$ be its Fourier series, where

$$A_0(x) = \frac{a_0}{2}, \qquad A_k(x) = a_k \cos kx + b_k \sin kx.$$

The best mean-square approximation is known to be realized by the partial sum s_n of the Fourier series,

$$\sum_{k=n+1}^{\infty} (a_k^2 + b_k^2) = \frac{1}{\pi} \int_{-\pi}^{\pi} (f - s_n)^2 \, dx \leqslant \frac{1}{\pi} \int_{-\pi}^{\pi} (f - t_n)^2 \, dx \leqslant 2\|f - t_n\|_C^2,$$

where $t_n \in \mathcal{T}_n$ is an arbitrary trigonometric polynomial. Hence,

$$E(f, \mathcal{T}_n)_C \geqslant \frac{1}{\sqrt{2}} \Big(\sum_{k=n+1}^{\infty} a_k^2 + b_k^2 \Big)^{1/2}.$$

2. For the de la Vallée Poussin sums $\sigma_{2n,n}(x)$, we have, by the Lebesgue inequality,

$$E(f, \mathcal{T}_{2n})_C \leqslant \|\sigma_{2n,n}(x) - f\|_C \leqslant \frac{2(2n+1)}{2n - n + 1} E(f, \mathcal{T}_n)_C \leqslant 4E(f, \mathcal{T}_n)_C.$$

Hence, if we know approximation of a function by the de la Vallée Poussin sums (and this is a fairly simple problem, because $\sigma_{2n,n}$ is a linear operator), then it is possible to estimate the best approximation from above and below.

3. For Fourier sums, we have the estimate

$$\|f - s_n\|_C \leqslant \left\{ \frac{4}{\pi^2} \ln n + O(1) \right\} E(f, \mathcal{T}_n)_C.$$

11.3. Approximation by rational functions

Let $R_{m,n} = R_{m,n}[a,b]$ be the set of all algebraic rational functions $R(x) = P(x)/Q(x)$, $\deg P \leqslant m$, $\deg Q \leqslant n$, which are supposed to be defined everywhere on $[a,b]$.

Given a function $f \in C[a,b]$, we approximate it by rational functions $R \in R_{m,n}$. Let us evaluate the best approximation

$$\inf_{R \in R_{m,n}} \|f - R\|_C = \rho_{m,n}(f).$$

We assume that the fraction is irreducible and that

$$\deg P = m - \mu, \qquad \deg Q = n - \nu.$$

Suppose that R is continuous on $[a,b]$; i.e., the poles of R do not lie on $[a,b]$. Let $a \leqslant x_1 < x_2 < \ldots < x_N \leqslant b$ and let $f(x_k) - R(x_k) = \lambda_k$ $(k = 1, \ldots, N)$. If $\operatorname{sign} \lambda_k \cdot \operatorname{sign} \lambda_{k+1} = -1$, $k = 1, \ldots, N-1$ (the signs of λ_k alternate), then the finite sequence $\{x_k\}$ will be said to form the *de la Vallée Poussin's alternant of length* N for the difference $f - R$ (see Fig. 11.2 with $N = 5$).

Theorem 11.2 (de la Vallée Poussin). *Suppose that $f - R$ has the de la Vallée Poussin's alternant of length $N = m + n - \min\{\mu, \nu\} + 2$. Then $\rho =: \rho_{m,n}(f) \geqslant \min_k |\lambda_k|$.*

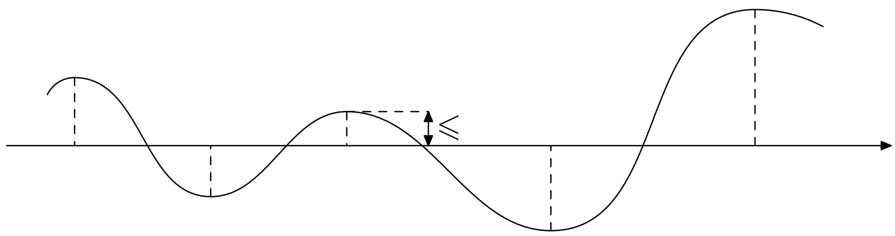

Fig. 11.2

Proof. Assume, to the contrary, that there exists a rational function $r = p/q \in R_{m,n}$ such that

$$\|r - f\|_C < \min_k |\lambda_k|.$$

Consider the difference $R(x) - r(x)$ at the points $x = x_k$:

$$R(x_k) - r(x_k) = R(x_k) - f(x_k) + f(x_k) - r(x_k) = -\lambda_k + (f(x_k) - r(x_k)).$$

We clearly have

$$\operatorname{sign}(R(x_k) - r(x_k)) = \operatorname{sign}(R(x_k) - f(x_k)) = -\operatorname{sign}\lambda_k.$$

Let $\Delta(x) = R(x) - r(x)$. Then the values of $\Delta(x)$ alternate at the points x_k. Hence Δ has at least $N-1$ zeros between these points. But

$$\Delta(x) = R(x) - r(x) = \frac{P(x)}{Q(x)} - \frac{p(x)}{q(x)} = \frac{Pq - Qp}{Qq},$$

where $\deg Pq \leqslant m - \mu + n$, $\deg Qp \leqslant m + n - \nu$. Since, by the hypothesis

$$N - 1 = m + n - \min(\mu, \nu) + 1 > m + n - \min(\mu, \nu),$$

the number $N-1$ of zeros in the denominator exceeds its degree, a contradiction. □

11.4. Chebyshev systems

A system (φ) of functions $\varphi_1(x), \ldots, \varphi_n(x)$ in $C[a,b]$ is called a *Chebyshev system* (or a Chebyshev system of order n) on the interval $[a,b]$ if, for any distinct points $x_1, x_2, \ldots, x_n \in [a,b]$, the determinant $\mathcal{D}(x_1, \ldots, x_n) = \det(\varphi_i(x_k))$ does not vanish.

In particular, for $n = 1$ the function φ_1 does not vanish on $[a,b]$; i.e., it is of constant sign.

Similarly one can define a Chebyshev system in the space $C(K)$ on an arbitrary compact set K.

Here are some properties of Chebyshev systems.

1. Chebyshev systems are precisely interpolating systems in the sense that the Lagrange interpolation problem is always uniquely solvable.

We recall that the Lagrange interpolation problem consists in finding, for a given system of knots $\{x_k\}$ and values $\{y_k\}$, a polynomial $\varphi(x) = \sum_{i=1}^n a_i\varphi_i(x)$ such that $\sum_{i=1}^n a_i\varphi_i(x_k) = y_k$, $k = 1, \ldots, n$.

2. Any Chebyshev system on an interval is linearly independent, inasmuch as any nontrivial polynomial in a Chebyshev system of order n has at most $n-1$ zeros.

Exercise. Any Chebyshev system on $[a,b]$ is *a fortiori* a Chebyshev system on any proper subinterval of $[a,b]$. Prove that the converse is not true.

3. The determinant

$$\mathcal{D}(\xi) = \mathcal{D}(\xi_0, \xi_1, \ldots, \xi_{n-1}) = \begin{vmatrix} \varphi_1(\xi) & \varphi_2(\xi) & \ldots & \varphi_n(\xi) \\ \varphi_1(\xi_1) & \ldots & \ldots & \varphi_n(\xi_1) \\ \ldots & \ldots & \ldots & \ldots \\ \varphi_1(\xi_{n-1}) & \ldots & \ldots & \varphi_n(\xi_{n-1}) \end{vmatrix}$$

is of constant sign on the set $\mathcal{M} = \{\xi :\ \xi_0 < \ldots < \xi_{n-1}\}$. This is because the function $\mathcal{D}(\xi)$ is continuous and has no zeros on $\mathcal{M}$ and since, for any $\xi, \xi' \in \mathcal{M}$, the n-tuple ξ can be continuously transformed in the n-tuple ξ' within $\mathcal{M}$.

4. Given an arbitrary set $\Sigma = \{\xi_i\}_{i=1,\ldots,n-1} \subset [a,b]$, $\xi_1 < \ldots < \xi_{n-1}$, there exists a polynomial φ in a Chebyshev system on this interval which such that, firstly, its set of

zeros coincides with Σ, and secondly, φ changes sign at the zeros. It suffices to consider $\varphi(x) = A\mathcal{D}(x, \xi_1, \dots, \xi_{n-1})$. Let $A = 1$, $\sigma = \operatorname{sign} \mathcal{D}(\xi)$, $\xi \in \mathcal{M}$, $i \in \{1, \dots, n-1\}$, $a < \xi_i < b$, and let x', x'' be points sufficiently close to ξ_i and such that $x' < \xi_i < x''$. Then $\varphi(x') = (-1)^{i-1}\mathcal{D}(\xi_1, \dots, \xi_{i-1}, x', \xi_i, \dots, \xi_{n-1})$. Since $(\xi_1, \dots, \xi_{i-1}, x', \xi_i, \dots, \xi_{n-1}) \in \mathcal{M}$, we have $\operatorname{sign} \varphi(x') = (-1)^{i-1}\sigma$. Similarly, $\operatorname{sign} \varphi(x'') = (-1)^{i}\sigma$. Hence φ changes sign at x_i.

5. There exists a polynomial in a Chebyshev system on $[a, b]$ such that it has constant sign on this interval (being strictly positive or negative).

In contrast to the case of algebraic polynomials (with respect to the system $1, x, \dots, x^{n-1}$), assertion 5) is far from being trivial in the general setting. First, we build a polynomial $P_0(x)$ which is nonnegative on $[a, b]$. To do so we can take for $P_0(x)$ the uniform limit of the convergent sequence of polynomials $\pm\frac{\varphi(x)}{\|\varphi\|_{C[a,b]}}$ as $\xi = (\xi_1, \dots, \xi_{n-1}) \to (a, a, \dots, a)$.

With the appropriately chosen sign, the pre-limiting polynomials are all positive on $(\xi_{n-1}, b]$, and hence, $P_0(x)$ is nonnegative on $[a, b]$. Let $x_1, \dots, x_r$ be the zeros of the polynomial P_0. Then $r \leqslant n-1$. Let $x_{r+1}, \dots, x_n$ be arbitrary distinct points different from $x_1, \dots, x_r$. Since (φ) is an interpolating system, there exists a polynomial Q with respect to this system such that $Q(x_1) = \ldots = Q(x_n) = 1$. Let $M = \|Q\|_{C[a,b]}$. We take $\varepsilon > 0$ so as to have $Q(x) > 0$ for all $i = 1, \dots, n$ and all x such that $|x - x_i| < \varepsilon$. Consider the set

$$E = \{x \in [a, b] \colon \exists\, i = 1, \dots, n-1 \quad |x - x_i| < \varepsilon\}.$$

Then $\delta := \min_{x \in [a,b] \setminus E} P_0(x) > 0$. Let $P(x) = P_0(x) + \frac{\delta}{2M} Q(x)$. We claim that $P(x) > 0$ for all $x \in [a, b]$. If fact, we have $P_0(x) \geqslant 0$ and $Q(x) > 0$ for $x \in E$. If $x \in [a, b] \setminus E$, then $P_0(x) \geqslant \delta$, $\frac{\delta}{2M} Q(x) \geqslant -\frac{\delta}{2}$.

11.5. Approximation of continuous functions by polynomials with respect to a Chebyshev system

Theorem 11.3 (see e.g. [7, Theorem 1.1]). *Any nonChebyhsev system (φ) consisting of n continuous linearly independent functions generates a nonChebyshev subspace; i.e., there exists a function f which has at least two polynomials φ_1^* and φ_2^* of best approximation with respect to this system.*

Proof. There exist points $x_1 < x_2 < \ldots < x_n$ on $[a, b]$ such that $\det(\varphi_i(x_k)) = 0$; i.e., the rows and columns of the determinant are linearly dependent. Hence there exist c_i, not all zeros, such that

$$\sum_{i=1}^{n} c_i \varphi_i(x_k) = 0, \qquad k = 1, \dots, n, \tag{11.1}$$

and there exist d_k, not all zeros, such that

$$\sum_{k=1}^{n} d_k \varphi_i(x_k) = 0, \qquad i = 1, \dots, n. \tag{11.2}$$

Hence, for any polynomial $\varphi = \sum_{i=1}^{n} a_i \varphi_i$ we have

$$\sum_{k=1}^{n} d_k \varphi(x_k) = 0, \qquad k = 1, \dots, n. \tag{11.3}$$

Let the function f be defined as follows: $f(x_k) = \operatorname{sign} d_k$ for $k = 1, \dots, n$ (assuming that $\operatorname{sign} 0 = 0$), f is linear on the intervals $[x_k, x_{k+1}]$ and is constant on the intervals $[a, x_1]$ and $[x_n, b]$. By the construction, $\|f\|_C = 1$.

From (11.3) it follows that, for any polynomial φ, there exists a number k such that $d_k \neq 0$ and $d_k \varphi(x_k) \leqslant 0$. Hence $\|f - \varphi\|_C \geqslant |f(x_k) - \varphi(x_k)| \geqslant 1 = \|f - 0\|_C$; i.e., the zero function is a polynomial of best approximation in the system (φ) for the function f just constructed.

By (11.1), there exists a polynomial $\varphi_0(x) \not\equiv 0$ such that $\varphi_0(x_k) = 0$, $k = 1, \dots, n$. We set

$$f_\varepsilon(x) = f(x) \cdot (1 - |\varepsilon \varphi_0(x)|),$$

where $\varepsilon > 0$ is such that $\|\varepsilon \varphi_0\|_C < 1$. For this function we have $f_\varepsilon(x_k) = \operatorname{sign} d_k$, $\|f_\varepsilon\|_C = 1$, and now, by the same argument as for f, it follows that the zero function is its polynomial of best approximation. In addition, for any point $x \in [a, b]$, we have

$$|f_\varepsilon(x) + \varepsilon \varphi_0(x)| \leqslant |f(x)| \cdot (1 - |\varepsilon \varphi_0(x)|) + |\varepsilon \varphi_0(x)| \leqslant (1 - |\varepsilon \varphi_0(x)|) + |\varepsilon \varphi_0(x)| = 1,$$

and hence $-\varepsilon \varphi_0$ is also a polynomial of best approximation to f_ε.

Thus, f_ε has at least two polynomials of best approximation, and hence (φ) generates a non-Chebyshev subspace. $\square$

Remark. A similar assertion can also be proved if we replace the interval $[a, b]$ by an arbitrary compact set K.

Lecture 12

Chebyshev systems. Haar's theorem

12.1. Chebyshev subspaces of $C(K)$

We are concerned with approximation of real-valued functions by finite-dimensional subspaces in the C-metric.

Let $f \in C[a,b]$ and let $L_n \subset C[a,b]$ be the subspace generated by a system of linearly independent functions $(\varphi) = \{\varphi_1, \dots, \varphi_n\}$. Our purpose is to find a polynomial $\varphi^*(f)$ of best approximation in the system (φ) (i.e., a polynomial that minimizes $E(f, L_n)_C$). Above we have shown that if (φ) is not a Chebyshev system, then there exists an element $C[a,b]$ for which there are at least two polynomials of best approximation.

Problem. On which compact sets K there are vector-valued Chebyshev systems with values in $\mathbb{R}^m$?

For $m = 1$ there is the well-known Mairhuber's theorem[1]: the space $C(K)$ on a compact Hausdorff space K contains a Chebyshev subspace of dimension $n+1$, $n \geqslant 1$, if and only if K is homeomorphic to a subset of the unit circle in $\mathbb{R}^2$. Moreover, K can be homeomorphic to the entire circle if and only if n is even.

Theorem 12.1 (Haar; see e.g. [36], [42, p. 215]). *Let K be a compact set. A linearly independent system (φ) generates a Chebyshev subspace in $C(K)$ if and only if (φ) is a Chebyshev system on K.*

Proof. The necessity has been addressed in the previous lecture: namely, it was proved (for $K = [a,b]$) that if a system is not Chebyshev then there exists a function having at least two polynomials of best approximation.

Sufficiency. Suppose that $(\varphi) = \{\varphi_1, \dots, \varphi_n\}$ is a Chebyshev system of function. We claim that it is a system of uniqueness. The system (φ) being linearly independent, the cardinality of K is at least n. If $\operatorname{card} K = n$, then the span of (φ) coincides with $C(K)$, and hence it is a Chebyshev set.

Let $f \in C(K)$ and let $\varphi(x) = \sum_{k=1}^{n} a_k \varphi_k(x)$ $(x \in K)$ be an arbitrary polynomial in a given system of functions. Consider the *set of points of maximal deviation* of f from φ:

$$M(f, \varphi) = \{x \in K \colon \|f - \varphi\|_C = |f(x) - \varphi(x)|\}.$$

[1] Also known as the Mairhuber–Curtis theorem (see e.g. [2]).

Since K is compact and since f, φ are continuous functions, this set is nonempty.

Now let φ^* be a polynomial of best approximation to f; i.e., $\|\varphi^* - f\| = E(f, L_n)_C$. We need the following auxiliary result.

Proposition 12.1. *Let $f \in C(K)$ and let φ^* be a polynomial of best approximation to f in a Chebyshev system of order n. Then the set $M(f, \varphi^*)$ cannot be too small, namely*

$$\operatorname{card} M(f, \varphi^*) \geqslant n + 1.$$

Proof. Assume the contrary. Suppose that, for some f and φ, we have $\operatorname{card} M(f, \varphi) \leqslant n$. In particular, $f \notin L_n$, $E_n(f, L_n)_C > 0$. We claim that such a polynomial φ is not a polynomial of best approximation. To do so we need to build a polynomial $h \in (\varphi)$ such that the deviation of f from $\varphi + \varepsilon h$ is smaller than that from φ,

$$\|f - (\varphi + \varepsilon h)\| < \|f - \varphi\|,$$

for some $\varepsilon > 0$. Let $M(f, \varphi)$ be composed of the points $x_1, \dots, x_n$ (the same argument applies if the number of points $\{x_k\}$ is smaller than n). Since (φ) is an interpolating system, there exists a polynomial h such that, for all x_k,

$$h(x_k) = f(x_k) - \varphi(x_k) (\neq 0).$$

We surround the points x_k by neighbourhoods U_k on which the functions h and $f - \varphi$ are of constant sign. Hence $|f - \varphi - \varepsilon h| < \|f - \varphi\|$ in U_k for small $\varepsilon > 0$. Outside the union of the U_k, we have $|f - \varphi| < \|f - \varphi\|_C$, and so $|f - \varphi - \varepsilon h| < \|f - \varphi\|_C$ for all sufficiently small ε. □

We now return to the proof of Haar's theorem. Assume, to the contrary, that there is a function $f \in C(K)$ having two polynomials φ_1^* and φ_2^* of best approximation:

$$\|f - \varphi_1^*\| = \|f - \varphi_2^*\| = E(f, L_n)_C = E.$$

Then, since the set of polynomials of best approximation is convex, we have $\|f - \varphi\| = E$ with $\varphi = t\varphi_1^* + (1 - t)\varphi_2^*$ for any $t \in [0, 1]$. In particular, for $t = \frac{1}{2}$, the polynomial $\widetilde{\varphi} = \frac{1}{2}\varphi_1^* + \frac{1}{2}\varphi_2^*$ is a polynomial of best approximation, and hence, by Proposition 12.1, there exist points x_k, $k = 1, 2, \dots, n + 1$, at which

$$\|f - \widetilde{\varphi}\| = |f(x_k) - \widetilde{\varphi}(x_k)| = E.$$

For φ_1^* and φ_2^*, we have

$$|f(x_k) - \varphi_1^*(x_k)| = E, \qquad |f(x_k) - \varphi_2^*(x_k)| = E$$

at these points, where the sign of the differences must be the same, i.e.,

$$f(x_k) - \varphi_1^*(x_k) = f(x_k) - \varphi_2^*(x_k) = \pm E.$$

Consequently,

$$h(x_k) = \varphi_1^*(x_k) - \varphi_2^*(x_k) = (f(x_k) - \varphi_2^*(x_k)) - (f(x_k) - \varphi_1^*(x_k)) = 0,$$

$k = 1, \dots, n + 1$, for the polynomial $h = \varphi_1^* - \varphi_2^*$, which is not identical zero by the assumption. Hence, a nonzero polynomial in the Chebyshev system has $n + 1$ zeros, which cannot be true. Hence, $h \equiv 0$, a contradiction. □

Remark. The condition card $M(f, \varphi^*) \geqslant n+1$ for a polynomial φ^* of best approximation is necessary but not sufficient.

Example 3. In $C[0,1]$ we approximate by constant functions. Let $n = 1$. Then we have $\varphi^* = \frac{1}{2}$ for a constant function of best approximation (see Fig. 12.1)

$$\operatorname{card} M(f, \varphi) \geqslant n + 1 = 2.$$

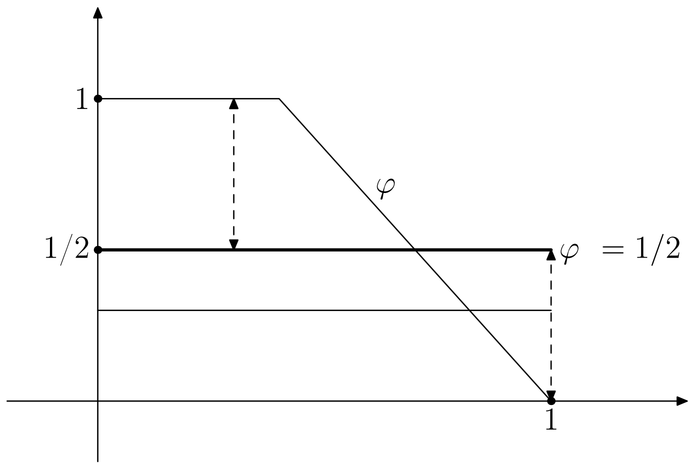

Fig. 12.1

For any constant function $\varphi < \frac{1}{2}$ there is a continuum of points of maximal deviation, and so this function does not deliver the best approximation.

12.2. Chebyshev's theorem

Let f be a continuous function on $[a, b]$, let (φ) be a Chebyshev system, and let $\varphi = \sum_{k=1}^{n} a_k \varphi_k(x)$ be a polynomial in this system.

Consider the set $M = M(f, \varphi)$ of the points of maximum deviation.

The set $M(f, \varphi)$ is said to have a *Chebyshev* $(n+1)$*-point alternant* if there exist points $x_1, x_2, \ldots, x_{n+1}$ in M such that

1) $a \leqslant x_1 < x_2 < \ldots < x_{n+1} \leqslant b$,

2) the signs of the differences $f(x_k) - \varphi(x_k)$ alternate $(k = 1, 2, \ldots, n+1)$.

Theorem 12.2 (Chebyshev; see e.g. [8, Ch 3, § 5]). *A necessary and sufficient condition for a polynomial φ in a Chebyshev system of order n on a closed interval to be a polynomial of least deviation from f is that the set of points of maximal deviation $M(f, \varphi)$ contains an alternant consisting of at least $n+1$ points.*

Proof. Necessity. Assume that $M(f, \varphi)$ does not contain an $(n+1)$-point alternant. We claim that then there exists a polynomial with smaller deviation.

