INTRODUCTION TO APPROXIMATION THEORY
STUDY GUIDE

K. A. Kopotun, A. V. Prymak and I. O. Shevchuk

Taras Shevchenko National University of Kyiv
University of Manitoba

January 24, 2018

Contents

1 Preface

Approximation Theory was founded by K. Weierstrass and P. L. Chebyshev.

In the XX century Approximation Theory has been significantly developed by the following members and the corresponding members of the National Academy of Scince of URSR: N. I. Akhiezer (1901-1980, Kharkiv), S. Banach (1892-1945, Lviv), S. N. Bernstein (1980-1968, Kharkiv, Leningrad), V. K. Dzyadyk (1919-1998, Kyiv), M. I. Korneychuk (1920-2003, Dnipropertrovsk, Kyiv), M. G. Krein (1907-1989, Odesa), Y. Y. Remez (1896-1975, Kyiv), and by other prominent mathematicians.

2 Introduction. Main problem of Approximation Theory

Let X be a generally "rich" or "large" function space, and $f \in X$ be a "complicated" function. (Unless otherwise noted, everywhere below, "a function" will mean "a real-valued function".) We need to find a "simple" function p from a relatively "small" subset P of the space X, which does not differ much from f.

In Approximation Theory, X is usually a normed space, commonly one of the following:

- $C[a,b]$ — the space of continuous functions on $[a,b]$ with the norm

$$\|f\|_{C[a,b]} := \max_{x\in[a,b]} |f(x)|.$$

- $\widetilde{C}$ — the space of continuous 2π-periodic functions on $\mathbb{R}$ with the norm

$$\|f\|_{\widetilde{C}} := \max_{x\in\mathbb{R}} |f(x)| = \max_{x\in[-\pi,\pi]} |f(x)|.$$

- $L_p[a,b]$, $1 \leqslant p < \infty$ — the space of measurable (Lebesgue-measurable) functions f on $[a,b]$ such that

$$\|f\|_{L_p[a,b]} := \left(\int_a^b |f(u)|^p\, du\right)^{1/p} < +\infty.$$

- $\widetilde{L}_p$, $1 \leqslant p < \infty$ — the space of measurable 2π-periodic functions f on $\mathbb{R}$ such that

$$\|f\|_{\widetilde{L}_p} := \left(\int_{-\pi}^{\pi} |f(u)|^p\, du\right)^{1/p} < +\infty.$$

Classical choices of "small" subsets P are:

- the space $\mathcal{P}_n$ of algebraic polynomials

$$p_n(x) := a_0 + a_1 x + \cdots + a_n x^n, \quad x \in \mathbb{R},$$

  of degree $\leqslant n$, for $C[a,b]$ and $L_p[a,b]$;

- the space $\mathcal{T}_n$ of trigonometric polynomials

$$\tau_n(t) := \frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos kt + b_k \sin kt), \quad t \in \mathbb{R},$$

  of degree $\leqslant n$, for $\widetilde{C}$ and $\widetilde{L}_p$;

- the set of splines (piecewise-polynomial functions), for all four classical spaces.

Remark 2.1. $\mathcal{P}_n$ is a finite dimensional space of dimension $n+1$, and $\mathcal{T}_n$ is a space of dimension $2n+1$.

Remark 2.2. Modern Approximation Theory also studies spaces $C(M)$, $L_p(M)$, where $M \subset \mathbb{R}^d$, $M \subset \mathbb{C}^d$, $\widetilde{C}(\mathbb{R}^d)$, $\widetilde{L}_p(\mathbb{R}^d)$, spaces BMO and many others. Notable approximation tools are also rational functions $r_{m,n} = p_m/q_n$, where $p_m \in \mathcal{P}_m$, $q_n \in \mathcal{P}_n$, wavelets, radial basis functions, etc.

3 Review, basic notions

Definition 3.1 (Metric space and metric). Let X be a nonempty set. A *metric* on X is a function $d : X \times X \mapsto \mathbb{R}$ such that, for all $x, y, z \in X$,

- $d(x, x) = 0$,
- $d(x, y) > 0$ if $x \neq y$,
- $d(x, y) = d(y, x)$,
- $d(x, y) \leq d(x, z) + d(z, y)$ (the triangle inequality).

If the above is satisfied then X is called a *metric space*. Notation: $\langle X, d\rangle$.

Definition 3.2 (Norm). Let X be a *linear space*. Then the norm $\|\cdot\| := \|\cdot\|_X : X \mapsto \mathbb{R}$ is a function such that

(i) $\|v\| \geq 0$ with "=" iff $v