Let $M(f,\varphi) = M_+ \bigcup M_-$, where

$$M_+ = \{x \in M(f,\varphi)\colon f(x) - \varphi(x) > 0\}, \qquad M_- = \{x \in M(f,\varphi)\colon f(x) - \varphi(x) < 0\}.$$

We can also assume that $M_+ \neq \varnothing$ and $M_- \neq \varnothing$, for otherwise there exists a polynomial $\varphi(x) - \varepsilon P_+(x)$ with smaller deviation, where $P_+(x) > 0$ on $[a,b]$, $\varepsilon > 0$ for $M_+ \neq \varnothing$, $\varepsilon < 0$ for $M_- \neq \varnothing$, and $|\varepsilon|$ is a sufficiently small number. That a polynomial $P_+(x)$ in a Chebyshev system exists was proved earlier. Let $\eta_1 = \min\{x\colon x \in M(f,\varphi)\}$. There is no loss of generality in assuming that $\eta_1 \in M_+$. We set

$$\eta_2 = \min\{x > \eta_1\colon x \in M_-\}, \quad \eta_3 = \min\{x > \eta_2\colon x \in M_+\},$$

and so on. This gives a system of points $\eta_1 < \ldots < \eta_k$; by the assumption, we have $1 < k \leqslant n$. Further, let

$$\zeta_1 = \max\{x \in M_+\colon x < \eta_2\}, \quad \zeta_2 = \max\{x \in M_-\colon x < \eta_3\}, \quad \ldots,$$
$$\zeta_k = \max\{x \in M\} \qquad (\zeta_k \geqslant \eta_k).$$

Finally, consider an arbitrary point $\xi_i \in (\zeta_i, \eta_{i+1})$, $i = 1,\ldots,k-1$. Surround the intervals $[\eta_i,\zeta_i]$, $i = 1,\ldots,k$, by intervals (a_i,b_i) not containing the points ξ_i (see Fig. 12.2). In the case $\eta_1 = a$ and/or $\zeta_k = b$, we take the intervals $[a,b_1)$ and $(a_k,b]$ instead of (a_1,b_1) and (a_k,b_k). By the above, if $k = n$, then there exists a polynomial $h(x) = \pm\mathcal{D}(x;\xi_1,\ldots,\xi_{n-1})$, which changes sign at the points $\xi_1,\ldots,\xi_{n-1}$.

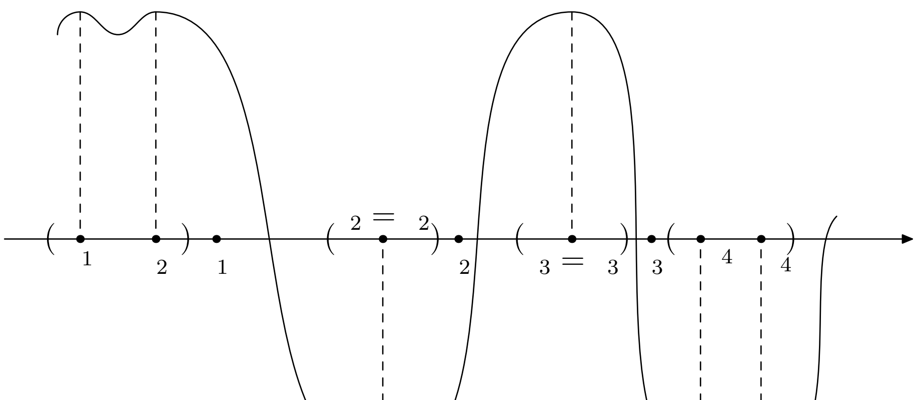

Fig. 12.2

By choosing the sign in the formula for $h(x)$ appropriately, one may assume that $h(x) > 0$ on the intervals (a_{2i-1},b_{2i-1}) and that $h(x) < 0$ on the intervals (a_{2i},b_{2i}). If $k < n$, then we put the missing distinct points of $\{\xi_i\}_{i=k}^{n-1}$ on the interval (ξ_1,a_2) in case there are an even number of them, and take $\xi_{n-1} = a$ for $a < \eta_1$ or $\xi_{n-1} = b$ for $\zeta_k < b$ if $n-k$ is odd; we accommodate the remaining points of $\{\xi_i\}_{i=k}^{n-2}$ on the

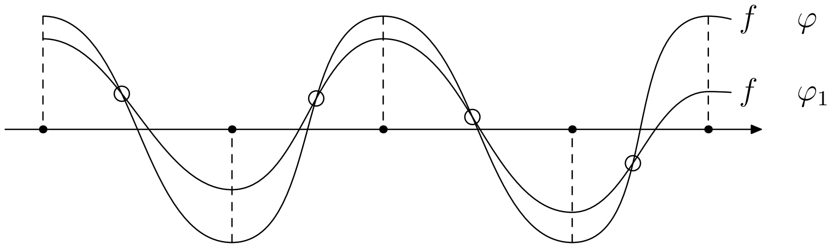

Fig. 12.3

interval (ξ_1, a_2). Clearly, the polynomial $h(x)$ constructed from these points preserves the property $\operatorname{sign} h(x) = \operatorname{sign}(f(x) - \varphi(x))$ for $x \in (a_i, b_i) \bigcap M$, $(i = 1, 2, \ldots, k)$, which was already proved for $k = n$.

It remains to construct the same polynomial in the remaining case, when $n - k$ is odd and $\eta_1 = a$, $\zeta_k = b$. In this case, on the already chosen points $\{\xi_k\}_{k=1}^{n-2}$ we build two polynomials $h_a(x)$ and $h_b(x)$, which vanish at these points and at the points $x = a$ and $x = b$, respectively. Taking $h = h_a + h_b$ we again obtain a polynomial with the required properties. Proceeding precisely as in the last part of the proof of Proposition 12.1, it follows that $\|f - \varphi - \varepsilon h\| < \|f - \varphi\|$ for small $\varepsilon > 0$.

Sufficiency. Assume there is an $(n+1)$-point alternant. We claim that in this case φ is a polynomial of least deviation from the function f. Assume the contrary. Then there exists another polynomial φ_1 of smaller deviation to f. Then the difference $\varphi_1 - \varphi$ alternates at the points of alternant (see Fig. 12.3). Hence $\varphi_1 - \varphi$ has n zeros, which is impossible. □

Remark. In case when a Chebyshev system is made up of polynomials of degree $\leqslant n$, a necessary and sufficient condition that a polynomial p_n be a polynomial of least deviation to a function f if that $M(f, p_n)$ should contain an $(n+2)$-point alternant.

12.3. De la Vallée Poussin's alternant

Let $(\varphi) = \{\varphi_k(x)\}_{k=1}^{n}$ be a Chebyshev system of continuous functions on the interval $[a, b]$, let $f \in C[a, b]$, and let $F \subset [a, b]$. We set

$$E(f, (\varphi), F) = \inf \|f - \varphi\|_{C(F)},$$

where the infimum is taken over polynomials φ in the system (φ). It is worth noting that Chebyshev's theorem is also valid for the restriction of f to a set F containing at least $n + 1$ points, namely: *a necessary and sufficient condition that a polynomial φ be a polynomial of best approximation to f on F is that the difference $f - \varphi$ have an at least $(n+1)$-point Chebyshev alternant on this set.*

Further let $F = F_{n+1}$ be an $(n+1)$-point de la Vallée Poussin's alternant; i.e., F_{n+1} is a finite subsequence $\{x_k\}$ of $[a, b]$, $x_1 < \ldots < x_{n+1}$, which is such that

$$f(x_k) - \varphi_F(x_k) = (-1)^k \sigma \lambda_k$$

for some polynomial φ_F, where $\sigma \in \{1, -1\}$, $\lambda_k > 0$, $k = 1, \dots, n+1$.

Let us consider the deviation of the polynomial from a function f defined on such a set $F_{n+1} \subset [a, b]$. We clearly have

$$E(f, (\varphi), [a, b]) \geqslant E(f, (\varphi), F_{n+1}).$$

Since φ_F is a polynomial in the system (φ), it may be assumed that on F_{n+1} we approximate the function f, $f(x_k) = (-1)^k \lambda_k$. Having this in mind, we shall try to calculate $E(f, (\varphi), F_{k+1})$ explicitly.

Taking $\varphi(x) = \sum_{i=1}^n a_i \varphi_i(x)$, we require that

$$f(x_k) - \varphi(x_k) = (-1)^k \rho, \qquad k = 1, \dots, n+1$$

on F_{n+1}. This is a linear system of equations in the coefficients a_i, $i = 1, \dots, n$, and in the unknown deviation ρ; so we have $n+1$ equations with $n+1$ unknowns:

$$(-1)^k \lambda_k = (-1)^k \rho + \sum_{i=1}^n a_i \varphi_i(x_k).$$

Let us find ρ. The determinant of this system

$$\begin{vmatrix} -1 & \varphi_1(x_1) & \dots & \varphi_n(x_1) \\ 1 & \dots & \dots & \dots \\ \vdots & & & \\ (-1)^{n+1} & \varphi_1(x_{n+1}) & \dots & \varphi_n(x_{n+1}) \end{vmatrix} = -\sum_{k=1}^{n+1} \mathcal{D}(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_{n+1})$$

does not vanish, since we assume that $x_1 < x_2 < \ldots < x_{n+1}$, and hence all the determinants under the summation sign must have the same sign. Indeed, one can transform from one determinant to another by changing the system of knots $\{x_k\}$ continuously, the determinant being nonzero. The deviation $E(f, (\varphi), F_{n+1}) = \rho$ is given by

$$\rho = \sum_{k=1}^{n+1} \lambda_k \mathcal{D}(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_{n+1}) \Big/ \sum_{k=1}^{n+1} \mathcal{D}(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_{n+1}).$$

Changing if necessary the sign of the numerator and denominator, it is found that ρ is some average of the numbers λ_k with positive weights. Hence, we have the following estimate for the de la Vallée Poussin's alternant F_{n+1}:

$$E(f, (\varphi), [a, b]) \geqslant \rho \geqslant \min_k \lambda_k,$$

where $\lambda_k = |f(x_k) - \varphi_F(x_k)|,\ x_k \in F_{n+1},\ k = 1, 2, \dots, n.$

Theorem 12.3 (decomposition theorem). *We have*

$$E(f, (\varphi), [a, b]) = \sup_{F_{n+1}} E(f, (\varphi), F_{n+1}) = E(f, (\varphi), F^*_{n+1}),$$

where F^*_{n+1} is a Chebyshev alternant for the function f on $[a, b]$.

Proof. Clearly, the first term is majorized by the second term. We prove the converse inequality. As F_{n+1} we take $F^*_{n+1} = \{x^*_1, \dots, x^*_{n+1}\}$, which is a Chebyshev alternant for the polynomial of best approximation on the whole interval. Then the function cannot be better approximated on this set than on the entire interval, because by Chebyshev's Equioscillation Theorem theorem the polynomials of best approximation coincide on $[a, b]$ and on F^*_{n+1}, whence the theorem. □

Lecture 13

Uniqueness theorems

13.1. Chebyshev rank of a subspace

Let M be a finite-dimensional subspace of a Banach space X and $x \in X$. As before, $Y(x)$ is the set of elements of best approximation in M to x. This is a closed convex set.

Given a fixed $y_0 \in Y(x)$, consider the set $\{y - y_0\}$, $y \in Y(x)$ (the metric projection onto Y). The number $r(x)$ of linearly independent elements of the set $\{y-y_0\}$, $y \in Y(x)$, is called the *dimension of the set of elements of best approximation*. It is clear that

$$0 \leqslant r(x) \leqslant n = \dim\ M.$$

For Chebyshev subspaces, $r(x) = 0$ for any $x \in X$. This is a characterizing property of Chebyshev subspaces.

Consider the following numerical characteristics of a given subspace M:

$$\sup_{x \in X} r(x) = \max_{x \in X} r(x) = R(M).$$

Clearly, $0 \leqslant R(M) \leqslant n$. The number $R(M)$ is called the *Chebyshev rank* of the subspace M. For Chebyshev subspaces, we have $R(M) = 0$.

The infimum $\inf_{x \in X} r(x)$ is always zero (and is attained on $x \in M$), and so there is no need to consider this characteristic of a subspace.

It turns out that $C[a, b]$ admits subspaces of any Chebyshev rank.

In $L[a, b]$ we encounter a different situation. In 1964, A. L. Garkavi proved that in a nonatomic $L_\mu[a, b]$ the rank of any finite-dimensional subspace coincides with its dimension.

Sometimes it is also advantageous to study the Chebyshev rank of a subspace M of X not with respect to the whole space X, but relative to some subset $\mathfrak{M}$ of X; in general, $\mathfrak{M}$ is not a subspace.

The quantity

$$\sup_{x \in \mathfrak{M}} r(x) = R(M, \mathfrak{M})_X \leqslant R(M)_X$$

is called the *Chebyshev rank* of a subspace M with respect to the subset $\mathfrak{M}$ of the space X.

In the classical spaces C and L, one can construct $\mathfrak{M}$ and M such that

$$R(M,\mathfrak{M})_X < R(M,X).$$

For example, if in $L[a,b]$ we take $\mathfrak{M} = C[a,b]$, then there are finite-dimensional Chebyshev subspaces with respect to $\mathfrak{M}$ (Jackson's theorem, to be proved).

13.2. Periodic case

Let $C_{2\pi}$ be the space of 2π-periodic continuous functions. However, in this setting, the algebraic polynomials are not good at all for the purpose of approximation. Instead one may speak about approximating such functions by trigonometric polynomials of order n:

$$t_n(x) = a_0 + \sum_{k=0}^{n}(a_k \cos kx + b_k \sin kx).$$

When counting zeros correctly (on $[0, 2\pi)$), this is a Chebyshev subspace of order $2n+1$, and an analogue of Chebyshev's theorem is valid: a trigonometric polynomial of order n is a polynomial of best approximation to a given function if and only if there is a $(2n+2)$-point Chebyshev alternant for its deviation with the function $2n+2$ (on any semi-interval of the form $[\alpha, \alpha+2\pi)$; existence of an alternant is independent of the choice of α).

13.3. Lacunary trigonometric series

The results in this subsection are given without proofs.

Suppose we are given two series

$$a_0 + \sum_{k=1}^{\infty}(a_{n_k} \cos n_k x + b_{n_k} \sin n_k x). \tag{13.1}$$

We assume that $1 \leqslant n_1 < n_2 < \ldots < n_k < \ldots$. Lacunary trigonometric series are series in which the terms that differ from zero are 'sufficiently sparse'.

Hadamard lacunary series: there exists $\lambda > 1$ such that

$$n_{k+1}/n_k \geqslant \lambda > 1 \qquad \forall\ k,$$

i.e., n_k increase at least as rapidly as a geometric progression with ratio $\lambda > 1$:

$$n_{k+1} \geqslant \lambda n_k, \qquad n_k \geqslant \lambda^{k-1} n_1 \geqslant \lambda^{k-1}.$$

Lacunary trigonometric series have a number of properties not shared by standard trigonometric series.

Let $f \in C_{2\pi}$. Suppose that its Fourier series has form (13.1) *and is lacunary in the sense of Hadamard, $n_{k+1}/n_k \geqslant \lambda > 1$. Then it converges uniformly* (*and even absolutely*) *to f* (A. N. Kolmogorov, 1924 and S. Sidon, 1927, respectively).

Suppose that $\rho_k = \sqrt{a_{n_k}^2 + b_{n_k}^2}$. If f is a continuous function with the Fourier series (13.1), then

$$\sum_{k=0}^{\infty} \rho_k \leqslant C(\lambda)\|f\|_C$$

and, if $\lambda > 1$ is fixed, then

$$\|f\|_C \asymp \sum_{k=0}^{\infty} \rho_k. \tag{13.2}$$

Exercise. Consider, for f in question, the symmetric difference operator with step $2h$

$$\Delta_h f(x) = f(x+h) - f(x-h) = 2\sum_{k=1}^{\infty} \sin n_k h(-a_{n_k} \sin n_k x + b_{n_k} \cos n_k x).$$

Apply (13.2) to estimate the norm $\|\Delta_h f\|_C$ and the modulus of continuity $\omega(f,\delta) = \sup_{|h|\leqslant\delta} \|\Delta_h f\|_C$ in terms of the Fourier coefficients of f.

13.4. Best approximation of functions representable by lacunary series

Consider the Weierstrass series

$$\sum_{k=1}^{\infty} a^k \cos b^k x,$$

where $b \in \mathbb{N}$, $b > 1$, $0 < a < 1$. With appropriate a and b this series is converges uniformly and absolutely to the Weierstrass function $f \in C_{2\pi}$, which is nowhere differentiable.

Theorem 13.1 (S. N. Bernstein). *If b is odd, then, for any n,*

$$s_n(x) = \sum_{k:\, b^k \leqslant n} a^k \cos b^k x$$

is a trigonometric polynomial that best approximates f in the uniform norm among all polynomials of degree at most n.

Proof. We have $f(x) - s_n(x) = \sum_{b^k > n} a^k \cos b^k x$. For $x = 0$ this function has a maximum

$$\|f - s_n\| = \sum_{b^k > n} a^k = \sum_{k=k_0}^{\infty} a^k, \tag{13.3}$$

where $b^{k_0} > n \geqslant b^{k_0 - 1}$. Let us find how many times the maximum is attained. For $b^k > n$, each of the functions $a^k \cos b^k x$ has period $\frac{2\pi}{b^{k_0}}$. At the point $x = \frac{\pi}{b^{k_0}}$, they all equal to -1 (b is odd), and hence,

$$f\left(\frac{\pi}{b^{k_0}}\right) - s_n\left(\frac{\pi}{b^{k_0}}\right) = -\|f - s_n\|.$$

Thus, on $[0, 2\pi)$ the difference $f - s_n$ has Chebyshev alternant of cardinality $2b^{k_0} \geqslant 2n + 2$. Hence, s_n is a polynomial of least deviation from f. ∎

Remark. For the sine series, we have

$$f(x) - s_n(x) = \sum_{b^k > n} a^k \sin b^k x$$

and so formula (13.3) fails. If we add π/b^{k_0} to x, then the sine terms will all change signs, and even though formula (13.3) does not hold, we still have a $2b^{k_0}$-point Chebyshev alternant on $[0, 2\pi)$, and so the theorem also remains valid in this case.

Theorem 13.2 (S. B. Stechkin). *Suppose that a function $f \in C_{2\pi}$ is expandable in a lacunary Fourier series* (13.1), $\frac{n_{k+1}}{n_k} \geqslant \lambda > 1$. *Then $E_n(f)_C \asymp \sum\limits_{n_k > n} \rho_k$; more precisely,*

$$C(\lambda) \sum_{n_k > n} \rho_k \leqslant E_n(f)_C \leqslant C_1(\lambda) \sum_{n_k > n} \rho_k,$$

where the numbers $C(\lambda)$, $C_1(\lambda)$ depend only on λ, $0 < C(\lambda) \leqslant C_1(\lambda) < \infty$.

Proof. Let us note one property of the Fourier coefficients. Since $\cos kx$ is orthogonal to any trigonometric polynomial t_n of order n, $n < k$, we have

$$a_k = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos kx \, dx = \frac{1}{\pi} \int_0^{2\pi} \{f - t_n^*\} \cos kx \, dx,$$

where t_n^* is the polynomial of least deviation from f in the uniform norm. Hence

$$|a_k| \leqslant E_n(f)_C \cdot \frac{4}{\pi}, \qquad k > n.$$

Similarly, for b_k, we have

$$|b_k| \leqslant E_n(f)_C \cdot \frac{4}{\pi}, \qquad k > n.$$

We also note that if $\sum A_{n_k}$ is a lacunary series with $\lambda > 1$, then between n and $2n$ there are a limited number of terms which depend only on λ (or order $\ln 2/\ln \lambda$).

Consider the difference $f - \sigma_{2n,n}$, where $\sigma_{2n,n} = \frac{1}{n+1} \sum_{k=n}^{2n} s_k$ are de la Vallée Poussin's sums (see Lecture 6). We already know that $\|f - \sigma_{2n,n}\| \leqslant 4E_n(f)_C$. Let us find out by how much the de la Vallée Poussin's sum s_n differs for any series with terms $A_k(x)$. We have (see Fig. 13.1).

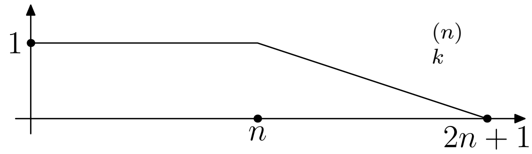

Fig. 13.1

$$\sigma_{2n,n} - s_n = \frac{1}{n}\sum_{k=n+1}^{2n} s_k - s_n = \sum_{k=0}^{2n} \lambda_k^{(n)} A_k(x) - \sum_{k=0}^{n} A_k(x) = \sum_{k=n+1}^{2n} \lambda_k^{(n)} A_k(x),$$

hence, as in the sum $\sum\limits_{k=n+1}^{2n}$ for a lacunary series, the number of terms depends only on λ,

$$|\sigma_{2n,n} - s_n| \leqslant \sum_{k=n+1}^{2n} |A_k(x)| \leqslant C(\lambda) E_n(f)_C.$$

Hence,

$$E_n(f)_C \leqslant \|f - s_n\| \leqslant \|f - \sigma_{2n,n}\| + \|\sigma_{2n,n} - s_n\| \leqslant C_1(\lambda) E_n(f)_C;$$

and so, for lacunary series, if $\lambda > 1$, then we have $\|f - s_n\| \asymp E_n(f)_C$.

Applying Sidon's theorem and inequality (13.2) to $f - s_n$, this establishes the relation $E_n(f)_C \asymp \sum\limits_{n_k > n} \rho_k$, which holds for any function having a lacunary Fourier series with $\lambda > 1$. □

Example 4. A function, for which the Fourier sums give the best order of approximation, can be constructed in the following way:

$$f = \sum_{n=1}^{\infty} a_n \cos nx, \qquad \left|\frac{a_{n+1}}{a_n}\right| < q < 1.$$

We have

$$\|\sigma_{2n,n} - s_n\| \leqslant \sum_{k=n+1}^{2n} |a_k| \leqslant C(q)|a_{n+1}| \leqslant C(q) E_n(f)_C.$$

Since $\|f - \sigma_{2n,n}\| \leqslant 4E_n(f)_C$, it follows that

$$\|f - s_n\| \leqslant \|f - \sigma_{2n,n}\| + \|\sigma_{2n,n} - s_n\| \leqslant 4E_n(f)_C + C(q)E_n(f)_C = C_1(q)E_n(f)_C.$$

Remark. For multivariate functions this problem is very difficult.

13.5. Approximation of function by finite-dimensional subspaces in $L[a, b]$

Let $L = L[a, b]$ be the space of Lebesgue integrable functions on $[a, b]$.

We already know that the one-dimensional subspace consisting of constant functions is not a Chebyshev subspace in L (see the example on p. 81).

Exercise. Prove that in $L[a, b]$ there are no finite-dimensional Chebyshev subspaces (except the trivial ones); see e.g. [12], [16].

The space $C[a, b]$ of continuous functions on $[a, b]$ is a dense linear manifold in L.

Theorem 13.3 (D. Jackson, see e.g. [47, p. 38]). *Let $(\varphi) = \{\varphi_1, \varphi_2, \dots, \varphi_n\}$ be a Chebyshev system of continuous functions $[a, b]$, and let L_n be the linear hull of (φ). Then the polynomial $\varphi^* \in L_n$ of best approximation in the L-norm is unique for any function $f \in C[a, b]$, $\|f - \varphi^*\|_L = E(f, L_n)_L$.*

We first need several auxiliary results.

Lemma 13.1. *If ψ_1, ψ_2 are polynomials of best approximation for a function $f \in C[a, b]$ in the space $L[a, b]$, then $(f(x) - \psi_1(x))(f(x) - \psi_2(x)) \geqslant 0$ for any point $x \in [a, b]$.*

Proof. Let ψ_1 and ψ_2 be polynomials of least deviation from f. Then $\psi = (\psi_1 + \psi_2)/2$ is also a polynomial of least deviation form f, and so

$$\int_a^b |f - \psi|\, dt = \frac{1}{2}\Big\{\int_a^b |f - \psi_1|\, dt + \int_a^b |f - \psi_2|\, dt\Big\},$$

hence,

$$\int_a^b |f - \psi_1 + f - \psi_2|\, dt = \int_a^b |f - \psi_1|\, dt + \int_a^b |f - \psi_2|\, dt.$$

For continuous functions the last equality is satisfied if and only if the differences $f - \psi_1$ and $f - \psi_2$ have the same signs. ☐

Lemma 13.2. *Let ψ_1, ψ_2 be distinct polynomials of least deviation for $f \in C[a, b]$ in $L[a, b]$ with respect to a Chebyshev system (φ). Also let $\alpha \in (0, 1)$, $\varphi_\alpha(t) = \alpha\psi_1(t) + (1 - \alpha)\psi_2(t)$. Then the difference $f - \psi_\alpha$ has at most $(n - 1)$ zeros.*

Proof. Let $f(t) - \varphi_\alpha(t) = 0$. By Lemma 13.1, $f(t) - \psi_1(t) = f(t) - \psi_2(t) = 0$. Hence any zero of the function $f - \varphi_\alpha$ is also a zero of $\psi_1 - \psi_2$. But the last function has at most $(n - 1)$, because (φ) is a Chebyshev system. ☐

Lemma 13.3. *Let $(\varphi) = \{\varphi_1, \varphi_2, \dots, \varphi_n\}$ be a Chebyshev system on $[a, b]$. Let $\varphi(t) = \sum_{k=1}^n a_k\varphi_k(t)$ be polynomials in this system such that $\|\varphi\|_L = \int_a^b |\varphi(t)|\, dt = \mathcal{D}$ for some $\mathcal{D}$. Then, for any measurable subset E of $[a, b]$,*

$$J(E) = \int_E |\varphi(t)|\, dt \leqslant K\mathcal{D} \operatorname{mes} E,$$

where K depends only on (φ).

Proof. Since $\int_a^b |\varphi(t)|\, dt$ is a continuous function of a_k, $k = 1, \dots, n$, it attains its minimum on the sphere $\sum_{k=1}^n a_k^2 = 1$. But this minimum cannot be zero, because (φ) is a Chebyshev system. Hence,

$$\int_a^b |\varphi(t)|\, dt \geqslant C > 0 \qquad \forall\ (a_1, \dots, a_n)\colon\ \sum a_k^2 = 1.$$

Now if $l = \big(\sum_{k=1}^n a_k^2\big)^{1/2} \neq 0$ and if $\int_a^b |\varphi(t)|\, dt = \mathcal{D}$, it follows that $\int_a^b \frac{|\varphi(t)|}{l}\, dt \geqslant C > 0$; i.e., $\mathcal{D} = \int_a^b |\varphi(t)|\, dt \geqslant Cl$. Hence $|a_k| \leqslant l \leqslant \frac{\mathcal{D}}{C}$. Using this bound, we have

$$\int_E |\varphi(t)|\, dt \leqslant \sum_{k=1}^n |a_k| \int_E |\varphi_k(t)|\, dt \leqslant K\mathcal{D} \operatorname{mes} E,$$

where K is some constant depending only on $\varphi_1, \varphi_2, \dots, \varphi_n$. ☐

Proof of Jackson's theorem. It is required to show that if $f \in C[a,b]$ and if $\{\varphi_1(t), \dots, \varphi_n(t)\}$ is a Chebyshev system, then the polynomial $\varphi(t) = \sum_{k=1}^{n} a_k \varphi_k(t)$ of least L-deviation from f is unique.

By Lemma 13.2, if $R(t) = f(t) - \varphi(t)$ has at least n changes of sign, then φ is a unique polynomial of least deviation from f.

We assume that there exist two polynomials ψ_1 and ψ_2 of least deviation from f. Let $R(t) = f(t) - \varphi_\alpha(t)$, where $\varphi_\alpha(t)$ is an arbitrary polynomial of the form

$$\varphi_\alpha(t) = \alpha\psi_1(t) + (1-\alpha)\psi_2(t), \qquad \alpha \in [0,1].$$

We claim that the difference $R(t)$ vanishes on a set of positive measure, which is independent of α. This will give us a contradiction with Lemma 13.2.

We renumber the points of sign change of the difference $R(t)$ as follows:

$$a < t_1 < \ldots < t_q < b, \qquad q \leqslant n-1.$$

If necessary, we enlarge this family to $n-1$ points on $[b-\delta, b]$, where $b - \delta > t_q$:

$$t_1 < \ldots < t_q < t_{q+1} < \ldots < t_{n-1} < b.$$

Since (φ) is a Chebyshev system, we have, by what was proved in Section 13.4, that there exists a polynomial $F(t)$ in this system vanishing exactly at those $n-1$ points, at which it changes sign. It may be assumed that $\|F\|_C = 1$. This can be achieved by taking $F(t) = \widetilde{F}(t)/\|\widetilde{F}\|_C$ or $F(t) = -\widetilde{F}(t)/\|\widetilde{F}\|_C$, where

$$\widetilde{F}(t) = \begin{vmatrix} \varphi_1(t) & \ldots & \varphi_n(t) \\ \varphi_1(t_1) & \ldots & \varphi_n(t_1) \\ \ldots & \ldots & \ldots \\ \varphi_1(t_{n-1}) & \ldots & \varphi_n(t_{n-1}) \end{vmatrix}.$$

The sign of $F(t)$ is chosen so as to have $\operatorname{sign} F(t) = \operatorname{sign} R(t)$ on $[a, b-\delta]$ (see Fig. 13.2)

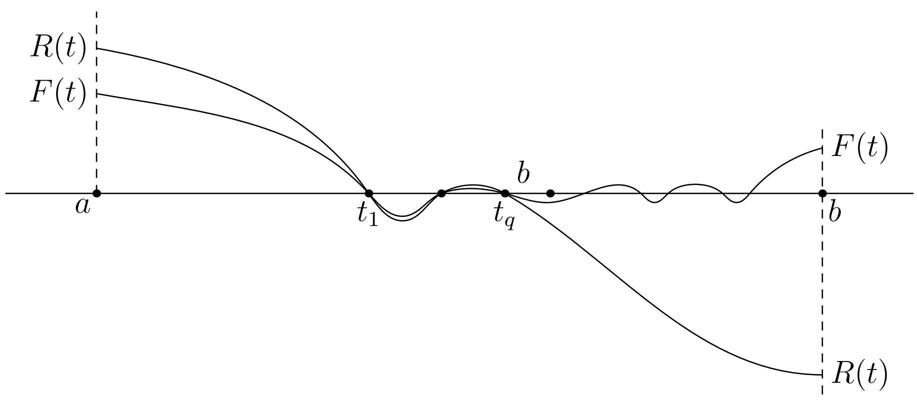

Fig. 13.2

Consider the difference $R(t) - \varepsilon F(t)$, $\varepsilon > 0$, and define three sets on $[a, b]$:

$$r: \quad |R(t)| > \varepsilon, \quad \operatorname{sign} R = \operatorname{sign} F,$$

$$s: \quad |R(t)| \leqslant \varepsilon, \quad \operatorname{sign} R = \operatorname{sign} F,$$

$$u: \quad \operatorname{sign} R(t) \neq \operatorname{sign} F.$$

By the construction of F, we have $u \subset [b - \delta, b]$ and $R(t) - \varepsilon F(t) = f(t) - \widetilde{\varphi}(t)$, where $\widetilde{\varphi}$ is some polynomial in the system (φ). Hence,

$$\int_a^b |R(t) - \varepsilon F(t)|\, dt \geqslant \int_a^b |R(t)|\, dt,$$

inasmuch as $\int_a^b |R(t)|\, dt = E(f, (\varphi))_L$. We further have

$$\int_r |R(t) - \varepsilon F(t)|\, dt = \int_r |R(t)|\, dt - \varepsilon \int_r |F(t)|\, dt,$$

because $\operatorname{sign} R(t) = \operatorname{sign} F(t)$ and $|R(t)| > \varepsilon$, $|F(t)| \leqslant 1$ for $t \in r$; also,

$$\int_s |R(t) - \varepsilon F(t)|\, dt \leqslant \int_s |R(t)|\, dt + \varepsilon \int_s |F(t)|\, dt.$$

It follows that

$$\int_a^b |R(t)|\, dt \leqslant \int_a^b |R(t) - \varepsilon F(t)|\, dt = \int_r + \int_s + \int_u$$

$$\leqslant \int_r |R(t)|\, dt - \varepsilon \int_r |F(t)|\, dt + \int_s |R(t)|\, dt + \varepsilon \int_s |F(t)|\, dt + \int_u |R(t)|\, dt + \varepsilon \int_u |F(t)|\, dt$$

$$= \int_a^b |R(t)|\, dt - \varepsilon \int_r |F(t)|\, dt + \varepsilon \int_{s \cup u} |F(t)|\, dt,$$

and so,

$$\int_r |F(t)|\, dt \leqslant \int_{s \cup u} |F(t)|\, dt.$$

Adding $\int_{s \cup u} |F(t)|\, dt$ to the both sides of this inequality and using Lemma 13.3, this gives

$$\mathcal{D} = \int_a^b |F(t)|\, dt \leqslant 2 \int_{s \cup u} |F(t)|\, dt \leqslant 2K\mathcal{D} \operatorname{mes}(s \cup u), \qquad K = K\big((\varphi)\big).$$

Hence $\operatorname{mes}(s \cup u) \geqslant c > 0$, where c is independent of $\mathcal{D}$. Taking $\delta < c/2$, it is found that $\operatorname{mes} s \geqslant c/2 > 0$, but for $\varepsilon \to 0$ the set s converges to the set on which the difference $R(t) = 0$. Hence, for any $\alpha \in [0, 1]$, the difference $R_\alpha = f(t) - \varphi_\alpha(t)$ vanishes on the set s_α, $\operatorname{mes} s_\alpha \geqslant c/2 > 0$. But this contradicts Lemma 13.2. □

Exercise. Where in the course of the proof we used continuity of the functions φ_k and f?

Remark. The measure in the proof of the theorem was the Lebesgue measure. Note that the argument just given does not work for arbitrary measures.

Lecture 14

Jackson's theorem

14.1. Approximation in $L_2(a,b)$

Throughout this lecture we assume that $X = L^2_{2\pi}$ and that $E_n(x)_{L^2}$ is the best approximation in $L^2_{2\pi}$ of a function $x \in L^2_{2\pi}$ by the subspace $\mathcal{T}_n$ (of dimension $2n+1$) of all trigonometric polynomials t_n with respect to the system

$$1, \cos t, \sin t, \cos 2t, \sin 2t, \dots .$$

This is a complete system in $L^2_{2\pi}$. This means

$$\forall\, x \in L^2_{2\pi}: \qquad E_n(x) \longrightarrow 0 \qquad (n \to \infty).$$

Given a fucntion $x \in L^2_{2\pi}$, consider its Fourier series

$$x(t) \sim \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kt + b_k \sin kt).$$

We set

$$s_n(t) = \frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos kt + b_k \sin kt).$$

By Parseval's formula,

$$E_n^2(x)_{L^2} = \|x(t) - s_n(t)\|^2_{L^2} = \sum_{k=n+1}^{\infty} (a_k^2 + b_k^2).$$

14.2. Bernstein's inequality

Given any trigonometric polynomial of degree n, we have (see e.g. [21, Ch. 3])

$$\|t_n'\|_{L^2} \leqslant n\|t_n\|_{L^2}.$$

Indeed,

$$\|t_n'\|^2_{L^2} = \sum_{k=1}^{n} k^2(a_k^2 + b_k^2) \leqslant n^2\Big(\frac{a_0^2}{2} + \sum_{k=1}^{n}(a_k^2 + b_k^2)\Big) = n^2\|t_n\|^2_{L^2}.$$

14.3. Moduli of oscillation and continuity

We set $\Delta_h x(t) = x(t+\frac{h}{2}) - x(t-\frac{h}{2})$. Then the quantity

$$\|\Delta_h x\|_{L^2} = \varkappa(h,x)$$

is called the *modulus of oscillation* of a function x with step h (in $L^2_{2\pi}$), and the quantity

$$\sup_{|h|\leqslant\delta} \varkappa(h,x) = \omega(\delta,x)_{L^2}$$

is called the *modulus of continuity* of x (in $L^2_{2\pi}$).

Remark. The modulus of continuity may cannot decrease too rapidly. If $\varkappa(h,x) = o(h)$ as $h \to 0$; i.e., $\|\Delta_h x/h\|_{L^2} \to 0$ as $h \to 0$, then its L_2-derivative vanishes almost everywhere, and so $x = \mathrm{const}$.

We set

$$\Delta_h^k x(t) = \Delta_h^{k-1}(\Delta_h x(t)), \qquad \left\|\Delta_h^k x\right\|_{L^2} = \varkappa_k(h,x).$$

Then the quantity

$$\sup_{|h|\leqslant\delta} \varkappa_k(h,x) = \omega_k(\delta,x)_{L^2}$$

is called the *modulus of continuity of order k*.

14.4. Jackson's theorem in $L^2_{2\pi}$

Theorem 14.1 (Jackson's inequality; see [45, Part II, § 9], [47]). *Let $x \in L^2_{2\pi}$. Then*

1) $E_n(x)_{L^2} \leqslant C\omega\big(\frac{\pi}{n}, x\big)_{L^2}$;
2) $E_n(x)_{L^2} \leqslant C_k\omega_k\big(\frac{\pi}{n}, x\big)_{L^2}$, $k \in \mathbb{N}$;

where C, C_k are absolute constants.

Proof. We have

$$E_n^2(x)_{L^2} = \sum_{k=n+1}^{\infty}(a_k^2+b_k^2), \qquad \Delta_h x(t) \sim \sum_{k=1}^{\infty}\Big(2\sin\frac{kh}{2}\Big)(-a_k\sin kt + b_k\cos kt),$$

and hence

$$\varkappa^2(h,x) = 4\sum_{k=1}^{\infty}\sin^2 k\frac{h}{2}(a_k^2+b_k^2).$$

But

$$\frac{1}{\delta_n}\int_0^{\delta_n}\sin^2 k\frac{h}{2}\,dh \geqslant c > 0 \qquad k \geqslant n, \quad \delta_n = \frac{\pi}{n},$$

and so

$$\frac{1}{\delta_n}\int_0^{\delta_n}\sum_{k=1}^{\infty}\sin^2 k\frac{h}{2}(a_k^2+b_k^2)\,dh \geqslant c\sum_{k=n+1}^{\infty}(a_k^2+b_k^2).$$

As a result,

$$\sum_{k=n+1}^{\infty}(a_k^2+b_k^2)\leqslant c_1\frac{1}{\delta_n}\int_0^{\delta_n}\sum_{k=1}^{\infty}\sin^2 k\frac{h}{2}(a_k^2+b_k^2)\;dh$$

$$=c_2\frac{1}{\delta_n}\int_0^{\delta_n}\varkappa^2(h,x)\;dh\leqslant c_2\omega(\delta_n,x)_{L^2}. \qquad \square$$

The proof of the second assertion of Theorem 14.1 follows the same lines, because

$$\Delta_h^k x(t)\sim\sum_{l=1}^{\infty}\Big(2\sin\frac{lh}{2}\Big)^k\Big(a_l\cos\Big(lx+\frac{k\pi}{2}\Big)+b_l\sin\Big(lx+\frac{k\pi}{2}\Big)\Big).$$

Corollary. *Let x be an absolutely continuous function with the derivative in $L^2_{2\pi}$. Then*

$$\Big\|\frac{\Delta_h x(t)}{h}\Big\|_{L^2}\leqslant K$$

and hence $E_n(x)=O(n^{-1})$, because $\varkappa(h,x)\leqslant Kh$. Therefore $\omega(\delta,x)_{L^2}\leqslant K\delta$.

14.5. Converse theorem

Theorem 14.2. *For any $n\in\mathbb{N}$ and any function $x\in L^2_{2\pi}$*

$$\omega^2\Big(\frac{\pi}{n},x\Big)_{L^2}\leqslant\frac{C}{n^2}\sum_{k=1}^{n}kE_{k-1}^2(x)_{L_2},$$

where $C=2\pi^2$.

Proof. We split $\|\Delta_h x\|^2$ into two sums

$$\|\Delta_h x(t)\|_{L^2}^2=\sum_{k=1}^{n-1}\Big(2\sin\frac{kh}{2}\Big)^2(a_k^2+b_k^2)+\sum_{k=n}^{\infty}\Big(2\sin\frac{kh}{2}\Big)^2(a_k^2+b_k^2)=I_1+I_2,$$

and evaluate these sums using the inequalities $|\sin x|\leqslant|x|$, $|\sin x|\leqslant 1$, $x\in\mathbb{R}$. We have

$$I_1\leqslant h^2\sum_{k=1}^{n-1}k^2(a_k^2+b_k^2),\qquad I_2\leqslant 4\sum_{k=n}^{\infty}(a_k^2+b_k^2),$$

and hence, for

$$\omega^2\Big(\frac{\pi}{n},x\Big)_{L^2}=\sup_{|h|\leqslant\frac{\pi}{n}}\|\Delta_h x\|_{L^2}^2$$

we have

$$\omega^2\Big(\frac{\pi}{n},x\Big)_{L^2}\leqslant\frac{\pi^2}{n^2}\sum_{k=1}^{n-1}k^2(a_k^2+b_k^2)+4E_{n-1}^2(x).$$

Applying the Abel transform, this gives

$$\sum_{k=1}^{n-1} k^2(a_k^2+b_k^2) \leqslant \sum_{k=1}^{n-1}(2k-1)E_{k-1}^2(x) - (n-1)^2E_{n-1}^2(x).$$

Therefore,

$$\omega^2\left(\frac{\pi}{n}, x\right)_{L^2} \leqslant \frac{\pi^2}{n^2}\sum_{k=1}^{n-1}(2k-1)E_{k-1}^2(x) - \frac{\pi^2}{n^2}(n-1)^2E_{n-1}^2(x) + 4E_{n-1}^2(x),$$

whence the result of Theorem 14.2 is apparent. □

Lecture 15

Differentiability and approximation in L^2

15.1. Proof of the second Jackson's theorem in L^2

Let $f \in L^2_{2\pi}$, $f(t) = a_0/2 + \sum_{k=1}^{\infty}(a_k \cos kt + b_k \sin kt)$, where the equality sign means that the left- and right-hand sides coincide as elements of $L^2_{2\pi}$ (here the series converges in $L^2_{2\pi}$ and its sum lies in $L^2_{2\pi}$).

We proceed to give a definition of the derivative that takes into account the structure of $L^2_{2\pi}$. Clearly,

$$\frac{\Delta_h f}{h} \in L^2_{2\pi}, \qquad h > 0,$$

$$\frac{\Delta_h f(t)}{h} = \sum_{k=1}^{\infty} \frac{2 \sin kh/2}{h}(-a_k \sin kt + b_k \cos kt).$$

If there exists a function $\varphi \in L^2_{2\pi}$ such that

$$\lim_{\substack{L_2 \\ h \to 0}} \frac{\Delta_h f}{h} = \varphi,$$

i.e., if

$$\left\| \frac{\Delta_h f}{h} - \varphi \right\|_{L_2} \to 0 \qquad (h \to 0),$$

then f will be said to be L^2-differentiable with the derivative φ, which will be also denoted by f' ($f' = \varphi$). We note that if a function $f \in L^2_{2\pi}$ is absolutely continuous, then the ordinary derivative, which exists almost everywhere, can be taken for the L^2-derivative f', provided it is square integrable.

If the L^2-derivative f' exists, then term-by-term differentiations of the series for the function f is justified, because $a_k(\varphi) = \lim_{h \to 0} a_k(\Delta_h f/h) = kb_k$, and similarly, $b_k(\varphi) = -ka_k$, since the Fourier series for φ can be obtained by formally differentiating the Fourier series for f under the summation sign.

Similarly to f', we define the L^2-derivatives of the second and other orders. So, if there exists the L^2-derivative $f^{(r)}$, then we have in $L^2_{2\pi}$

$$f^{(r)}(t) = \sum_{k=1}^{\infty} k^r \Big\{ a_k \cos\Big(kt + \frac{r\pi}{2}\Big) + b_k \sin\Big(kt + \frac{r\pi}{2}\Big)\Big\},$$

$$\|f^{(r)}\|^2_{L^2_{2\pi}} = \sum_{k=1}^{\infty} k^{2r}(a_k^2 + b_k^2) = \sum_{k=1}^{\infty} k^{2r}\rho_k^2, \qquad \rho_k^2 = a_k^2 + b_k^2.$$

Consider the difference

$$f(t) - s_n(t, f) = \sum_{k=n+1}^{\infty} A_k(t), \qquad A_k(t) = A_k(t, f) = a_k \cos kt + b_k \sin kt.$$

We have

$$E_n^2(f)_{L^2} = \|f - s_n\|^2_{L^2_{2\pi}} = \sum_{k=n+1}^{\infty} \rho_k^2,$$

$$E_n^2(f^{(r)})_{L^2_{2\pi}} = \sum_{k=n+1}^{\infty} k^{2r}\rho_k^2 \geqslant (n+1)^{2r} \sum_{k=n+1}^{\infty} \rho_k^2 = (n+1)^{2r} E_n^2(f)_{L^2_{2\pi}},$$

that is,

$$E_n(f)_{L^2_{2\pi}} \leqslant \frac{1}{(n+1)^r}\, E_n(f^{(r)})_{L^2_{2\pi}} \tag{15.1}$$

or

$$\|f - s_n\|_{L^2_{2\pi}} \leqslant \frac{1}{(n+1)^r}\, \|f^{(r)} - s_n^{(r)}\|_{L^2_{2\pi}}.$$

We rewrite the last inequality in a different form, taking into account that

$$\{\varphi \perp t_n \quad \forall\, t_n \in \tau_n\} \qquad \Longleftrightarrow \qquad s_n(\varphi) \equiv 0.$$

Hence, if $\varphi \perp t_n$, we have the following inequality (known as the Favard or Bohr–Favard inequality)

$$\|\varphi\|_{L^2} \leqslant \frac{1}{(n+1)^r}\|\varphi^{(r)}\|_{L^2}; \tag{15.2}$$

in other words, if the spectrum of a function lies sufficiently far from zero, then the $L^2_{2\pi}$-norm of the function is sufficiently small compared with the L^2-norm of the derivative.

Remark. In other metrics, where the best approximation is normally not delivered by Fourier sums, inequalities (15.1) and (15.2) are not related.

We have earlier proved Bernstein's inequality, from which it follows that

$$\|t_n^{(r)}\|_{L^2} \leqslant n^r \|t_n\|_{L^2}; \tag{15.3}$$

i.e., inequality (15.3) holds for the functions whose spectrum is separated from the infinity.

Inequality (15.3) becomes equality if and only if

$$t_n(t) = A_n(t) = a_n \cos nt + b_n \sin nt;$$

also, the Bohr–Favard's inequality becomes equality if and only if $\varphi = A_{n+1}(t)$.

When speaking about derivatives $f^{(r)}$ of order r, we shall assume for the remainder of this lecture, that they are taken in the L^2-sense or assume that $f^{(r-1)}$ is absolutely continuous and $f^{(r)} \in L^2_{2\pi}$.

By Jackson's inequality,

$$E_n(f^{(r)})_{L^2} \leqslant C\omega\Big(\frac{\pi}{n}, f^{(r)}\Big)_{L^2}.$$

This establishes the second Jackson's inequality, which gives an estimate of the best approximation for an r times differentiable function:

$$E_n(f)_{L^2} \leqslant \frac{C}{(n+1)^r}\,\omega\Big(\frac{\pi}{n}, f^{(r)}\Big)_{L^2}.$$

Using the bound

$$E_n(f^{(r)})_{L^2} \leqslant C_k\omega_k\Big(\frac{\pi}{n}, f^{(r)}\Big)_{L^2},$$

we have the following inequality for r times differentiable functions

$$E_n(f)_{L^2} \leqslant \frac{C_k}{(n+1)^r}\,\omega_k\Big(\frac{\pi}{n}, f^{(r)}\Big)_{L^2}.$$

Suppose that f has the derivative $f^{(r)}$ and that $t_n(t) = \frac{\alpha_0}{2} + \sum_{k=1}^n (\alpha_k \cos kt + \beta_k \sin kt)$ is an approximating polynomial. We estimate the difference $\|f^{(r)} - t_n^{(r)}\|$ via $\|f - t_n\|$ and $f^{(r)}$. We have

$$\|f - t_n\|^2_{L^2} = \frac{(a_0 - \alpha_0)^2}{2} + \sum_{k=1}^{n}\{(a_k - \alpha_k)^2 + (b_k - \beta_k)^2\} + \sum_{k=n+1}^{\infty}(a_k^2 + b_k^2),$$

$$\begin{aligned}\|f^{(r)} - t_n^{(r)}\|^2_{L^2} &= \sum_{k=1}^{n} k^{2r}\{(a_k - \alpha_k)^2 + (b_k - \beta_k)^2\} + \sum_{k=n+1}^{\infty} k^{2r}\rho_k^2 \\ &\leqslant n^{2r}\|f - t_n\|^2_{L^2} + E_n^2(f^{(r)})_{L^2} \leqslant \Big(n^r\|f - t_n\|_{L^2} + E_n(f^{(r)})_{L^2}\Big)^2,\end{aligned}$$

and hence

$$\|f^{(r)} - t_n^{(r)}\|_{L^2} \leqslant n^r\|f - t_n\|_{L^2} + E_n(f^{(r)})_{L^2}. \tag{15.4}$$

Remark. Inequality (15.4) can be extended to other metrics $L^p_{2\pi}$ $(1 < p < \infty)$.

Consider now the case when a polynomial gives a good approximation to a function lying in $L^2_{2\pi}$; by this we mean that it provides approximation which is best in order:

$$\|f - t_n\|_{L^2} \leqslant AE_n(f)_{L^2}. \tag{15.5}$$

Hence, by Favard's inequality,

$$\|f - t_n\|_{L^2} \leqslant \frac{A}{(n+1)^r}E_n(f^{(r)})_{L^2}$$

and so, by (15.4),

$$\|f^{(r)} - t_n^{(r)}\|_{L^2} \leqslant (A+1)^r E_n(f^{(r)})_{L^2}, \tag{15.6}$$

i.e., the derivative of a well approximating polynomial also gives the best possible order of approximation to the derivative of the function.

15.2. Differential properties of functions and properties of approximating polynomials

From (15.6) it follows that if a polynomials from some class provides a good approximation (in the above sense) to an r times differentiable function, then its derivatives of order r are uniformly bounded:

$$\|t_n^{(r)}\|_{L_2} \leqslant C_r \|f^{(r)}\|_{L^2};$$

here the constant $C_r = C_r(A)$ depends only on r and of the constant A occurring in inequality (15.5). For the best approximation polynomials $s_n = s_n(f)$ (in $L^2_{2\pi}$) one can take $C_r = 1$, because in view of the equality $s_n^{(r)}(f) \equiv s_n(f^{(r)})$ and Parseval's formula,

$$\|s_n^{(r)}\|_{L^2} \leqslant \|f^{(r)}\|_{L^2},$$

the bound being uniform in n. In spaces $L^p_{2\pi}$, $1 < p < \infty$, the norms of partial Fourier sums for the best approximation polynomials $t_n^* = t_n^*(f)$ are bounded in $L^p_{2\pi}$, and hence

$$\|t_n^{*(r)}\|_{L^p} \leqslant C_r \|f^{(r)}\|_{L^p}.$$

Let $\omega(\delta, f)_{L^2}$ be the modulus of continuity of f. What can be said about $\omega(\delta, s_n)_{L^2}$? We have

$$\|\Delta_h s_n(f)\|^2_{L^2} = 4 \sum_{k=1}^{n} \sin^2 \frac{kh}{2} (a_k^2 + b_k^2) = \|s_n(\Delta_h f)\|^2_{L^2}.$$

Hence

$$\|\Delta_h s_n\|_{L^2} \leqslant \|\Delta_h f\|_{L^2},$$

and so, for any $n \in \mathbb{N}$,

$$\omega(\delta, s_n)_{L^2} \leqslant \omega(\delta, f)_{L^2};$$

in other words, the best approximation polynomial has the same differential properties, uniformly in n, as the function.

Similarly, for any n, given an r times continuously L^2-differentiable function, we have

$$\omega(\delta, s_n^{(r)})_{L^2} \leqslant \omega(\delta, f^{(r)})_{L^2}.$$

For a polynomial $t_n \in \mathcal{T}_n$, consider $\|\Delta_h t_n\|^2_{L^2}$, $\|\Delta_h^r t_n\|^2_{L^2}$, $\|t_n^{(r)}\|^2_{L^2}$. For the finite difference, the following bound in terms of derivatives is valid:

$$\|\Delta_h^r t_n\|_{L^2} \leqslant |h|^r \|t_n^{(r)}\|_{L^2}. \tag{15.7}$$

Exercise. Prove this estimate. (Hint: Write t_n in terms of the repeated integral of $t_n^{(r)}$.)

Now let us estimate the derivative in terms of finite differences. We have

$$\|\Delta_h^r t_n\|_{L^2}^2 = 2^{2r}\sum_{k=1}^{n}\sin^{2r}\frac{kh}{2}\rho_k^2, \qquad \|t_n^{(r)}\|_{L^2}^2 = \sum_{k=1}^{n} k^{2r}\rho_k^2,$$

where $\rho_k^2 = a_k^2(t_n) + b_k^2(t_n)$. We note that this readily establishes (15.7). Our purpose is to obtain an inequality of the form

$$\|t_n^{(r)}\|_{L^2} \leqslant c_n(h)\|\Delta_h^r t_n\|_{L^2}.$$

To do so it suffices to find a constant $c_n(h)$ such that

$$k^{2r} \leqslant 2^{2r}\sin^{2r}\frac{kh}{2}c_n^2(h) \qquad \forall\ k = 1,\dots,n$$

or

$$\frac{k}{2} \leqslant \widetilde{c_n}(h)\Big|\sin\frac{kh}{2}\Big| \qquad \forall\ k = 1,\dots,n,$$

where

$$\widetilde{c_n}(h) = \max_{k=1,2,\dots,n}\frac{k/2}{|\sin kh/2|}.$$

In order that $\widetilde{c_n}(h)$ be finite the quantity $\left|\sin\frac{kh}{2}\right|$ must be positive for any k. Hence, one must have $\frac{k|h|}{2} < \pi$; i.e., $|h| < \frac{2\pi}{n}$. Since the function $\frac{\sin x}{x}$ is decreasing on the interval $[0,\pi]$, we have, for $\frac{k|h|}{2} < \pi$,

$$\Big|\frac{\sin kh/2}{kh/2}\Big| = \frac{\sin kh/2}{kh/2} \geqslant \frac{\sin nh/2}{nh/2} \qquad \Big(k = 1,2,\dots,n; \quad h \leqslant \frac{2\pi}{n}\Big).$$

Hence, for $\frac{k|h|}{2} < \pi$,

$$\max_{k=1,\dots,n}\frac{k/2}{|\sin kh/2|} = \widetilde{c_n}(h) = \frac{n}{2|\sin nh/2|}.$$

The behaviour of the function $\widetilde{c_n}(h)$, $0 < h < \frac{2\pi}{n}$, is depicted in Fig. 15.1. As a result,

$$\|t_n^{(r)}\|_{L^2} \leqslant \Big(\frac{n}{2\sin n|h|/2}\Big)^r \Big\|\Delta_h^{(r)} t_n\Big\|_{L^2},$$

provided that $|h| < \frac{2\pi}{n}$. This is Stechkin's inequality (see [44], [21, Ch. III], [8, Ch. 4]).

In particular, for $h = \pi/n$,

$$\|t_n^{(r)}\|_{L^2} \leqslant \Big(\frac{n}{2}\Big)^r \Big\|\Delta_h^r t_n\Big\|_{L^2}$$

and since

$$\Big\|\Delta_h^r t_n\Big\|_{L^2} \leqslant 2^r\|t_n\|_{L^2},$$

this is a generalization of Bernstein's inequality of Sec. 14.2.

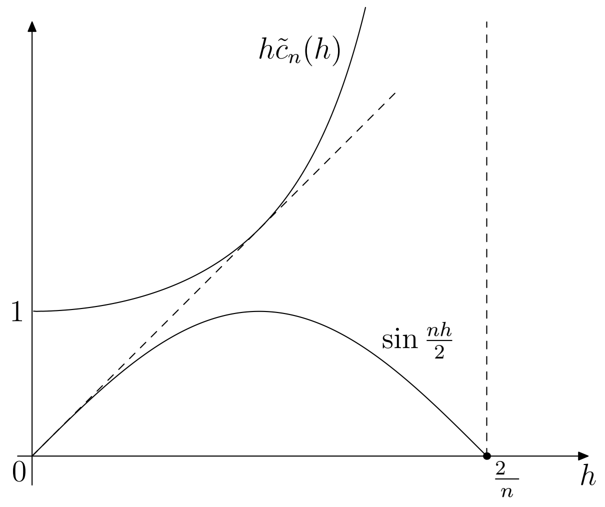

Fig. 15.1 The graph of the function $h\widetilde{c_n}(h)$.

15.3. Differential properties of approximating polynomials

Given a function $f \in L^2_{2\pi}$, suppose that we know its modulus of continuity $\omega(\delta, f)$ in $L^2_{2\pi}$. Let t_n be a well-approximating polynomial in the sense that

$$\|f - t_n\|_{L^2} \leqslant A\omega\Big(\frac{\pi}{n}, f\Big)_{L^2} \tag{15.8}$$

(the fact that such polynomials exist follows by Jackson's inequality).

We shall be concerned with $\omega(\delta, t_n) = \omega(\delta, t_n)_{L^2}$. We have

$$\|\Delta_h t_n\|_{L^2} \leqslant \|\Delta_h f\|_{L^2} + \|\Delta_h(f - t_n)\|_{L^2} \leqslant \omega(h, f) + 2\|f - t_n\|_{L^2} \leqslant \omega(h, f) + 2A\omega\Big(\frac{\pi}{n}, f\Big).$$

Suppose that $h \geqslant \pi/n$. Then

$$\|\Delta_h t_n\|_{L^2} \leqslant (2A + 1)\omega(h, f),$$

because $\omega\big(\frac{\pi}{n}\big) \leqslant \omega(h)$ as $h \geqslant \pi/n$.

If $h = \pi/n$, we have

$$\|\Delta_{\frac{\pi}{n}} t_n\|_{L^2} \leqslant (2A + 1)\omega\Big(\frac{\pi}{n}, f\Big).$$

Let us estimate the norm of the derivative. By Stechkin's inequality,

$$\|t_n'\|_{L^2} \leqslant \frac{n}{2}\|\Delta_{\frac{\pi}{n}} t_n\|_{L^2} \leqslant \frac{1}{2} n\omega\Big(\frac{\pi}{n}, f\Big) = o(n),$$

whereas Bernstein's inequality only gives

$$\|t_n'\|_{L^2} \leqslant n\|t_n\|_{L^2} = O(n).$$

Now let us estimate $\|\Delta_h t_n\|_{L^2}$ for all h, $0 < h < \pi/n$. First we prove the inequality

$$\omega(\lambda\delta, f) \leqslant (\lambda + 1)\omega(\delta, f).$$

If k is integer, then we have

$$\Delta_{kh} f = \sum_{\nu=0}^{k-1} \Delta_h f\Big(t + \nu h + \frac{1-k}{2}h\Big)$$

and so

$$\omega(k\delta, f) \leqslant k\omega(\delta, f).$$

If now $k \leqslant \lambda < k+1$, then

$$\omega(\lambda\delta, f) \leqslant \omega\big((k+1)\delta, f\big) \leqslant (k+1)\omega(\delta, f) \leqslant (\lambda+1)\omega(\delta, f).$$

Hence, for $0 < h < \pi/n$,

$$\|\Delta_h t_n\|_{L^2} \leqslant h\|t_n'\|_{L^2} \leqslant \frac{1}{2} nh\, \omega\Big(\frac{\pi}{n}, f\Big) = \\ = \frac{1}{2} nh\, \omega\Big(\frac{\pi}{nh} \cdot h, f\Big) \leqslant \frac{1}{2} nh\Big(\frac{\pi}{nh} + 1\Big)\, \omega(h, f) \leqslant \pi\omega(h, f).$$

Consequently, for any $h > 0$,

$$\|\Delta_h t_n\|_{L^2} \leqslant C\omega(h, f)_{L^2},$$

and so

$$\omega(h, t_n)_{L^2} \leqslant C\omega(h, f)_{L^2},$$

where $C = \max\{2A + 1, \pi\}$ depends only on the constant A of (15.8) and is independent of both n and h. In other words, differential properties of the well-approximating polynomials are the same, uniformly in n, as the differential properties of the functions.

Similarly, one can prove that

$$\omega_k(h, t_n)_{L^2} \leqslant C_k \omega_k(h, f)_{L^2}, \qquad C_k = C_k(A).$$

Clearly, the converse assertion also holds, inasmuch as

$$\Delta_h f = \Delta_h(f - t_n) + \Delta_n(t_n) \quad \text{and} \quad \|f - t_n\|_{L^2} \leqslant A\, \omega\Big(\frac{\pi}{n}, f\Big).$$

As a result, in order that polynomials $t_n(t)$ $(n \in N)$ with a given modulus of continuity $\omega(\delta)$ (i.e., $\omega(\delta, t_n)_{L^2} \leqslant \omega(\delta)$) give good approximation, uniform in n, to a function f in the sense of estimate (15.8) is it is necessary and sufficient that the function itself have the same modulus of continuity (or more precisely, $\omega(\delta, f)_{L^2} \leqslant C(\omega(\delta))$.

Lecture 16

Jackson's inequality in L^2 with exact constant. Norms of de la Vallée Poussin sums

16.1. Chernykh's theorem

Above we have proved the Jackson's inequality in $L^2_{2\pi}$:

$$E_n(f)_{L^2} \leqslant C\omega\Big(\frac{\pi}{n}, f\Big)_{L^2}.$$

This inequality can be strengthened, because the proof also applies to $E_{n-1}(f)_{L^2}$. Namely,

$$E_{n-1}(f)_{L^2} \leqslant C\omega\Big(\frac{\pi}{n}, f\Big)_{L^2}.$$

We shall be concerned with the following problem.

Problem. What is the best possible constant C for which Jackson's inequality remains valid for all n?

So, we are interested the following extremal problems

$$\sup_{\substack{f\in L^2\\ f\not\equiv\mathrm{const}}} \frac{E_{n-1}(f)_{L^2}}{\omega\Big(\frac{\pi}{n}, f\Big)_{L^2}} = C_n \quad \text{and} \quad \sup_n C_n = C.$$

If $f \equiv \mathrm{const}$, we have $E_{n-1}(f)_{L^2} = 0$, $\omega(\frac{\pi}{n}, f)_{L^2} = 0$, and so the inequality in question holds with any constant.

It appears that C_n is independent of n, $C_n = C = \frac{1}{\sqrt{2}}$, and the final form of Jackson's inequality in $L^2_{2\pi}$ is as follows:

Theorem 16.1 (N. I. Chernykh). *For any $f \in L^2_{2\pi}$, $f \not\equiv \mathrm{const}$, and any $n \in \mathbb{N}$*

$$E_{n-1}(f)_{L^2} < \frac{1}{\sqrt{2}}\,\omega\Big(\frac{\pi}{n}, f\Big)_{L^2}.$$

Here the constant $1/\sqrt{2}$ *is exact*[1]*; i.e., it cannot be decreased, even though the inequality becomes equality only for* $f \equiv \text{const}$.

However, it is possible to build another functional, which is smaller than $\omega(\frac{\pi}{n}, f)$, but is such that the estimate for $E_{n-1}(f)$ in terms of this functional still holds. It is given by Chernykh's theorem, from which Jackson's inequality with exact constant follows as a corollary.

Theorem 16.2 (N. I. Chernykh). *For any* $f \in L^2_{2\pi}$ *and any* $n \in \mathbb{N}$,

$$E^2_{n-1}(f)_{L^2} \leqslant \frac{n}{4} \int_0^{\frac{\pi}{n}} \|\Delta_t f\|^2_{L^2} \sin nt \, dt = J_n; \tag{16.1}$$

here the inequality becomes equality only for functions $f \in L^2_{2\pi}$ *of the form*

$$\alpha_0 + \sum_{k=1}^{\infty} \big(\alpha_k \cos(2k+1)nx + \beta_k \sin(2k+1)nx\big).$$

Likewise $\omega(\frac{\pi}{n}, f)$, J_n is also a structural characterization of a function. Clearly, Jackson's inequality (with constant $C = 1/\sqrt{2}$) is a corollary to Theorem 16.2. Indeed,

$$\omega(h, f) = \sup_{|t| \leqslant h} \|\Delta_t f\|_{L^2} = \sup_{0 \leqslant t \leqslant h} \|\Delta_t f\|_{L^2} \geqslant \|\Delta_t f\|_{L^2}.$$

Hence,

$$J_n \leqslant \frac{n}{4} \omega^2\Big(\frac{\pi}{n}, f\Big) \int_0^{\frac{\pi}{n}} \sin nt \, dt = \frac{1}{2}\, \omega^2\Big(\frac{\pi}{n}, f\Big),$$

i.e., Jackson's inequality can be put in the form

$$E_{n-1}(f)_{L^2} \leqslant \frac{1}{\sqrt{2}}\, \omega\Big(\frac{\pi}{n}, f\Big).$$

We shall find out later when the equality holds.

Proof of Theorem 16.2. As usual, let $\rho_k^2 = a_k^2 + b_k^2$, $k = 1, \ldots$. Then

$$E^2_{n-1}(f)_{L^2} = \sum_{k=n}^{\infty} \rho_k^2 \tag{16.2}$$

and

$$\|\Delta_t f\|^2_{L^2} = 4 \sum_{k=1}^{\infty} \sin^2 \frac{kt}{2} \rho_k^2 = 2 \sum_{k=1}^{\infty} \rho_k^2 (1 - \cos kt)$$

$$\geqslant 2 \sum_{k=n}^{\infty} \rho_k^2 (1 - \cos kt) = 2E^2_{n-1}(f)_{L^2} - 2 \sum_{k=n}^{\infty} \rho_k^2 \cos kt.$$

---

[1] Later it was proved in 1979 by V. V. Arestov and N. I. Chernykh that the factor π/n also cannot be decreased without increasing the multiple $2^{-1/2}$ of ω.

Hence

$$E_{n-1}^2(f)_{L^2} \leqslant \frac{1}{2}\|\Delta_t f\|_{L^2}^2 + \sum_{k=n}^{\infty} \rho_k^2 \cos kt. \tag{16.3}$$

Multiplying both parts of inequality (16.3) by $\sin t$ and integrating from 0 to $\frac{\pi}{n}$ gives

$$E_{n-1}^2(f)_{L^2} \int_0^{\frac{\pi}{n}} \sin nt\, dt = \frac{2}{n} E_{n-1}^2(f)_{L^2}$$

$$\leqslant \frac{1}{2} \int_0^{\frac{\pi}{n}} \|\Delta_t f\|_{L^2}^2 \sin nt\, dt + \int_0^{\frac{\pi}{n}} \sum_{k=n}^{\infty} \rho_k^2 \cos kt \sin nt\, dt.$$

The series in the last integral converging absolutely and uniformly, the change of integration and summation is permissible, and so

$$\frac{2}{n} E_{n-1}^2(f)_{L^2} \leqslant \frac{1}{2} \int_0^{\frac{\pi}{n}} \|\Delta_t f\|_{L^2}^2 \sin nt\, dt + \sum_{k=n}^{\infty} \rho_k^2 \int_0^{\frac{\pi}{n}} \cos kt \sin nt\, dt, \qquad k \geqslant n.$$

Let us evaluate the integrals under the summation sign:

$$\int_0^{\frac{\pi}{n}} \cos kt \sin nt\, dt = \left\{ \begin{array}{ll} 0, & k = n \\ \frac{2n}{n^2-k^2} \cos^2 \frac{k\pi}{2n}, & k > n \end{array} \right\} \leqslant 0 \qquad (k \geqslant n). \tag{16.4}$$

Hence,

$$\sum_{k=n}^{\infty} \rho_k^2 \int_0^{\frac{\pi}{n}} \cos kt \sin nt \leqslant 0 \tag{16.5}$$

and

$$\frac{2}{n} E_{n-1}^2(f)_{L^2} \leqslant \frac{1}{2} \int_0^{\frac{\pi}{n}} \|\Delta_t f\|_{L^2}^2 \sin nt\, dt.$$

This proves inequality (16.1).

Let us find out when inequality (16.1) becomes equality. This happens if the equality holds in all the computations given. Let us check all these places.

1) Since $\int_0^{\frac{\pi}{n}} \cos kt \sin nt\, dt > 0, \qquad 1 \leqslant k < n$, the condition

$$\sum_{k=1}^{n-1} \rho_k^2 \int_0^{\frac{\pi}{n}} \cos kt \sin nt\, dt = 0$$

(see (16.2) and (16.4)) means that $\rho_k = 0$ for all k, $k < n$, except for $k = 0$.

2) Inequalities (16.4) and (16.5) become equalities only if $\rho_k = 0$ for all $k \neq (2m + 1)n$.

Hence the inequality in Theorem 16.2 is equality if and only if the function f has the Fourier series of the form

$$\frac{a_0}{2} + \sum_{k=0}^{\infty} a_{(2k+1)n} \cos(2k+1)nx + b_{(2k+1)n} \sin(2k+1)nx.$$

These functions are periodic with period $\frac{2\pi}{n}$.

Let us now explore the case when the inequalities

$$E_{n-1}^2(f)_{L^2} \leqslant \frac{n}{4}\int_0^{\frac{\pi}{n}} \|\Delta_t f\|_{L^2}^2 \sin nt\,dt \leqslant \frac{1}{2}\,\omega^2\Big(\frac{\pi}{n}, f\Big)$$

reduce to equalities, or what is the same, when Jackson's inequality in Theorem 16.1 becomes equality. The equality holds if and only if, for all t, $0 \leqslant t \leqslant \frac{\pi}{n}$, we have

$$\|\Delta_t f\|_{L^2}^2 = \omega\Big(\frac{\pi}{n}, f\Big).$$

Since $\|\Delta_t f\|_{L^2}^2 \to 0$ as $t \to 0$, it follows that

$$\|\Delta_t f\|_{L^2}^2 = \omega\Big(\frac{\pi}{n}, f\Big) = 0,$$

i.e., $\|\Delta_t f\|_{L^2}^2 = 0$. Hence $\rho_k = 0$ for all $k > 0$, but this means that $f \equiv \text{const}$.

We have already proved that

$$\sup_{\substack{f\in L^2\\ f\not\equiv\text{const}}} \frac{E_{n-1}(f)_{L^2}}{\omega\Big(\frac{\pi}{n}, f\Big)} \leqslant \frac{1}{\sqrt{2}}.$$

We claim that this inequality reduces to equality. Consider the 2π-periodic function defined by $f_\varepsilon(x) = 1$ if $0 \leqslant x \leqslant \varepsilon$ and $f_\varepsilon(x) = 0$ if $\varepsilon < x < 2\pi$ $(\varepsilon < \pi)$. The average value of $f_\varepsilon(x)$ over $(0, 2\pi)$ is equal to $\frac{\varepsilon}{2\pi}$.

To find $E_0^2(f_\varepsilon)$ we have

$$E_0^2(f_\varepsilon) = \frac{1}{\pi}\int_0^{2\pi}\Big(f_\varepsilon(x) - \frac{\varepsilon}{2\pi}\Big)^2 dx = \frac{\varepsilon}{\pi}\Big(1 - \frac{\varepsilon}{2\pi}\Big).$$

We now find an upper bound for $\omega^2(\pi, f_\varepsilon)$. For any t, we have

$$\|\Delta_t f_\varepsilon\|^2 = \frac{1}{\pi}\int_0^{2\pi}[f_\varepsilon(x+t) - f_\varepsilon(x)]^2 dx \leqslant \frac{1}{\pi}\int_0^{2\pi}[f_\varepsilon^2(x+t) + f_\varepsilon^2(x)]dx = \frac{2\varepsilon}{\pi}.$$

Letting $\varepsilon \to 0+$, this gives

$$\frac{E_0^2(f_\varepsilon)}{\omega^2(f_\varepsilon, \pi)} \geqslant \frac{1 - \varepsilon/2\pi}{2} \to \frac{1}{2}. \tag{16.6}$$

This establishes that the constant in Theorem 16.1 is exact for $n = 1$. Given an arbitrary $n \in \mathbb{N}$, $n \geqslant 2$, we consider the periodization of f_ε:

$$f_{\varepsilon,n}(x) = \frac{1}{n}\sum_{k=0}^{n-1} f_\varepsilon\Big(x - \frac{2\pi k}{n}\Big), \qquad \varepsilon < \frac{\pi}{n}.$$

Since this function is $\frac{2\pi}{n}$-periodic, the supports of the terms are disjoint, and $f_{\varepsilon,n} = \frac{1}{n} f_\varepsilon$ on $[0, \frac{2\pi}{n})$, we have

$$\frac{1}{2}\,a_0(f_{\varepsilon,n}) = \frac{1}{2}\,a_0(f_\varepsilon) = \frac{\varepsilon}{2\pi},$$

$$E_{n-1}^2(f_{\varepsilon,n}) = \frac{n}{\pi}\int_0^{2\pi/n}\Big(f_{\varepsilon n}(x) - \frac{\varepsilon}{2\pi}\Big)^2 dx = \frac{\varepsilon}{n\pi}\Big(1 - \frac{\varepsilon n}{2\pi}\Big).$$

A similar estimate can be obtained for the modulus of continuity. For any t,

$$\|\Delta_t f_{\varepsilon,n}\|_{L^2}^2 \leqslant \frac{2\varepsilon}{n\pi},$$

which again leads us to an estimate of form (16.6) for $n \geqslant 2$. □

Remark 16.1. A complete space H of 2π-periodic summable functions f, which is invariant under translations by any $h \in \mathbb{R}$, and in which the norm satisfies the following properties

$$\|f(x+h)\|_H = \|f(x)\|_H, \quad \forall\, h \in (-\infty, \infty),$$
$$\|\Delta_t f\|_H \to 0 \ \text{ as } \ t \to 0 \qquad \forall\, f \in H$$

is called a *uniform space.*

It is readily verified that such a space contains, as a dense subset, the class of trigonometric polynomials (but not necessarily all ones); also, Jackson's inequality

$$E_{n-1}(f)_H \leqslant C\omega\Big(\frac{\pi}{n}, f\Big)_H$$

holds for best approximations from polynomials from the subspace $\mathcal{T}_n \cap H$, provided this subspace is nonempty. The assertion regarding trigonometric polynomials in H follows from the definition of a uniform space.

Problem. Stechkin's conjecture: if $C = \frac{1}{\sqrt{2}}$ and if H is a uniform function space of sufficiently large dimension ($\geqslant 3$), then H is a Hilbert space.

Problem. Does the inequality

$$E_{n-1}^p(f)_{L^p} \leqslant \int_0^{\frac{\pi}{n}} \|\Delta_t f\|_{L^p}^p \varphi_n(t)\, dt, \qquad p > 1$$

hold for some φ_n such that $\int_0^{\frac{\pi}{n}} \varphi_n(t)dt < C$, where $C > 0$ is independent of n?

As a corollary to this inequality we would have the Jackson's inequality for $L_{2\pi}^p$.

For differentiable 2π-periodic functions the following inequality holds:

$$E_{n-1}(f)_{L^2} \leqslant \frac{1}{n^k} E_{n-1}(f^{(k)})_{L^2}.$$

Hence, by the Jackson's inequality in the form of Chernykh,

$$E_{n-1}(f)_{L^2} \leqslant \frac{1}{n^k} E_{n-1}(f^{(k)})_{L^2} \leqslant \frac{1}{\sqrt{2}} \frac{\omega\Big(\frac{\pi}{n}, f^{(k)}\Big)_{L^2}}{n^k}.$$

Consequently, the following result is a corollary to Theorem 16.1.

Corollary. *If a function $f \in C_{2\pi}$ has the absolutely continuous derivative of order $k-1$ and if $f^{(k)} \in L_{2\pi}^2$, then*

$$E_{n-1}(f)_{L^2} \leqslant \frac{1}{\sqrt{2}} \cdot \frac{1}{n^k} \omega\Big(\frac{\pi}{n}, f^{(k)}\Big)_{L^2}.$$

This can be put in a different form:

$$\sup_{\substack{f:\ f^{(k)}\in L^2\\ f^{(k)}\not\equiv\text{const}}} \frac{E_n(f)_{L^2} n^k}{\omega\left(\frac{\pi}{n}, f^{(k)}\right)_{L^2}} \leqslant \frac{1}{\sqrt{2}}.$$

Problem. In the inequality

$$E_{n-1}(f)_{L^2} \leqslant C_k \omega_k\left(\frac{\pi}{n}, f\right)_{L^2}$$

the exact constant is known only for $k = 1$. For the remaining k, the exact constant C_k is not yet calculated.

Problem. For functions of m variables defined on the torus $\mathbb{T}^m$ it is not clear how to formulate the corresponding problems: in this form

$$E_n(f)_{L^2(\mathbb{T}^m)} \leqslant C_m \omega\left(\frac{\pi}{n}, f\right)_{L^2(\mathbb{T}^m)}$$

or most likely in this way

$$E_n(f)_{L^2(\mathbb{T}^m)} \leqslant C_m \omega\left(\frac{\gamma_m}{n}, f\right)_{L^2(\mathbb{T}^m)}?$$

It is required to find[2] the order of the rate of growth for γ_m as $m \to \infty$.

To all these problems the variational methods can be applied, and in fact, they are variational problems.

16.2. Norms of de la Vallée Poussin sums. Nikol'skii's theorem

In this section we shall be concerned with best approximation of 2π-periodic functions (i.e., functions in the space $C = C_{2\pi}$) by trigonometric polynomials with respect to the norm $\|f\|_C = \max\{|f(x)| : \ x \in (-\infty, \infty)\}$.

Given $f \in C = C_{2\pi}$, consider the de la Vallée Poussin sums[3]

$$\frac{1}{p+1} \sum_{k=n-p}^{n} s_k(f) = \sigma_{n,p}(f); \tag{16.7}$$

here p is an integer, $0 \leqslant p \leqslant n$. This is a linear operator from C into C. Let us examine its norm $\|\sigma_{n,p}\|_{C\to C} =: L_{n,p}$.

[2] V. Yudin has proved in 1981 this inequality with the exact constant $1/\sqrt{2}$ for the best approximation by trigonometric polynomials with the spectrum in the disc of radius n with the smallest possible (as was later proved by D. Gorbachev) constant γ_m for $C_m = 1/\sqrt{2}$.

[3] Under the notation of Lecture 6, the de la Vallée Poussin's means, as defined here and in Lecture 17, would be denoted by $\sigma_{n,n-p}$. This has to be taken into account when employing the results of Lecture 6.

Theorem 16.3 (Nikol'skii [32], see also [19]). *Let* $0 \leqslant p \leqslant n$. *Then*

$$L_{n,p} = \frac{4}{\pi^2} \ln \frac{n}{p+1} + O(1).$$

Remark. If $p = 0$, then $\sigma_{n,0} = s_n$, and so we arrive at the standard asymptotic formula for the Lebesgue constants for Fourier sums. If $c\,n \leqslant p \leqslant n$, $0 < c < 1$, then $L_{n,p} = O(1)$. The norms $L_{n,p}$ increase if and only if $p = o(n)$, and in this case, the formula is asymptotical.

Lemma 16.1. *For* $r \geqslant 1$, *we have*

$$\int_0^\pi \frac{|\sin rt|}{t}\, dt = \frac{2}{\pi} \ln r + O(1). \tag{16.8}$$

Proof. We write

$$\int_0^\pi \frac{|\sin rt|}{t}\, dt = \sum_{k=0}^{k_0} \int_{\frac{k\pi}{r}}^{\frac{(k+1)\pi}{r}} \frac{|\sin rt|}{t}\, dt + \int_{\frac{(k_0+1)\pi}{r}}^{\pi} \frac{|\sin rt|}{t}\, dt,$$

where a nonnegative integer k_0 is chosen from the condition $k_0 + 1 \leqslant r < k_0 + 2$, or what is the same, $\frac{(k_0+1)\pi}{r} \leqslant \pi < \frac{(k_0+2)\pi}{r}$. The last integral is bounded for $r \geqslant 1$:

$$\int_{\frac{(k_0+1)\pi}{r}}^{\pi} \frac{|\sin rt|}{t}\, dt = O(1).$$

Changing variables, we obtain

$$\sum_{k=0}^{k_0} \int_{\frac{k\pi}{r}}^{\frac{(k+1)\pi}{r}} \frac{|\sin rt|}{t}\, dt = \sum_{k=0}^{k_0} \int_0^{\frac{\pi}{r}} \frac{|\sin rt|}{t + \frac{k\pi}{r}}\, dt = \int_0^{\frac{\pi}{r}} \sin rt \cdot \sum_{k=0}^{k_0} \frac{1}{t + \frac{k\pi}{r}}\, dt.$$

Here the term with $k = 0$ is independent of r:

$$\int_0^{\frac{\pi}{r}} \frac{\sin rt}{t}\, dt = \int_0^\pi \frac{\sin t}{t}\, dt = O(1).$$

For the remaining terms, we have

$$\frac{r}{\pi} \sum_{k=1}^{k_0} \frac{1}{k+1} \leqslant \sum_{k=1}^{k_0} \frac{1}{t + \frac{k\pi}{r}} \leqslant \frac{r}{\pi} \sum_{k=1}^{k_0} \frac{1}{k}.$$

As a corollary, we obtain

$$\sum_{k=1}^{k_0} \frac{1}{t + \frac{k\pi}{r}} = \frac{r}{\pi} \sum_{k=1}^{k_0} \frac{1}{k} + O(r), \qquad t \in \left[0, \frac{\pi}{r}\right].$$

Since $k_0 = r + O(1)$, it follows that

$$\int_0^{\frac{\pi}{r}} \sin rt \sum_{k=0}^{k_0} \frac{1}{t + \frac{k\pi}{r}}\, dt = \frac{2}{\pi} \sum_{k=1}^{k_0} \frac{1}{k} + O(1) = \frac{2}{\pi} \ln k_0 + O(1) = \frac{2}{\pi} \ln r + O(1).$$

Thus relation (16.8) follows. □

This result will be employed in the next lecture to prove Nikol'skii's theorem 16.3.

Lecture 17

Norms of de la Vallée Poussin's sums (continuation)

17.1. Proof of Nikol'skii's theorem

Our aim is to prove Nikol'skii's theorem 16.3. We use the following representation of the de la Vallée Poussin sums $\sigma_{n,p}$ in terms of Fejer sums σ_n:

$$\sigma_{n,p}(f) = \frac{1}{p+1} \sum_{k=n-p}^{n} s_k(f) = \frac{n+1}{p+1}\sigma_n(f) - \frac{n-p}{p+1}\sigma_{n-p-1}(f);$$

for $p = n$ we assume that $\sigma_{n-p-1}(f) = \sigma_{-1}(f) \equiv 0$. For Fejer sums, we have the following integral representation (see Sec. 5.2)

$$\sigma_n(f) = \sigma_{n,n}(f) = \frac{1}{2\pi(n+1)} \int_0^{2\pi} f(x+t) \left(\frac{\sin \frac{n+1}{2} t}{\sin \frac{t}{2}} \right)^2 dt.$$

As a result, for the de la Vallée Poussin sums, we have the representation

$$\begin{aligned} \sigma_{n,p}(f) &= \frac{1}{2\pi(p+1)} \int_0^{2\pi} f(x+t) \left\{ \left(\frac{\sin \frac{n+1}{2} t}{\sin \frac{t}{2}} \right)^2 - \left(\frac{\sin \frac{n-p}{2} t}{\sin \frac{t}{2}} \right)^2 \right\} dt \\ &= \frac{1}{\pi(p+1)} \int_0^{2\pi} f(x+t) \frac{\sin \frac{2n+1-p}{2} t \cdot \sin \frac{p+1}{2} t}{2 \sin^2 \frac{t}{2}} \, dt, \end{aligned}$$

and so

$$\sigma_{n,p}(f) = \int_0^{2\pi} f(x+t) K(t) \, dt, \quad \text{where} \quad K(t) = \frac{1}{\pi(p+1)} \cdot \frac{\sin \frac{2n+1-p}{2} t \cdot \sin \frac{p+1}{2} t}{2 \sin^2 \frac{t}{2}}.$$

Consequently, the de la Vallée Poussin sum is a convolution operator with continuous kernel K. In the space C, the norm of this operator is calculated as follows:

$$\|\sigma_{n,p}\| = \|\sigma_{n,p}\|_C^C = \int_0^{2\pi} |K(t)| \, dt.$$

Since the kernel K is even, we have

$$\|\sigma_{n,p}\| = 2\int_0^\pi |K(t)|\,dt = \frac{2}{\pi(p+1)}\int_0^\pi \left|\frac{\sin\frac{2n+1-p}{2}\,t\cdot\sin\frac{p+1}{2}\,t}{2\sin^2\frac{t}{2}}\right| dt.$$

As a result, our problem reduces to examining this integral. Let

$$m = \frac{p+1}{2}, \qquad r = \frac{2n+1-p}{p+1}.$$

It follows that $m \geqslant \frac{1}{2}$, $r \geqslant 1$. We note that $rm = \frac{2n+1-p}{2}$. In these notation, we have

$$\|\sigma_{n,p}\| = \frac{1}{\pi m}\int_0^\pi \frac{|\sin rmt\cdot\sin mt|}{2\sin^2\frac{t}{2}}\,dt.$$

Since

$$\frac{1}{\sin^2\frac{t}{2}} - \frac{4}{t^2} = O(1), \qquad 0 < t \leqslant \pi,$$

and since the numerator of the expression under the integral sign is bounded, we have

$$\begin{aligned}\|\sigma_{n,p}\| &= \frac{2}{\pi m}\int_0^\pi \frac{|\sin rmt\cdot\sin mt|}{t^2}\,dt + O\left(\frac{1}{m}\right)\\ &= \frac{2}{\pi m}\int_0^\pi \frac{|\sin rmt\cdot\sin mt|}{t^2}\,dt + O(1).\end{aligned}$$

Changing this variable t to $u = mt$, the last expression takes the form

$$\frac{2}{\pi}\int_0^{\pi m} \frac{|\sin ru\cdot\sin u|}{u^2}\,du + O(1).$$

Since $\int_\pi^\infty u^{-2}\,du < \infty$,

$$\frac{2}{\pi}\int_0^{\pi m} \frac{|\sin ru\cdot\sin u|}{u^2}\,du = \frac{2}{\pi}\int_0^{\pi} \frac{|\sin ru\cdot\sin u|}{u^2}\,du + O(1).$$

Further, since

$$\frac{\sin u}{u^2} - \frac{1}{u} = O(1), \qquad u \in (0,\pi],$$

it follows that

$$\|\sigma_{n,p}\| = \frac{2}{\pi}\int_0^{\pi} \frac{|\sin ru\cdot\sin u|}{u^2}\,du + O(1) = \frac{2}{\pi}\int_0^{\pi} \frac{|\sin ru|}{u}\,du + O(1),$$

and hence, by Lemma 16.1,

$$\|\sigma_{n,p}\| = \frac{4}{\pi^2}\ln r + O(1).$$

Since $0 \leqslant p \leqslant n$, we have $n+1 \leqslant 2n+1-p \leqslant 2n+1$. As a result, $2n+1-p \asymp n$ and

$$\ln(2n+1-p) = \ln n + O(1).$$

Finally, for de la Vallée Poussin sums, we have

$$\|\sigma_{n,p}\| = \frac{4}{\pi^2}\ln\frac{n}{p+1} + O(1). \qquad \square$$

Remark. Let H be a uniform space (see Remark 16.1). In this case, for all $f \in H$,

$$\|\sigma_{n,p}(f)\|_H = \left\| \int_0^{2\pi} f(x+t)K(t)\,dt \right\|_H$$
$$\leqslant \int_0^{2\pi} |K(t)| \|f(x+t)\|_H\,dt = \|f\|_H \int_0^{2\pi} |K(t)|\,dt \qquad \forall\, f \in H,$$

and so

$$\|\sigma_{n,p}\|_H = \|\sigma_{n,p}\|_{H\to H} \leqslant \int_0^{2\pi} |K(t)|\,dt = \|\sigma_{n,p}\|_C.$$

This is why the case of space C is of particular importance.

Corollary. *For* $1 \leqslant q < \infty$,

$$\|\sigma_{n,p}\|_{L_q} = \|\sigma_{n,p}\|_{L_q\to L_q} \leqslant \frac{4}{\pi^2} \ln \frac{n}{p+1} + O(1). \tag{17.1}$$

For $q = 1$ this inequality is in fact an equality, and for $1 < q < \infty$ this bound holds, but is not exact.

17.2. Application of de la Vallée Poussin sums to approximation of continuous functions

For a natural number k, let $C^{(k)} = C^{(k)}_{2\pi}$ be the class of k times continuously differentiable (on the whole real axis) 2π-periodic functions.

Problem. Given a function $f \in C^{(k)}$ and a trigonometric polynomial t_n, the norm of the difference $\|f - t_n\|_C$ is known. How to evaluate $\|f^{(k)} - t_n^{(k)}\|_C$?

The following theorem gives one of the possible answers.

Theorem 17.1 (on the differentiation of approximating polynomials). *There exists a constant* A_k *such that the inequality*

$$\|f^{(k)} - t_n^{(k)}\|_C \leqslant A_k \left\{ n^k \|f - t_n\|_C + E_n(f^{(k)})_C \right\}$$

holds for every function $f \in C^{(k)}_{2\pi}$ *and every trigonometric polynomial* t_n.

Corollary. *Given* $f \in C^{(k)}_{2\pi}$ *suppose that* $n^k \|f - t_n\|_C \to 0$ *as* $n \to \infty$. *Then* $\|f^{(k)} - t_n^{(k)}\|_C \to 0$ *as* $n \to \infty$.

The mere condition $\|f - t_n\|_C = o(n^{-k})$ is not sufficient to ensure that $f \in C^{(k)}$.

Remark. These assertions also hold in the spaces $L^p = L^p_{2\pi}$, $1 \leqslant p < \infty$, because the proof of Theorem 17.1 depends only on the properties of the de la Vallée Poussin sums, property (17.1), and Bernstein's inequality (which in the case of $C_{2\pi}$ can be strengthened as follows).

Lemma 17.1 (generalized Bernstein's inequality (Stechkin [44], see also [8, Ch. 4, § 12])). *Let t_n be a trigonometric polynomial of order n. Then*

$$\|t_n'\|_C \leqslant \frac{n}{2\sin\frac{nh}{2}}\|\Delta_h t_n\|_C, \qquad 0<h<\frac{2\pi}{n},$$

and, as a corollary,

$$\|t_n^{(k)}\|_C \leqslant \Big(\frac{n}{2\sin\frac{nh}{2}}\Big)^k\|\Delta_h^k t_n\|_C, \qquad k\in\mathbb{N}.$$

Putting $h=\pi/n$ in these inequalities and taking into account that $\|\Delta^k f(x)\|_C \leqslant 2^k\|f\|_C$, we arrive at the classical Bernstein's inequality

$$\|t_n'\|_C \leqslant \frac{n}{2}\|\Delta_{\frac{\pi}{n}} t_n\|_C \leqslant n\|t_n\|_C; \tag{17.2}$$

and, as a corollary,

$$\|t_n^{(k)}\|_C \leqslant n^k\|t_n\|_C.$$

Proof of Lemma 17.1. Let x_0 be a point at which the maximum of $|t_n'|$ is attained. For definiteness, we assume that $t_n'(x_0)>0$. Let

$$M_1=\max|t_n'(x)|=t_n'(x_0)$$

and let

$$\varphi_n(x)=\varphi_n(x,t_n)=M_1\cos n(x-x_0)$$

be the comparison function.

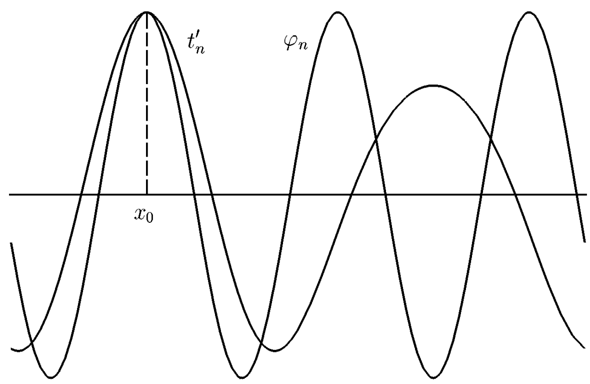

Fig. 17.1

Consider the interval $[x_0-\frac{\pi}{n},x_0+\frac{\pi}{n}]=I$, which is one of the smallest periods of φ_n. Let us prove that (see Fig. 17.1)

$$\varphi_n(x)\leqslant t_n'(x) \qquad \forall\ x\in[x_0-\pi/n,x_0+\pi/n]=I.$$

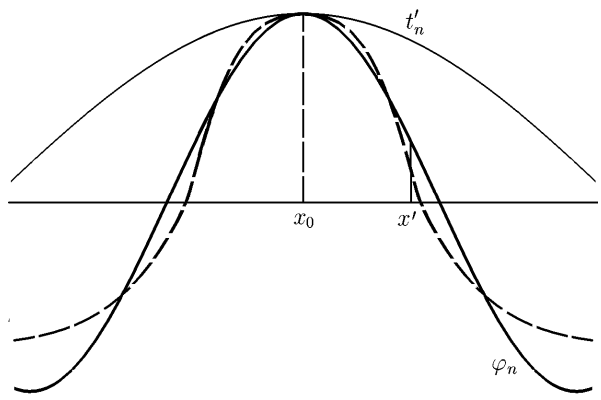

Fig. 17.2

We proceed by *reduction ad absurdum*. Suppose there is a point $x' \in I$ at which $t_n'(x') < \varphi_n(x')$. Then, besides x_0, there is at least one more zero x'' of the function $t_n' - \varphi_n$ in the interval I (see Fig. 17.2).

Let us count the zeros of the difference $t_n' - \varphi_n$. We consider the intervals on which $\cos n(x - x_0)$ changes once from -1 to 1. On each of these intervals, the graph of φ_n intersects the graph of t_n'; also, if these graphs are met at an extremal point of φ_n, then this point is a double root of the difference $t_n' - \varphi_n$. Taking into account multiplicities, the graph of the polynomial $\varphi_n(x) = M_1 \cos n(x - x_0)$ intersects the graph of t_n' at least as many times as $\cos n(x - x_0)$ changes from -1 to 1. Hence, there are $2n - 2$ zeros of the difference $t_n' - \varphi_n$ outside the main interval (provided that this difference does not vanish at the end points of I); moreover, on I there are two more zeros: one at the point x_0 (a double zero) and the other at the point $x'' \in \operatorname{int} I$ (by the assumption). As a result, under the additional assumption, there are at least $2n-2+2+1 = (2n+1)$ zeros (counting multiplicity) of the difference $t_n' - \varphi_n$ on the period. Clearly, if x'' coincides with one of the end points of the interval I or if the function $t_n' - \varphi_n$ vanishes at both its end points, then the total number of its zeros over the period does not decrease. However, $t_n' - \varphi_n$ is a (nonzero) trigonometric polynomial of order n, and so may not have that many zeros (over the period). This gives a contradiction which shows that

$$M_1 \cos n(x - x_0) \leqslant t_n'(x) \qquad \forall\, x \in \left[x_0 - \frac{\pi}{n}, x_0 + \frac{\pi}{n}\right].$$

Integrating this inequality in x over $[x_0 - \frac{h}{2}, x_0 + \frac{h}{2}] \subset I$ for $0 < h < \frac{2\pi}{n}$, we get

$$M_1 \int_{x_0-\frac{h}{2}}^{x_0+\frac{h}{2}} \cos n(x - x_0)\, dx = \frac{M_1}{n} 2 \sin \frac{nh}{2}$$

$$\leqslant \int_{x_0-\frac{h}{2}}^{x_0+\frac{h}{2}} t_n'(x)\, dx = t_n\left(x_0 + \frac{h}{2}\right) - t_n\left(x_0 - \frac{h}{2}\right) \leqslant \|\Delta_h t_n\|_C$$

and, since $M_1 = \|t_n'\|_C$, Lemma 17.1 is proved for $k = 1$. Now suppose that, for $0 < h < \frac{2\pi}{n}$,

$$\|t_n^{(k-1)}\| \leqslant \left(\frac{n}{2 \sin \frac{nh}{2}}\right)^{k-1} \|\Delta_h^{k-1} t_n\|_C.$$

Hence, by what has been just proved,

$$\|t_n^{(k)}\|_C \leqslant \frac{n}{2\sin\frac{nh'}{2}}\|\Delta_{h'}t_n^{(k-1)}\|_C = \left(\frac{n}{2\sin\frac{nh'}{2}}\right)\|(\Delta_{h'}t_n)^{(k-1)}\|_C$$

$$\leqslant \left(\frac{n}{2\sin\frac{nh'}{2}}\right)\left(\frac{n}{2\sin nh}\right)^{k-1}\|\Delta_h^{k-1}(\Delta_{h'}t_n)\|,$$

for $0 < h' < \frac{2\pi}{n}$. Taking $h' = h \in \left(0, \frac{2\pi}{n}\right)$, we have the result stated.

Proof of Theorem 17.1. We represent $f - t_n$ as follows

$$f - t_n = f - \sigma_{n+p,p}(f) - (t_n - \sigma_{n+p,p}(f)).$$

Differentiating k times, this gives

$$f^{(k)} - t_n^{(k)} = f^{(k)} - \sigma_{n+p,p}(f^{(k)}) - (t_n - \sigma_{n+p,p}(f))^{(k)},$$

and hence

$$\|f^{(k)} - t_n^{(k)}\|_C \leqslant \|f^{(k)} - \sigma_{n+p,p}(f^{(k)})\|_C + \|(t_n - \sigma_{n+p,p}(f))^{(k)}\|_C.$$

Applying Lebesgue's inequality (for the de la Vallée Poussin's method), we have

$$\|f^{(k)} - \sigma_{n+p,p}(f^{(k)})\|_C \leqslant (\|\sigma_{n+p,p}\|_C^C + 1)E_n(f^{(k)})_C.$$

Further, applying Bernstein's and Lebesgue's inequality, this gives

$$\begin{aligned}\|(t_n - \sigma_{n+p,p}(f))^{(k)}\|_C &\leqslant (n+p)^k\|t_n - \sigma_{n+p,p}(f)\|_C \\ &\leqslant (n+p)^k(\|f - t_n\|_C + \|f - \sigma_{n+p,p}(f)\|_C) \\ &\leqslant (n+p)^k\left\{\|f - t_n\|_C + (\|\sigma_{n+p,p}\|_C^C + 1)\|f - t_n\|_C\right\} \\ &= (\|\sigma_{n+p,p}\|_C^C + 2)(n+p)^k\|f - t_n\|_C.\end{aligned}$$

Consequently, we have

$$\|f^{(k)} - t_n^{(k)}\|_C \leqslant (\|\sigma_{n+p,p}\|_C + 2)\left\{(n+p)^k\|f - t_n\|_C + E_n(f^{(k)})_C\right\}. \tag{17.3}$$

We put here $p = n$. By Nikol'skii's theorem, the norms $\|\sigma_{2n,n}\|$ are bounded in n. Hence, by (17.3),

$$\|f^{(k)} - t_n^{(k)}\|_C \leqslant A_k\left\{n^k\|f - t_n\|_C + E_n(f^{(k)})_C\right\}. \qquad \square$$

Remark. From the proof we have $A_k = O(2^k)$. One can also prove that $A_k = O(\ln(k+1))$ by suitably choosing the parameter p. Using (17.3) and Nikol'skii's theorem, we obtain

$$\|f^{(k)} - t_n^{(k)}\|_C \leqslant \left\{\frac{4}{\pi^2}\ln\frac{n+p}{p+1} + O(1)\right\}\left(\frac{n+p}{n}\right)^k\left(n^k\|f - t_n\|_C + E_n(f^{(k)})_C\right).$$

By varying $p = p(n)$ we make the quantity

$$\left\{\frac{4}{\pi^2}\ln\frac{n+p}{p+1} + O(1)\right\}\left(\frac{n+p}{n}\right)^k \tag{17.4}$$

as small as possible. Consider three cases.

The case $k \leqslant n$. Let p be an integer such that $\frac{n}{k} - 1 \leqslant p \leqslant \frac{n}{k}$. Then

$$\left(\frac{n+p}{n}\right)^k \leqslant \left(1 + \frac{n/k}{n}\right)^k = \left(1 + \frac{1}{k}\right)^k \leqslant e,$$

$$\frac{n+p}{p+1} = \frac{n-1}{p+1} + 1 \leqslant k\frac{n-1}{n} + 1 < k+1.$$

For this specification of the parameter p quantity (17.4) is of order $O(\ln(k+1))$, which gives $A_k = O(\ln(k+1))$.

The case $k \geqslant n$. We set $p = 0$. In this case we have, for quantity (17.4),

$$\frac{4}{\pi^2}\ln(n+1) + O(1) \leqslant \frac{4}{\pi^2}\ln(k+1) + O(1) = O(\ln(k+1)).$$

Consequently, in either case,

$$A_k = O(\ln(k+1)).$$

Problem. Prove that this order of A_k is sharp.

Remark. We have proved Bernstein's inequality (17.2) in the space $C_{2\pi}$. Furthermore, it also holds in any uniform space (see [39], and also [8]). Hence Theorem 17.1 on the differentiation of approximating polynomials holds not only in C but also in any uniform space (see the remark in Section 17.1).

Lecture 18

Approximation of functions represented by integral transforms

18.1. Approximation of functions in the space $L_{2\pi}$

We first consider the problem of best approximation of functions in the space $L = L_{2\pi}$ equipped with the norm $\|f\|_L = \frac{1}{\pi}\int_0^{2\pi} |f(t)|dt$, by trigonometric polynomials

$$t_{n-1}(x) = \frac{\alpha_0}{2} + \sum_{k=1}^{n-1}(\alpha_k \cos kx + \beta_k \sin kx)$$

of order $n-1$, $n \geqslant 1$.

Theorem 18.1. *Let $f \in L_{2\pi}$. Then the following assertions hold:*

1) *Let t^*_{n-1} be a trigonometric polynomial and let $R = f - t^*_{n-1}$. Suppose that the function* $\operatorname{sign} R$ *is orthogonal to the subspace $\mathcal{T}_{n-1}$; i.e.,*

$$\operatorname{sign} R \perp t_{n-1} \qquad \forall\, t_{n-1} \in \mathcal{T}_{n-1}, \tag{18.1}$$

then t^*_{n-1} is a polynomial of best approximation to f in $L_{2\pi}$.

2) *If t^*_{n-1} is a polynomial of best approximation to the function f and if the difference $f - t^*_{n-1}$ is different from zero a.e., then the orthogonality condition* (18.1) *holds.*

Proof. Suppose that orthogonality condition (18.1) is satisfied for a trigonometric polynomial t^*_{n-1}; i.e., the function $\operatorname{sign} R$ is orthogonal to any polynomial of order $n-1$. Then, for any polynomial $t_{n-1} \in \mathcal{T}_{n-1}$,

$$\begin{aligned}\|f - t^*_{n-1}\|_L &= \frac{1}{\pi}\int_0^{2\pi} |f(x) - t^*_{n-1}(x)|\,dx = \frac{1}{\pi}\int_0^{2\pi} (f(x) - t^*_{n-1}(x))\,\operatorname{sign} R(x)\,dx \\ &= \frac{1}{\pi}\int_0^{2\pi} \big((f(x) - t_{n-1}(x)) + (t_{n-1}(x) - t^*_{n-1}(x))\big)\,\operatorname{sign} R(x)\,dx \\ &= \frac{1}{\pi}\int_0^{2\pi} (f(x) - t_{n-1}(x))\,\operatorname{sign} R(x)\,dx \leqslant \frac{1}{\pi}\int_0^{2\pi} |f(x) - t_{n-1}(x)|\,dx = \|f - t_{n-1}\|_L.\end{aligned}$$

Hence t^*_{n-1} is a polynomial of best approximation from $\mathcal{T}_{n-1}$ to the function f in $L_{2\pi}$.

Now let t^*_{n-1} be a polynomial of best approximation to f. Suppose that $t^*_{n-1} \neq f$ a.e. Given an arbitrary polynomial t_{n-1} of order $n-1$, consider the function

$$\Phi(\lambda) = \|f - (t^*_{n-1} - \lambda t_{n-1})\|_L = \frac{1}{\pi}\int_0^{2\pi} |f(x) - t^*_{n-1}(x) + \lambda t_{n-1}(x))|\,dx$$

of a real variable λ. We claim that the function Φ is differentiable at the point $\lambda = 0$ and calculate the derivative $\Phi'(0)$.

For real numbers $a \neq 0$ and b, the function $\varphi(\lambda) = |a + \lambda b|$ of variable λ is differentiable at the point $\lambda = 0$. Also, $\varphi'(0) = b\ \operatorname{sign} a$ and

$$\left|\frac{|a+\lambda b| - |a|}{\lambda}\right| \leqslant |b|, \qquad \lambda \neq 0.$$

Applying Lebesgue's Dominated Convergence Theorem, it is easily seen that the function Φ is differentiable at $\lambda = 0$, and

$$\Phi'(0) = \frac{1}{\pi}\int_0^{2\pi} t_{n-1}(x)\ \operatorname{sign}(f(x) - t^*_{n-1}(x))\,dx.$$

Since t^*_{n-1} is an extremal polynomial, the function Φ attains its minimum at $\lambda = 0$. Hence $\Phi'(0) = 0$, and so orthogonality property (18.1) holds. □

Note that if condition (18.1) holds, then

$$E_{n-1}(f)_L = \frac{1}{\pi}\int_0^{2\pi} (f(x) - t^*_{n-1}(x))\ \operatorname{sign} R(x)\,dx = \frac{1}{\pi}\int_0^{2\pi} f(x)\ \operatorname{sign} R(x)\,dx$$

and, finally,

$$E_{n-1}(f)_L = \frac{1}{\pi}\int_0^{2\pi} f(x)h^*(x)\,dx, \tag{18.2}$$

where $h^* = \operatorname{sign} R$. In this setting, the function h^* is such that $\|h^*\|_{L^\infty} \leqslant 1$ and $h^* \perp t_{n-1}$ for any polynomial t_{n-1}.

Theorem 18.2. *Let $f \in L_{2\pi}$. Then*

$$E_{n-1}(f)_L \geqslant \frac{1}{\pi}\int_0^{2\pi} f(x)h(x)\,dx \tag{18.3}$$

for any function $h \in L^\infty_{2\pi}$ such that

$$\|h\|_{L^\infty} \leqslant 1 \quad \text{and} \quad h \perp t_{n-1} \qquad \forall\ t_{n-1} \in \mathcal{T}_{n-1}.$$

Proof. Let t^*_{n-1} be a polynomial of best approximation to f in L. Since h is orthogonal to any polynomial of order $n-1$,

$$\frac{1}{\pi}\int_0^{2\pi} f(x)h(x)\,dx = \frac{1}{\pi}\int_0^{2\pi} (f(x) - t^*_{n-1}(x))h(x)\,dx$$
$$\leqslant \frac{1}{\pi}\int_0^{2\pi} |f(x) - t^*_{n-1}(x)|\,dx = E_{n-1}(f)_L.$$ □

Corollary 1. *Let t^*_{n-1} be a polynomial of best $L_{2\pi}$-approximation to a function f. Suppose that the $h^* = \operatorname{sign}(f - t^*_{n-1})$ satisfies orthogonality condition* (18.1). *Then*

$$E_{n-1}(f)_L = \frac{1}{\pi}\int_0^{2\pi} f(x)h^*(x)\,dx$$

$$= \max\left\{\frac{1}{\pi}\int_0^{2\pi} f(x)h(x)\,dx : \ h \in L^\infty_{2\pi}, \ \|h\|_{L^\infty} \leqslant 1; \ h \perp t_{n-1} \ \forall\, t_{n-1} \in \mathcal{T}_{n-1}\right\}.$$

Corollary 2. *Let a polynomial t^*_{n-1} be such that the difference $f - t^*_{n-1} \neq 0$ a.e. and that the function $h^* = \operatorname{sign}(f - t^*_{n-1})$ is orthogonal to any polynomial of order $n-1$. Then t^*_{n-1} is a polynomial of best $L_{2\pi}$-approximation to f, and inequality* (18.3) *becomes equality if and only if $h = h^*$.*

Considering any function $h \in L^\infty_{2\pi}$ such that

$$\|h\|_{L^\infty} \leqslant 1 \quad \text{and} \quad h \perp t_{n-1} \qquad \forall\, t_{n-1} \in \mathcal{T}_{n-1}$$

and using (18.3) we obtain a lower bound for $E_{n-1}(f)_L$. We carefully pick a function h. Suppose that h has period $\omega = \frac{2\pi}{n}$ and that $\int_{-\pi}^{\pi} h(x)\,dx = 0$. Then the Fourier series of such a function is as follows

$$h(x) \sim \sum_{k=1}^{\infty} \alpha_{nk}\cos nkx + \beta_{nk}\sin nkx;$$

consequently, the only candidates for nonzero coefficients are those whose indices multiply n. In particular,

$$h \perp t_{n-1} \qquad \forall\, t_{n-1} \in \mathcal{T}_{n-1}.$$

Hence, as h in (18.3) we may take any such function (satisfying, in addition, the condition $\|h\|_{L^\infty} \leqslant 1$).

Let $h(x) = \operatorname{sign}\sin(nx + \alpha)$. This function satisfies all the above properties: its period ω is $\frac{2\pi}{n}$, $\frac{1}{\pi}\int_{-\pi}^{\pi} h(x)\,dx = 0$, and $\|h\|_{L^\infty} \leqslant 1$. Hence,

$$E_{n-1}(f)_L \geqslant \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\,\operatorname{sign}\sin(nx+\alpha)\,dx \qquad \forall\, \alpha \in (-\infty, \infty). \tag{18.4}$$

Consider the cases $\alpha = 0$ and $\alpha = \frac{\pi}{2}$. The expansion

$$\operatorname{sign}\sin x = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{\sin(2k+1)x}{2k+1}$$

is well known. Hence,

$$\operatorname{sign}\sin nx = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{\sin(2k+1)nx}{2k+1}. \tag{18.5$'$}$$

Similarly,

$$\operatorname{sign}\cos x = \operatorname{sign}\sin\left(x + \frac{\pi}{2}\right) = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k\cos(2k+1)x}{2k+1},$$

$$\operatorname{sign}\cos nx = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k\cos(2k+1)nx}{2k+1}. \tag{18.5''}$$

Using (18.4), we have, for any function $f \in L$ (respectively, for $\alpha = 0$ and $\alpha = \frac{\pi}{2}$),

$$E_{n-1}(f)_L \geqslant \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{b_{(2k+1)n}}{2k+1}, \tag{18.6}$$

$$E_{n-1}(f)_L \geqslant \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k a_{(2k+1)n}}{2k+1}; \tag{18.7}$$

here $a_{(2k+1)n}$ and $b_{(2k+1)n}$ the corresponding Fourier coefficients of f. Note that if $\alpha = \pi$, then $h(x) = -\operatorname{sign}\sin nx$, and if $\alpha = \frac{3}{2}\pi$, then $h(x) = -\operatorname{sign}\cos nx$. Hence, the similar estimates hold:

$$E_{n-1}(f)_L \geqslant -\frac{4}{\pi}\sum_{k=0}^{\infty}\frac{b_{(2k+1)n}}{2k+1}, \tag{18.6'}$$

$$E_{n-1}(f)_L \geqslant -\frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k a_{(2k+1)n}}{2k+1}. \tag{18.7'}$$

The question arises: When the last four inequalities become equalities? Let t^*_{n-1} be a polynomial of best approximation to the function f. We have $f(x) - t^*_{n-1}(x) \neq 0$ a.e. on $x \in (-\pi, \pi)$. The equality in (18.6) is possible only if (a.e.)

$$\operatorname{sign}(f - t^*_{n-1}) = \operatorname{sign}\sin nx; \tag{18.8}$$

the equality in (18.7) is possible only if

$$\operatorname{sign}(f - t^*_{n-1}) = \operatorname{sign}\cos nx. \tag{18.9}$$

Similarly, (18.6′) and (18.7′) become equalities if, respectively,

$$\operatorname{sign}(f - t^*_{n-1}) = -\operatorname{sign}\sin nx, \tag{18.8'}$$

$$\operatorname{sign}(f - t^*_{n-1}) = -\operatorname{sign}\cos nx. \tag{18.9'}$$

Assume that the function f is continuous on the interval $(-\pi, \pi)$. Then it follows by (18.9) and (18.9′) that the difference $f - t^*_{n-1}$ changes sign only at those points at which $\cos nx$ vanishes; hence t^*_{n-1} interpolates f at the zeros of $\cos nx$. Similarly, it follows by (18.8) and (18.8′) that t^*_{n-1} interpolates f at the zeros of $\sin nx$. These interpolation conditions are necessary conditions (for a continuous function) for inequalities (18.7), (18.7′), (18.6), (18.6′) respectively, to become equalities. Moreover, sometimes these conditions are also sufficient. Indeed, if a polynomial t_{n-1} interpolates the function f (for example, only at the zeros of $\cos nx$) and if the difference $f - t_{n-1}$ changes sign only at these zeros and nowhere else, then (18.7) (or (18.7′), respectively), becomes equality, and hence t^*_{n-1} is a polynomial of best $L_{2\pi}$-approximation to f. In these cases,

$$E_{n-1}(f)_L = \frac{4}{\pi}\left|\sum_{k=0}^{\infty}\frac{b_{(2k+1)n}}{2k+1}\right|, \tag{18.6''}$$

$$E_{n-1}(f)_L = \frac{4}{\pi}\left|\sum_{k=0}^{\infty}\frac{(-1)^k a_{(2k+1)n}}{2k+1}\right|. \tag{18.7''}$$

Hence, we have at our disposal a way of obtaining a polynomial of best $L_{2\pi}$-approximation to $f \in C_{2\pi}$. Namely, we take $h(x) = \operatorname{sign}\sin nx$ or $h(x) = \operatorname{sign}\cos nx$ according to whether if f is odd or even (in other cases, we consider the function $\operatorname{sign}\sin(nx+\alpha)$ and try to choose the parameter α appropriately). Hence we construct a polynomial that interpolates f at the zeros of $\sin nx$ or $\cos nx$, respectively. If the signs of the difference satisfy corresponding conditions (18.8), (18.8′), (18.9) or (18.9′), then the construction of a polynomial of best approximation is complete. Hence the problem of constructing a polynomial of best $L_{2\pi}$-approximation reduces to testing the sign of $f - t^*_{n-1}$. If the sign-test-operation is skipped, then (18.7), (18.7′) and (18.6) or (18.6′) give a lower bound for $E_{n-1}(f)_L$.

18.2. Approximation of classes of functions in $C_{2\pi}$

A function $K \in L_{2\pi}$ will be referred to as an *integrable kernel*. Consider the class $\mathfrak{M} = \mathfrak{M}_K$ of functions which are represented as integral transforms with this kernel — these are functions of the form

$$f(x) = c + \frac{1}{\pi}\int_0^{2\pi} K(t)\varphi(x+t)\,dt, \tag{18.10}$$

where φ is an arbitrary 2π-periodic function in the space $L^\infty = L^\infty_{2\pi}$, $\|\varphi\|_{L^\infty} \leqslant 1$, and $c = c(f)$ is a real constant. Any function from the class $\mathfrak{M}$ is continuous and 2π-periodic. The quantity

$$\sup_{f\in\mathfrak{M}}\ \min_{t_{n-1}\in\mathcal{T}_{n-1}} \|f - t_{n-1}\|_C = E_{n-1}(\mathfrak{M}_K)_C$$

is the best approximation of the class $\mathfrak{M}$ in $C_{2\pi}$ by trigonometric polynomials of order $n-1$.

For any trigonometric polynomial $\widetilde{t}_{n-1}$, the function

$$t_{n-1}(x) = c + \frac{1}{\pi}\int_0^{2\pi} \widetilde{t}_{n-1}(t)\varphi(x+t)\,dt$$

is also a trigonometric polynomial of order $n-1$. For a suitable $c = c(f)$, we have

$$|f(x) - t_{n-1}(x)| = \left|\frac{1}{\pi}\int_0^{2\pi}\{K(t) - \widetilde{t}_{n-1}(t)\}\varphi(x+t)\,dt\right| \leqslant \frac{1}{\pi}\int_0^{2\pi}|K(t) - \widetilde{t}_{n-1}(t)|\,dt,$$

and hence

$$\|f - t_{n-1}\|_C \leqslant \frac{1}{\pi}\int_0^{2\pi}|K - \widetilde{t}_{n-1}|\,dt \qquad \forall\, \widetilde{t}_{n-1} \in \mathcal{T}_{n-1}.$$

Therefore,

$$E_{n-1}(f)_C \leqslant E_{n-1}(K)_L \qquad \forall\, f \in \mathfrak{M}_K.$$

Consequently, the following inequality holds:

$$\sup_{f\in\mathfrak{M}_K} E_{n-1}(f)_C \leqslant E_{n-1}(K)_L. \tag{18.11}$$

Here the inequality becomes equality holds for a wide class of kernels K; below we shall verify this in an important special case.

The above argument applies to any uniform space H of 2π-periodic functions. In this case, we again have

$$\|f - t_{n-1}\|_H = \left\| \frac{1}{\pi}\int_0^{2\pi} \{K(t) - \widetilde{t}_{n-1}(t)\}\varphi(x+t)\,dt \right\|_H$$

$$\leqslant \frac{1}{\pi}\int_0^{2\pi} |K(t) - \widetilde{t}_{n-1}(t)|\;\|\varphi(\cdot+t)\|_H\,dt = \frac{1}{\pi}\int_0^{2\pi} |K(t) - \widetilde{t}_{n-1}(t)|\,dt\cdot\|\varphi\|_H,$$

and hence

$$E_{n-1}(f)_H \leqslant E_{n-1}(K)_L\cdot\|\varphi\|_H.$$

Let $\mathfrak{M}_{K,H}$ be the class of functions in H of the form (18.10) with $\varphi\in H$ and $\|\varphi\|_H\leqslant 1$. Тогда получим

$$\sup_{f\in\mathfrak{M}_{K,H}} E_{n-1}(f)_H \leqslant E_{n-1}(K)_L.$$

In general the inequality is strict.

18.3. Approximation in the mean of Bernoulli kernels by trigonometric polynomials

Now we consider the class $W_1^{(r)}$, $r\geqslant 1$, of functions $f\in C_{2\pi}$ whose derivative $f^{(r-1)}$ of order $r-1$ is a 1-Lipschitz function; that is,

$$|f^{(r-1)}(x') - f^{(r-1)}(x'')| \leqslant |x'-x''| \qquad \forall\, x', x''\in[0,2\pi].$$

A function $f\in W_1^{(r)}$ has the derivative $f^{(r)}$ of order r a.e.; also $|f^{(r)}(x)|\leqslant 1$ a.e., and

$$\int_0^{2\pi} f^{(r)}(x)\,dx = 0,$$

that is, the mean value of the derivative $f^{(r)}$ over the period is zero. A function $f\in W_1^{(r)}$ has the following integral representation

$$f(x) = \frac{a_0}{2} + \frac{1}{\pi}\int_{-\pi}^{\pi} K_r(t) f^{(r)}(x+t)\,dt; \tag{18.12}$$

here $\frac{a_0}{2} = \frac{1}{2\pi}\int_0^{2\pi} f(x)\,dx$ is the mean value of f, and

$$K_r(t) = \sum_{n=1}^{\infty} \frac{\cos\left(nt + \frac{r\pi}{2}\right)}{n^r}. \tag{18.13}$$

The function K_r is known as the *Bernoulli kernel.*

For $r = 1$, we have

$$K_1(t) = -\sum_{n=1}^{\infty} \frac{\sin nt}{n} = \frac{t-\pi}{2}, \qquad t \in (0, 2\pi).$$

For $r > 1$, the kernel K_r can be found by integrating the kernel K_{r-1}, when a constant is chosen in such a way that the mean value is zero:

$$\frac{1}{2\pi} \int_0^{2\pi} K_r(t)dt = 0.$$

In Section 18.2 we have proved, for a function $f \in W_1^{(r)}$,

$$E_{n-1}(f)_C \leqslant E_{n-1}(K_r)_L.$$

Let us find the best approximation $E_{n-1}(K_r)_L$. Instead of K_r, it is convenient to work with the function

$$K_r(t+\pi) = \sum_{n=1}^{\infty} (-1)^n \frac{\cos\left(nt + \frac{r\pi}{2}\right)}{n^r},$$

which we again denote by K_r. Due to 2π-periodicity, this does not change the best approximation. In this notation we have

$$K_1(t) = \frac{t}{2}, \qquad t \in (-\pi, \pi).$$

The function K_r is odd or even according to whether r is odd or even.

For the kernels K_r, we construct interpolation polynomials U_{n-1} with the nodes at zeros of $\sin nx$, of $\cos nx$ respectively, according to whether the number r is odd, even respectively. Now our task is to prove that the difference $R = K_r - U_{n-1}$ changes sign only at the interpolation points. For $r = 1$ the function K_1 has discontinuities at the points $(2k+1)\pi$; these points should be looked at as points of changes of sign of the difference R. For $r = 2s+1$, $s > 0$, the points $\pm\pi$ are also additional interpolation points.

Lemma 18.1. *For the difference $R = K_r - U_{n-1}$ of the kernel K_r and the interpolation polynomial U_{n-1}, the following holds:*

$$\operatorname{sign}(K_r(t) - U_{n-1}(t)) = \pm \operatorname{sign} \sin nt, \quad t \in (-\pi, \pi), \quad \text{if } r \text{ is odd;}$$
$$\operatorname{sign}(K_r(t) - U_{n-1}(t)) = \pm \operatorname{sign} \cos nt, \quad t \in (-\pi, \pi), \quad \text{if } r \text{ is even.}$$

Proof. The difference $R = K_r - U_{n-1}$ on $(-\pi, \pi)$ has $(2n-1)$ zeros for odd r and $2n$ zeros for even r. It suffices to verify that all these zeros are simple and that there are no other zeros. To show this, it suffices in turn to show that, for odd r, the difference $R = K_r - V_{n-1}$ has on $(-\pi, \pi)$ at most $(2n-1)$ zeros (counting multiplicities) for an arbitrary odd trigonometric polynomial V_{n-1} of order $n-1$, and for odd r and even polynomial V_{n-1}, the difference $R = K_r - V_{n-1}$ has on $(-\pi, \pi)$ at most $2n$ zeros (counting multiplicities). We proceed by induction on $r \geqslant 1$.

1. Let $r = 1$. We claim that the difference $R = K_1 - V_{n-1}$ of the kernel K_1 (which is odd in this case) and an odd trigonometric polynomial V_{n-1} has at most $(2n-1)$ zeros on $(-\pi, \pi)$ (counting multiplicities). Assume the contrary. The derivative of the difference

$$R'(t) = \frac{1}{2} - V'_{n-1}(t) = \frac{1}{2} - \sum_{k=1}^{n-1} \alpha_k \cos kt, \qquad t \in (-\pi, \pi)$$

is a (nonzero) polynomial in cosines of order $n-1$. Such a polynomial has at most $(2n-2)$ zeros on $(-\pi, \pi)$. If on $(-\pi, \pi)$ the difference R would have at least $2n$ zeros, then by Rolle's theorem, the derivative R' would have $(2n-1)$ zeros, which is impossible. This contradiction shows that $R = K_r - V_{n-1}$ has at most $(2n-1)$ zeros on $(-\pi, \pi)$.

2. Let $r = 2$. Both the kernel K_2 and the polynomial V_{n-1} are even. We claim that the difference $R = K_2 - V_{n-1}$ has at most $2n$ zeros on $(-\pi, \pi)$. We have $R' = K'_2(t) - V'_{n-1}$, where V'_{n-1} is an odd polynomial of order $n-1$. By the above, this difference has at most $(2n-1)$ zeros on $(-\pi, \pi)$. Hence, R has at most $2n$ zeros.

3. We claim that if $r > 1$ then the required property of zeros is satisfied for $r-1$, then it also holds for r. Let $r > 1$ be odd, $r = 2s+1$, $s \geqslant 1$. Then both K_{2s+1} and V_{n-1} are odd, and hence both K_{2s+1} and V_{n-1} vanish at the points $\pm\pi$. But then the difference $R = K_{2s+1} - V_{n-1}$ also vanishes at these points, $R(\pm\pi) = 0$. We claim that the number of zeros m on $(-\pi, \pi)$ of the difference R is at most $2n-1$. Since $R(\pm\pi) = 0$, it follows by Rolle's theorem, that the derivative R' has $m+1$ zeros on $(-\pi, \pi)$. By the inductive assumption, $R' = K_{r-1} - V'_{n-1}$ has at most $2n$ zeros on $(-\pi, \pi)$. Hence $R = K_{2s+1} - V_{n-1}$ has at most $2n-1$ zeros.

If $r = 2s$ is even, then $R' = K_{2s-1} - V'_{n-1}$ is an odd function, which by the above has at most $(2n-1)$ zeros. By Rolle's theorem, R has at most $2n$ zeros. □

Now we can write down a polynomial of best approximation in the mean to the Bernoulli kernel (18.13) by trigonometric polynomials.

Theorem 18.3. *A sufficient condition for a polynomial U_{n-1} of order $n-1$ to be a polynomial of best L-approximation to K_r is that U_{n-1} interpolates the kernel K_r at the zeros of* $\sin nx$ *or* $\cos nx$ *according to whether r is odd or even. Moreover,*

$$E_{n-1}(f)_L = \|K_r - U_{n-1}\|_L = \frac{M_r}{n^r}, \tag{18.14}$$

where

$$M_r = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^{(r+1)k}}{(2k+1)^{r+1}}. \tag{18.15}$$

Proof. The fact that a polynomial U_{n-1} is extremal was proved above. Further, applying (18.6″), (18.7″), (18.5′), and (18.5″), and using expansion (18.13), it follows that

$$E_{n-1}(f)_L = \left| \frac{1}{\pi} \int_{-\pi}^{\pi} K_r(t) \operatorname{sign} \sin nt \, dt \right| = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{1}{n^r (2k+1)^{r+1}}, \qquad r \text{ is odd},$$

$$E_{n-1}(f)_L = \left| \frac{1}{\pi} \int_{-\pi}^{\pi} K_r(t) \operatorname{sign} \cos nt \, dt \right| = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{n^r (2k+1)^{r+1}}, \qquad r \text{ is even}. \quad □$$

Lecture 19

Favard's theorem and its applications

19.1. Favard's Theorem on approximation of differentiable functions

In the last lecture we discussed the problem of approximation by trigonometric polynomials of functions which are represented by integral transforms with kernels $K \in L_{2\pi}$. In this connection, we were interested in finding the best approximation

$$E_{n-1}(K)_L = \min_{t_{n-1}} \|K - t_{n-1}\|_L = \|K - t^*_{n-1}\|_L$$

of a kernel K by trigonometric polynomials in $L_{2\pi}$. It frequently happens that a polynomial of best approximation is the one which interpolates the kernel K at equidistant nodes spaced at $\frac{\pi}{n}$. In particular, this holds for the Bernoulli kernel K_r, which gives an integral representation of r times differentiable functions—in particular, of functions of the class $W_1^{(r)}$:

$$f(x) = \frac{a_0}{2} + \frac{1}{\pi}\int_0^{2\pi} K_r(t)\varphi(x+t)\,dt;$$

here $\varphi = f^{(r)}$. In this case, as we have shown in Theorem 18.3,

$$E_{n-1}(K)_L = \|K_r - t^*_{n-1}\|_L = \frac{M_r}{n^r},$$

where M_r are constants defined by (18.15).

Our purposes in this lecture are as follows:

1) given a function $f \in C_{2\pi}^{(r)}$, to evaluate the best approximation $E_{n-1}(f)_C$ by means of trigonometric polynomials of order $n-1$;

2) to calculate

$$\sup_{W_1^{(r)}} E_{n-1}(f)_C = E_{n-1}(W_1^{(r)})_C.$$

Clearly, if the latter problem is solved, then, for any function $f \in C^{(r)}$,

$$E_{n-1}(f)_C \leqslant E_{n-1}(W_1^{(r)})_C \cdot \|f^{(r)}\|_C.$$

Let t^*_{n-1} be a polynomial of best approximation in the mean to kernel K_r. Then

$$\int_0^{2\pi} \{K_r(\theta) - t^*_{n-1}(\theta)\}\varphi(x+\theta)\,d\theta = f(x) - t_{n-1}(x), \tag{19.1}$$

where t_{n-1} is some trigonometric polynomial of order $n-1$. Hence,

$$E_{n-1}(f)_C \leqslant \|f - t_{n-1}\|_C \leqslant \int_0^{2\pi} |K_r(t) - t^*_{n-1}(t)|\,dt \cdot \|f^{(r)}\|_C, \tag{19.2}$$

giving the bound

$$E_{n-1}(W_1^{(r)})_C \leqslant E_{n-1}(K)_L. \tag{19.3}$$

Actually equality holds, and the following result is valid (see, e.g., [8, Ch. 7, § 4], [24, Ch. 8, § 3]).

Theorem 19.1 (Favard). *For any $n \geqslant 1$, $r \geqslant 1$,*

$$E_{n-1}(W_1^{(r)})_C = E_{n-1}(K_r)_L.$$

Proof. In view of (19.3), to prove Theorem 19.1 it suffices to show that

$$E_{n-1}(W_1^{(r)})_C \geqslant E_{n-1}(K_r)_L.$$

To do so we construct a function $f^* \in W_1^{(r)}$ such that $E_{n-1}(f^*)_C = E_{n-1}(K_r)_L$. Let t^*_{n-1} be a polynomial of best approximation in the mean to K_r in (19.1). We claim that the function $\varphi^* = \operatorname{sign}(K_r - t^*_{n-1})$ is the derivative of order $r \geqslant 1$ of some function in $W_1^{(r)}$.

A necessary and sufficient condition for a bounded, measurable and 2π-periodic function φ to be the derivative of order $r \geqslant 1$ of some function in $W^{(r)}$ is that the function φ have a zero mean value. In our case, $\operatorname{sign}(K_r - t^*_{n-1}) = \operatorname{sign}\sin(nx+\alpha)$ for an appropriate α, and hence $\int_0^{2\pi} \operatorname{sign}\{K_r - t^*_{n-1}\}\,dx = 0$. Therefore, the function $\varphi^* = \operatorname{sign}(K_r - t^*_{n-1})$ is the derivative of order r of some function $f^* \in W_1^{(r)}$; this function can be recovered as follows (see (18.12)) keeping in mind the convention about the symbol K_r)

$$f^*(x) = \int_0^{2\pi} K_r(\theta - \pi)\,\varphi^*(x+\theta)\,d\theta.$$

For the function $\varphi^*(x) = \operatorname{sign}\sin(nx+\alpha)$, we have

$$f^*\left(x + \frac{\pi}{n}\right) = -f^*(x), \qquad x \in (-\infty, \infty);$$

hence the function f^* has a Chebyshev $2n$-alternant on $[0, 2\pi)$. Consequently, a polynomial of best uniform approximation to the function f^* vanishes identically, and so

$$E_{n-1}(f^*)_C = \|f^*\|_C = \int_0^{2\pi} |K_r(t) - t^*_{n-1}(t)|\,dt = E_{n-1}(K_r)_L.$$

Hence, $E_{n-1}(W_1^{(r)})_C = E_{n-1}(K_r)_L$, the result of Theorem 19.1. □

We have thus proved that

$$E_{n-1}(W_1^{(r)})_C = E_{n-1}(K_r)_L = n^{-r}M_r.$$

The constants M_r were first calculated by Jean Favard and hence are known as Favard constants. We have

$$M_2 = \frac{\pi}{8} \leqslant M_r \leqslant M_1 = \frac{\pi}{2}, \qquad r \geqslant 1; \qquad \lim_{r\to+\infty} M_r = \frac{4}{\pi}.$$

Given any function $f \in C_{2\pi}^{(r)}$, we can now write

$$E_{n-1}(f)_C \leqslant \frac{M_r}{n^r}\|f^{(r)}\|_C. \tag{19.4}$$

This inequality is known as the Favard or Bohr–Favard inequality .

Applying the Favard inequality to the function $f - t_{n-1}$, where t_{n-1} is an arbitrary polynomial of order $n-1$, we get

$$E_{n-1}(f)_C \leqslant \frac{M_r}{n^r}\|f^{(r)} - t_{n-1}^{(r)}\|_C.$$

The quantity $\|f^{(r)} - t_{n-1}^{(r)}\|$ in general exceeds $E_{n-1}(f^{(r)})_C$ for any t_{n-1}, because $t_{n-1}^{(r)}$ has a zero mean value. Choosing a best possible $t_{n-1}^{(r)}$, we can evaluate $\|f^{(r)} - t_{n-1}^{(r)}\|_C$ only via $2E_{n-1}(f^{(r)})_C$, because the constant term of the polynomial of best approximation to the derivative $f^{(r)}$ has a bound of the form $E_{n-1}(f^{(r)})_C$. Note that more accurate estimates allow us to eliminate the extra factor 2.

Indeed, for any function $f \in C_{2\pi}^{(r)}$ and any trigonometric polynomial τ_{n-1}, the following representation holds

$$f(x) - t_{n-1}(x) = \int_0^{2\pi} \{K_r(t) - t_{n-1}^*\}\{\varphi(x+t) - \tau_{n-1}(x+t)\}\,dt, \qquad \varphi = f^{(r)},$$

in which t_{n-1} is the trigonometric polynomial of order $n-1$ defined by the polynomial τ_{n-1}. Choosing for τ_{n-1} the polynomial of best uniform approximation to the function $\varphi = f^{(r)}$ in $C_{2\pi}$, we obtain a polynomial t_{n-1} such that

$$\|f - t_{n-1}\|_C \leqslant \frac{M_r}{n^r}E_{n-1}(f^{(r)})_C.$$

Therefore,

$$E_{n-1}(f)_C \leqslant \frac{M_r}{n^r}E_{n-1}(f^{(r)})_C. \tag{19.5}$$

Inequality (19.5) also holds for other (classical) spaces, but the constant M_r is not sharp in any of the spaces $L_{2\pi}^p$ $(1 \leqslant p < \infty)$.

We next proceed to examine applications of the Favard inequality.

19.2. Extension of Bernstein's inequality to differentiable functions

Let $f \in C_{2\pi}^{(r)}$. We want to evaluate the norm $\|f^{(r)}\|_C$ via $\|f\|_C$.
For $f = t_n$, Bernstein's inequality

$$\|f^{(r)}\|_C \leqslant n^r \|f\|_C$$

holds; this inequality becomes an equality, in particular, for the functions $f(x) = \sin nx$.
For $f \neq t_n$, Bernstein's inequality does not hold. However, the following result can be proved.

Theorem 19.2. *For $r \geqslant 1$ and for functions $f \in C_{2\pi}^{(r)}$, the following generalized Bernstein's inequality holds*:

$$\|f^{(r)}\|_C \leqslant n^r \|f\|_C + A_r E_n(f^{(r)})_C; \tag{19.6}$$

here $A_r > 0$ depends only on r.

If r is fixed and $n \to \infty$, then $E_n(f^{(r)})_C \to 0$ and $A_r E_n(f^{(r)})_C$ is small for large n.
We first prove a theorem on simultaneous approximation of function and its derivatives, which appears to be interesting in itself.

Theorem 19.3. *Let $t_n = t^*(f)$ be a polynomial of best approximation to $f \in C_{2\pi}^{(r)}$, $r \in \mathbb{N}$. Then*

$$\|f^{(k)} - t_n^{(k)}\|_C \leqslant C_r E_n(f^{(k)})_C \qquad k = 1, 2, \ldots, r \tag{19.7}$$

where $C_r > 0$ depends only on r.

Proof. To obtain Theorem 19.3 we consider the de la Vallée Poussin sum $\sigma(f) = \sigma_{n+p,n}(f)$, where $p = [\frac{n}{r}]$. Using Bernstein's, Lebesgue's and Favard's inequalities and applying Nikol'skii's theorem, we have, for $0 \leqslant k \leqslant r$,

$$\begin{aligned}
\|f^{(k)} - t_n^{(k)}\|_C &\leqslant \|f^{(k)} - \sigma(f)^{(k)}\|_C + \|(\sigma(f) - t_n)^{(k)}\|_C \\
&\leqslant \|f^{(k)} - \sigma(f)^{(k)}\|_C + (n+p)^k \|\sigma(f) - t_n\|_C \\
&= \|f^{(k)} - \sigma(f^{(k)})\|_C + (n+p)^k \|\sigma(f - t_n)\|_C \\
&\leqslant (\|\sigma\| + 1) E_n(f^{(k)})_C + (n+p)^k \|\sigma\| E_n(f)_C \\
&\leqslant (\|\sigma\| + 1) \Big\{ E_n(f^{(k)})_C + (n+p)^k E_n(f)_C \Big\} \\
&\leqslant (\|\sigma\| + 1) \Big\{ E_n(f^{(k)})_C + M_k \Big(\frac{n+p}{n+1}\Big)^k E_n(f^{(k)}) \Big\},
\end{aligned}$$

where $\|\sigma\| = \|\sigma_{n+p,n}\|$. Since $p \leqslant \frac{n}{r}$, $0 \leqslant k \leqslant r$ and $M_k \leqslant \frac{\pi}{2}$, it follows that

$$\|f^{(k)} - t_n^{(k)}\|_C \leqslant A\big(\|\sigma\| + 1\big) E(f^{(k)})_C,$$

where A is an absolute constant. Also, under the assumptions made,

$$\|\sigma\| = O\left(\ln \frac{n+p}{n+1}\right) = O(\ln(r+1)).$$

Hence, for $t_n = t_n^*(f)$ the following bound holds

$$\|f^{(k)} - t_n^{(k)}\|_C \leqslant O\big(\ln(r+1)\big) E_n(f^{(k)})_C,$$

with an absolute constant in the O-symbol. □

Proof of Theorem 19.2. Assume that a polynomial t_n gives the simultaneous approximation of both the function and its derivatives; more precisely, assume that property (19.7) is satisfied for t_n. Then, using (19.7), Bernstein's and Favard's inequalities, it follows that

$$\begin{aligned}
\|f^{(r)}\|_C &\leqslant \|f^{(r)} - t_n^{(r)}\|_C + \|t_n^{(r)}\|_C \\
&\leqslant C_r E_n(f^{(r)})_C + n^r \|t_n\|_C \\
&\leqslant C_r E_n(f^{(r)})_C + n^r \|f\|_C + n^r \|f - t_n\|_C \\
&\leqslant n^r \|f\|_C + C_r E_n(f^{(r)})_C + n^r C_r E_n(f)_C \\
&\leqslant n^r \|f\|_C + (1 + M_r) C_r E_n(f^{(r)})_C,
\end{aligned}$$

where M_r is the Favard constant. □

19.3. Application of Favard's inequality to estimating the norm of integrals

Theorem 19.4. *Suppose that $f \in C_{2\pi}^{(r)}$ is orthogonal to any polynomial $t_{n-1} \in \mathcal{T}_{n-1}$. In other words, the spectrum of f begins at value n. Then*

$$\|f\|_C \leqslant \frac{M_r}{n^r} \|f^{(r)}\|_C.$$

Proof. Indeed, since $f \perp t_{n-1}$ for any $t_{n-1} \in \mathcal{T}_{n-1}$ it follows that $f^{(r)} \perp t_{n-1}$. Hence

$$f(x) = \frac{1}{\pi} \int_0^{2\pi} \{K_r(t) - t_{n-1}^*(t)\} f^{(r)}(x+t)\, dt,$$

whence follows the inequality stated. □

19.4. Kolmogorov's inequality

Given a function $f \in C^{(r)}$, we compare the norms $\|f\|_C$, $\|f^{(k)}\|_C$, $\|f^{(n)}\|_C$ $(0 < k < n)$. The following result holds (see, e.g., [20, § 2.5.], [8, Theorem 7.1]).

Theorem 19.5 (Kolmogorov's inequality). *For any $0 < k < n$, there exists constants $K_{n,k}$ such that*

$$\|f^{(k)}\|_C \leqslant K_{n,k} \|f\|_C^{\frac{n-k}{n}} \cdot \|f^{(n)}\|_C^{\frac{k}{n}} \qquad \forall\, f \in C_{2\pi}^{(n)}.$$

Remark. The constants $K_{n,k}$ are uniformly bounded with respect to n and k.

Proof. Consider the de la Vallée Poussin sums $\sigma = \sigma_{m,p}(f)$ of a function $f \in C_{2\pi}^{(n)}$, see (16.7). We have

$$\|f^{(k)}\|_C \leqslant \|f^{(k)} - \sigma(f)^{(k)}\|_C + \|\sigma^{(k)}(f)\|_C.$$

Applying Lebesgue's inequality for the de la Vallée Poussin sums and using Favard's inequality for the kth derivative, it is found that

$$\begin{aligned}\|f^{(k)} - \sigma^{(k)}(f)\|_C &= \|f^{(k)} - \sigma(f^{(k)})\|_C \leqslant (\|\sigma\| + 1)E_{m-p}(f^{(k)})_C \\ &\leqslant (\|\sigma\| + 1)\frac{M_{n-k}}{(m-p+1)^{n-k}}\|f^{(n)}\|_C, \qquad \|\sigma\| = \|\sigma\|_C^C.\end{aligned}$$

By Bernstein's inequality,

$$\|\sigma^{(k)}(f)\|_C \leqslant m^k\|\sigma(f)\|_C \leqslant m^k\|\sigma\| \cdot \|f\|_C,$$

and hence

$$\|f^{(k)}\|_C \leqslant (\|\sigma\| + 1)M_{n-k}\left\{\frac{1}{(m-p+1)^{n-k}}\|f^{(n)}\|_C + m^k\|f\|_C\right\}.$$

To appropriately choose the parameter m (assuming, for example, that $p \leqslant \frac{m}{2}$), we need in essence to minimize the quantity

$$\min_X \left(X^{-(n-k)}\|f^{(n)}\|_C + X^k\|f\|_C\right).$$

Since the first summand decreases and the second one increases, it is convenient to choose X from the condition

$$X^{-(n-k)}\|f^{(n)}\|_C = X^k\|f\|_C,$$

and so it is natural to take m to be equal to

$$X = \left(\frac{\|f^{(n)}\|_C}{\|f\|_C}\right)^{\frac{1}{n}}.$$

With m so chosen, if we take $p = \frac{m}{2}$, then the result follows. But m (and p as well) must be integer. Hence, we take m defined by the condition $m \leqslant X \leqslant m+1$. If $X < 1$, we set $m = 0$, $p = 0$. If $X \geqslant 1$, then, for example, for $p = [\frac{m}{2}]$, we have the required order. So, finally,

$$\|f^{(k)}\|_C \leqslant K_{n,k}\|f\|_C^{\frac{n-k}{n}} \cdot \|f^{(n)}\|_C^{\frac{k}{n}},$$

the result stated. □

Lecture 20

Kolmogorov's inequality. Approximation by smooth functions. The Steklov function. Jackson's inequality

20.1. The second proof of Kolmogorov's inequality

In Lecture 19 we proved Kolmogorov's inequality relating the norms of the derivatives of differentiable periodic functions. More precisely, we proved following result (Theorem 19.5).

For any $0 < k < n$*, there exists a constant* $K_{n,k}$ *such that*

$$\|f^{(k)}\|_C \leqslant K_{n,k}\|f\|_C^{\frac{n-k}{n}} \cdot \|f^{(n)}\|_C^{\frac{k}{n}}, \qquad f \in C_{2\pi}^{(n)}. \tag{20.1}$$

For $k = 0$ and $k = n$ this inequality also holds with constant 1.

Our purpose here is to derive inequality (20.1) from the generalized Bernstein's inequality (19.6). We replace n by l. Let $0 < k < l$. Given a function $f \in C_{2\pi}^{(k)}$, we have, by the generalized Bernstein's inequality (Theorem 19.2),

$$\|f^{(k)}\|_C \leqslant n^k\|f\|_C + A_k E_n(f^{(k)})_C. \tag{20.2}$$

This inequality was originally proved only for natural n. However, since $f^{(k)} \perp \text{const}$, this inequality also holds for $n = 0$. Applying Favard's inequality (19.4) to $f^{(k)}$, we have

$$E_n(f^{(k)})_C \leqslant \frac{M_{l-k}}{(n+1)^{l-k}}\|f^{(l)}\|_C,$$

where M_{l-k} are Favard's constants; by the above, $M_{l-k} \leqslant \pi/2$. Hence, by (20.2),

$$\|f^{(k)}\|_C \leqslant n^k\|f\|_C + C'_k\frac{\|f^{(l)}\|_C}{(n+1)^{l-k}} \qquad \forall\ n = 0, 1, \dots \tag{20.3}$$

Given a real parameter $h > 0$, choose a nonnegative integer n so as to have $\frac{1}{n+1} < h \leqslant \frac{1}{n}$ for $0 < h \leqslant 1$ and $n = 0$ for $h > 1$. Then, by (20.3),

$$\|f^{(k)}\|_C \leqslant h^{-k}\|f\|_C + C_k' h^{l-k}\|f^{(l)}\| \qquad \forall\, h > 0$$

and

$$\|f^{(k)}\|_C \leqslant C_k(h^{-k}\|f\|_C + h^{l-k}\|f^{(l)}\|_C) \qquad \forall\, h > 0, \tag{20.4}$$

where $C_k = \max\{1, C_k'\}$. We choose h so that the right-hand side of the last inequality attains the smallest possible value or is close to it.

In this setting, it is frequently advantageous to use the following approach, which we shall refer to as the *least value principle* for a sum of functions. Assume that u is a decreasing and v is an increasing function on some interval I. What is the minimum value (on I) of the sum $u + v$. Let $\overline{h}$ be a point of I at which $u(\overline{h}) = v(\overline{h})$ (if any). The inequality $\inf\{u(t) + v(t):\ t \in I\} \leqslant u(\overline{h}) + v(\overline{h})$ (see Fig. 20.1) holds in general. However, it frequently happens that $u(\overline{h}) + v(\overline{h})$ is either fairly close to or sometimes even coincides with the minimum value of the sum $u + v$ on I. Hence, sometimes it suffices to substitute the smallest value of the sum by $u(\overline{h}) + v(\overline{h})$. This is the idea underlying the least value principle.

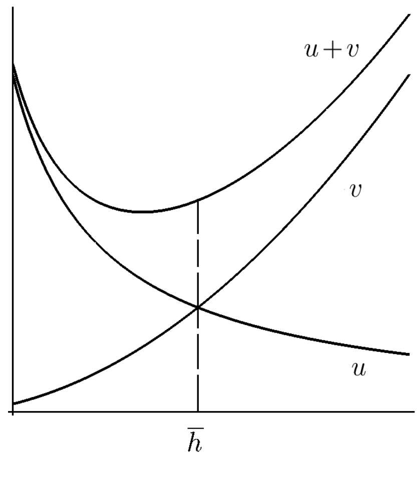

Fig. 20.1

The term h^{-k} on the right of (20.4) decreases, and h^{l-k} increases. According to the above principle, we take a point $h = \overline{h}$ so as to have $(\overline{h})^{-k}\|f\|_C = (\overline{h})^{l-k}\|f^{(l)}\|_C$. Hence,

$$\overline{h} = \left(\frac{\|f\|_C}{\|f^{(l)}\|_C}\right)^{\frac{1}{l}}.$$

Substituting this value h to (20.4), we arrive at Kolmogorov's inequality,

$$\|f^{(k)}\|_C \leqslant 2C_k\|f\|_C^{\frac{l-k}{l}}\,\|f^{(l)}\|_C^{\frac{k}{l}}.$$

We have obtained the result of the theorem with constant independent of n: $K_{n,k} \leqslant \pi A_k$.

Remark. 1) The same argument applies to give in any uniform space of 2π-periodic functions (see Remark 16.1), because the generalized Bernstein's inequality also holds for these spaces. In particular, this inequality holds in $L^p_{2\pi}$ (see, e.g., [14, Theorem 6.4], [8, Ch. 5]).

2) Inequality (20.1) also holds for functions defined on the entire line.

3) The significance of Kolmogorov's inequality is as follows: if both a function and its highest derivative $f^{(l)}$ lie in some space, then all its intermediate derivatives also lie in this space.

4) The sharp value of $K_{n,k}$ in Kolmogorov's inequality (20.1) is equal to $F_{n-k}/(F_n)^{\frac{n-k}{k}}$, where F_r is the Favard constant defined by (18.15) (see [18]).

5) Kolmogorov's inequality gives the solution of the following problem. Given numbers $M_l \geqslant 0$ and $M_0 \geqslant 0$, consider all l times differentiable functions f such that

$$\|f\|_C = M_0, \qquad \|f^{(l)}\|_C = M_l.$$

The question is as follows: What is the range of $\|f^{(k)}\|_C$, and in particular, what is the value of $M_k = \max \|f^{(k)}\|_C$? This range for M_k is described by Kolmogorov's inequality.

20.2. Term by term differentiation of sequences of functions

Consider the following problem. Suppose we are given a sequence of periodic k times differentiable functions $\{f_n\}$ converging uniformly to a function f $(f_n \rightrightarrows f)$. When the limit function is k times differentiable and $f_n^{(k)} \rightrightarrows f^{(k)}$?

Suppose that the derivatives of some order $l > k$ all exist and are uniformly bounded: $\|f_n^{(l)}\|_C \leqslant A$, $n \geqslant 1$. Kolmogorov's inequality for $f_n - f_m$ is as follows:

$$\|f_n^{(k)} - f_m^{(k)}\|_C \leqslant K\|f_n - f_m\|_C^{\frac{k}{l}} \cdot (2A)^{\frac{l-k}{l}};$$

the last inequality showing that the sequence of derivatives $\{f_n^{(k)}\}$ is a Cauchy sequence. Hence $\{f_n^{(k)}\}$ converges to some function φ. But then, by the theorem on uniform convergence of differentiable functions, the function f is k times differentiable, and $\varphi = f^{(k)}$.

It follows that the differentiation of order k, $0 < k < l$, is well-defined on the class of l times differentiable functions, whose derivative order l are uniformly bounded; more precisely, the differentiation operator of order k is continuous. This means that the differentiation operator is stable on this class: a minor error in the original function f should cause only a small error in calculating the derivatives $f^{(k)}$ $(0 < k < l)$ under a suitable method of their recovering from $\widetilde{f}(x)(\approx f(x))$.

Let $f \in C^{(l)}_{2\pi}$ and let t_n be some trigonometric polynomial. Applying Kolmogorov's inequality to $f - t_n$, gives

$$\|f^{(k)} - t_n^{(k)}\|_C \leqslant K\|f - t_n\|_C^{\frac{k}{l}} \cdot \|f^{(l)} - t_n^{(l)}\|_C^{\frac{l-k}{l}},$$

whence it follows that if both a function and its highest derivative are well approximable by a trigonometric polynomial and its highest derivative, then the intermediate derivatives are also well approximable by the corresponding derivatives of the polynomial.

Taking the Fourier sums for t_n, we obtain the bound

$$\|f^{(k)} - s_n^{(k)}\|_C \leqslant K_1 \|f - s_n\|_C^{\frac{k}{l}} \cdot \bigl(\ln(n+1) E_n(f^{(l)})\bigr)^{\frac{l-k}{l}}, \qquad n \geqslant 1,$$

which relates approximations of the derivative and the function by Fourier sums.

We note that, in Kolmogorov's inequality (20.1) in the space $C_{2\pi}$, one cannot substitute best approximations for norms: in the space $C_{2\pi}$ a polynomial of best approximation of the derivative may not be the derivative of the polynomial of best approximation to the function. However, such a change is possible in L_2.

Considerable research has been devoted to the inequalities

$$\|f^{(k)}\|_{L^q} \leqslant K \|f\|_{L^p}^{\alpha} \cdot \|f^{(l)}\|_{L^r}^{\beta}, \tag{20.5}$$

where k, l are integers, $0 \leqslant k < l$, $L^p = L^p(I)$ and I is the whole line or half-line. These inequalities are more general than (20.1) (for an overview, consult the surveys [1] and [46]).

Multiplicative inequality (20.5) is equivalent to the family of additive inequalities

$$\|f^{(k)}\|_{L^q} \leqslant A \|f\|_{L^p} + B \|f^{(l)}\|_{L^r} \tag{20.6}$$

with arbitrary $A > 0$ and appropriately chosen $B > 0$.

For the case of a finite interval I inequality (20.5) does not hold (say, for $f(x) = x^{l-1}$) and inequality (20.6) holds only for $0 < A^* \leqslant A < \infty$ with appropriate B. Some inequalities of type (20.6) with sharp constants were obtained in [3].

20.3. Jackson's inequality

20.3.1. Intermediate approximations (approximations by smooth functions)

Any periodic continuous function can be approximated to any desired degree of accuracy by smooth functions—in particular, by the Fejer sums:

$$\forall\, f \in C_{2\pi} \qquad \forall\, \varepsilon > 0 \qquad \exists\, n: \qquad \|f - \sigma_n(f)\|_C < \varepsilon.$$

We specialize the normalization parameters. Let l be a natural number and let $M > 0$. Consider the class $C^{(l)}(M)$ of all l times continuously differentiable 2π-periodic functions φ with $\|\varphi^{(l)}\|_C \leqslant M$. The class $C^{(l)}(M)$ (with fixed M) is not suitable for approximation of arbitrary functions to any desired degree of accuracy. The following problem arises.

Problem. Find

$$\inf_{\varphi \in C^{(l)}(M)} \|f - \varphi\|_C = E(f, C^{(l)}(M)).$$

This problem is infinite-dimensional and nonlinear.

20.3.2. On smoothing of functions

For $l = 1, 2$, the classical solution of the smoothing problem can be given with the help of Steklov functions to be defined later.

Given $h > 0$, we average the function $f \in C_{2\pi}$ over an interval of length h with centre at the point x. As a result, we obtain the function

$$f_h(x) = \frac{1}{h} \int_{-\frac{h}{2}}^{\frac{h}{2}} f(x+t)\,dt, \tag{20.7}$$

known as the *Steklov function*. The smoothed function (20.7) is known to be continuously differentiable, and also

$$f_h'(x) = \frac{1}{h} \left\{ f\left(x + \frac{h}{2}\right) - f\left(x - \frac{h}{2}\right) \right\}.$$

We also have $\|f_h'\|_C \leqslant \frac{2}{h}\|f\|_C$ and

$$\|f - f_h\|_C \leqslant \frac{1}{h} \int_{-\frac{h}{2}}^{\frac{h}{2}} \|f(x) - f(x+t)\|_C\,dt \leqslant \frac{2}{h} \int_0^{\frac{h}{2}} \omega(f,t)\,dt \leqslant \omega\left(f, \frac{h}{2}\right).$$

Hence, the Steklov function approximates the initial function and is continuously differentiable. However, with good accuracy of approximation the derivative f_h' is in general fairly large.

Given a real number $M > 0$, we find a parameter h so that $M = \frac{2}{h}\|f\|_C$; in other words, $h = \frac{2\|f\|_C}{M}$. As a result, we have the following theorem on approximation by Steklov functions.

Theorem 20.1. *For any function $f \in C_{2\pi}$ and any constant $M > 0$, there exists a continuously differentiable function φ, $\|\varphi'\|_C \leqslant M$, such that*

$$\|f - \varphi\|_C \leqslant \omega\left(f, \frac{\|f\|_C}{M}\right).$$

A similar analysis applies to the case $l = 2$. Consider the second order difference

$$\Delta_t^2 f(x) = f(x+t) - 2f(x) + f(x-t).$$

Integrating twice, we get

$$\frac{1}{h^2} \int_0^h \int_0^{t_1} \Delta_t^2 f(x)\,dt\ dt_1 = \varphi_h(x) - f(x), \tag{20.8}$$

where

$$\varphi_h(x) = \frac{1}{h^2} \int_0^h \int_0^{t_1} (f(x+t) + f(x-t))\,dt\ dt_1. \tag{20.9}$$

Making the change of variables $x+t = u$ in the fist and $x-t = u$ in the second integral, we obtain

$$\varphi_h(x) = \frac{1}{h^2} \int_0^h \int_{x-t_1}^{x+t_1} f(u)\ du\ dt_1.$$

This function is differentiable in x, and

$$\varphi_h'(x) = \frac{1}{h^2}\int_0^h (f(x+t_1) - f(x-t_1))\,dt_1 = \frac{1}{h^2}\left(\int_x^{x+h} f(v)\,dv - \int_{x-h}^{x} f(v)\,dv\right).$$

This expression is again differentiable, and hence the function φ_h is twice differentiable, and

$$\varphi_h''(x) = \frac{1}{h^2}\left(f(x+h) - 2f(x) + f(x-h)\right). \tag{20.10}$$

Hence, the function φ_h is twice continuously differentiable, and in view of (20.10) and (20.8), has the following properties:

$$\|\varphi_h''\|_C \leqslant \frac{1}{h^2}\|\Delta_h^2 f\|_C \leqslant \frac{4}{h^2}\|f\|_C, \qquad \|f-\varphi_h\|_C \leqslant \frac{1}{2}\,\omega_2(h,f). \tag{20.11}$$

20.3.3. Jackson's inequality in $C_{2\pi}$

Having the smoothing procedure and Favard's inequality at our disposal allows us to evaluate the best approximation of an arbitrary function in terms of its smoothness properties.

Theorem 20.2 (D. Jackson (see, e.g., [45, Part II, Ch. 6], [14, Ch. 14])**.** *For any $k \geqslant 0$ there exists a constant C_k such that, for any $f \in C^{(k)}$,*

$$E_n(f)_C \leqslant \frac{C_k}{(n+1)^k}\,\omega_2\left(\frac{1}{n}, f^{(k)}\right).$$

Proof. By Favard's inequality (19.5),

$$E_n(f)_C \leqslant \frac{M_k}{(n+1)^k} E_n(f^{(k)})_C;$$

now it suffices to estimate $E_n(f^{(k)})_C$. Consider the function φ defined in (20.9) for the derivative $f^{(k)}$ and parameter $h = \frac{1}{n}$. Hence, by (20.11),

$$\|\varphi''\|_C \leqslant n^2 \left\|\Delta_{\frac{1}{n}}^2 f^{(k)}\right\|_C \leqslant n^2\omega_2\left(\frac{1}{n}, f^{(k)}\right)$$

and so, by Favard's inequality (19.4),

$$E_n(\varphi)_C \leqslant \frac{M_2}{(n+1)^2} n^2 \omega_2\left(\frac{1}{n}, f^{(k)}\right).$$

We have

$$E_n(f^{(k)})_C \leqslant E_n(f^{(k)} - \varphi)_C + E_n(\varphi)_C \leqslant \|f^{(k)} - \varphi\|_C + E_n(\varphi)_C,$$

and hence, by (20.11),

$$E_n(f^{(k)})_C \leqslant \left(\frac{1}{2} + M_2\right)\omega_2\left(\frac{1}{n}, f^{(k)}\right). \qquad \square$$

For $k = 0$ this gives Jackson's inequality for nondifferentiable functions:

$$E_n(f)_C \leqslant C\omega_2\left(\frac{1}{n}, f\right).$$

Remark. The proof just given can be carried out in any uniform space.

References

1. V. V. Arestov, *Approximation of unbounded operators by bounded operators and related extremal problems*, Russian Math. Surveys, 51, no. 6 (1996), 1093–1126.

2. D. Braess, *Nonlinear approximation theory*, Springer, 1986.

3. V. I. Burenkov, *Exact constants in inequalities for norms of intermediate derivatives on a finite interval. II*, Trudy Mat. Inst. Steklov, 173 (1986), 38–49.

4. J. A. Clarkson, *Uniformly convex spaces*, Trans. Amer. Math. Soc. 40 (1936), 396–414.

5. P. J. Davis, *Interpolation and approximation*, Blaisdell, Waltham, MA, 1963.

6. M. M. Day, *Normed linear spaces*, Springer, Berlin, 1973.

7. R. DeVore, *The approximation of continuous functions by positive linear operators*, Springer, Berlin, 1972.

8. R. DeVore, G. G. Lorentz, *Constructive approximation*, Springer, Berlin, 1993.

9. J. Diestel, J. J. Uhl, *Vector measures*, Amer. Math. Soc, 1977.

10. P. Dodos, *Banach spaces and descriptive set theory: Selected topics*, Springer, Berlin, 2010.

11. P. Enflo, *A counterexample to the approximation property in Banach spaces*, Acta Math., 130 (1973), 309–317.

12. A. L. Garkavi, *On best approximation by the elements of infinite-dimensional subspaces of a certain class*, Sb. Math., 62(104):1 (1963), 104–120.

13. A. L. Garkavi, *On Chebyshev and almost Chebyshev subspaces*, Izv. Akad. Nauk SSSR Ser. Mat., 28:4 (1964), 799–818.

14. V. I. Ivanov, *Introduction to Approximation Theory*, Tula State. Univ., Tula, 1999 (in Russian).

15. O. G. Jørsboe, L. Mejlbro, *The Carleson–Hunt theorem on Fourier series*, Springer, Berlin, 1982.

16. S. Ya. Havinson, Z. S. Romanova, *Approximation properties of finite-dimensional subspaces in* L^1, Sb. Math., 89(131):1(9) (1972), 3–15.

17. *A. N. Kolmogorov. Selected Works*, Ed. V. M. Tikhomirov, Kluwer, 1991.

18. A. N. Kolmogorov, *Une généralisation de l'inégalité de M. J. Hadamard entre les bornes supérieures des dérivées successives d'une fonction*, C.R. Acad. Sci. Paris 207 (1938), 764–765.

19. N. P. Korneĭchuk, *S. M. Nikol'skii and the development of research on approximation theory in the USSR*, Russian Math. Surveys, 40(5) 83–156 (1985).

20. N. P. Korneĭchuk, *Splines in approximation theory*, Nauka, Moscow, 1984 [in Russian].

21. N. P. Korneĭchuk, V. F. Babenko, A. A. Ligun, *Extremal properties of polynomials and splines*, Nova Science, New York, 1996.

22. B. R. Kripke, T. J. Rivlin, Approximations in the metric of $L^1(x,\mu)$, *Trans. Amer. Math. Soc.* 119, (1965), 101–122.

23. V. I. Krylov, *Approximate calculation of integrals*, Moscow, Nauka, 1967; Dover, New York, 2005.

24. G. G. Lorentz, *Approximation of functions*, Holt, Rinehart and Winston, New York, 1966.

25. G. G. Lorentz, S. D. Riemenschneider, K. Jetter, *Birkhoff interpolation*, Addison Wesley, Mass., 1983.

26. V. A. Markov *Über Polynome die einen geyebenen Intervalle moglichst wenig von null abweichen*, St. Petersburg, 1892; Math. Ann. 77 (1916), 213–258.

27. G. Mastroianni, G. V. Milovanović, *Interpolation processes Basic theory and applications*, Springer, Berlin, 2008.

28. R. Megginson, *An introduction to Banach space theory*, Springer, Berlin, 1998.

29. D. P. Milman, *On some criteria for the regularity of spaces of the type* (B), Dokl. Akad. Nauk SSSR (N.S.) 20 (1938), 243–246.

30. I. P. Natanson, *Constructive function theory, I*, Frederic Ungar, New York, 1964.

31. I. P. Natanson, *Constructive function theory, III*, Frederic Ungar, New York, 1965.

32. S. M. Nikol'skii, *Sur certaines méthodes d'approximation au moyen de sommes trigonométriques*, Izv. Akad. Nauk SSSR Ser. Mat., 4(6) (1940), 509–520.

33. G. Nürnberger, *Approximation by splines*, Springer, Berlin, 1989.

34. B. J. Pettis, *A proof that every uniformly convex space is reflexive*, Duke Math. J. 5 (1939), 249–253.

35. J. R. Rice, *The approximation of functions. Linear theory*, Addison–Weseley, Reading, 1964.

36. J. R. Rice, *The approximation of functions. Nonlinear and multivariate theory*, Addison–Weseley, Reading, 1969.

37. T. J. Rivlin, *An introduction to the approximation of functions*, Blaisdell, Waltham, MA, 1969.

38. T. J. Rivlin, *The Chebyshev polynomials*, Wiley, New York, 1974.

39. K. Scherer, *On Bernstein's inequalities in Banach space*, Math. Notes, 17(6) (1975), 925–937.

40. H. S. Shapiro, *Topics in approximation theory*, Springer, Berlin, 1971.

41. I. Singer, *Some remarks on approximative compactness*, Rev. Roumaine Math. Pures Appl. 9 (1964), 167–177.

42. I. Singer, *Best approximation in normed linear spaces by elements of linear subspaces* Springer, Berlin, 1970.

43. I. Singer, *The theory of best approximation and functional analysis*, SIAM, Philadelphia (1974).

44. S. B. Stechkin, *Generalization of some inequalities of S. N. Bernstein*, Dokl. Math. 60 (1948), 1511–1514.

45. A. I. Stepanets, *Methods of approximation theory*, Brill Academic Publ., 2005.

46. V. Tikhomirov, A. Kochurov, *Kolmogorov-type inequalities on the whole line or half line and the Lagrange principle in the theory of extremum problems*, Eurasian Math. J., 2(3) (2011), 125–142.

47. A. F. Timan, *Theory of approximation of functions of a reaf variable*, Macmillan, New York, 1963.

48. V. A. Zalgaller, *k-dimensional directions which are singular for a convex body F in $\mathbb{R}^n$*, Zap. Nauchn. Sem. LOMI, 27 (1972), 67–72.

49. R. Webster, *Convexity*, Oxford University, Oxford, 1993.

50. E. T. Whittaker, G. N. Watson, *A course of modern analysis*, Cambridge, 1996.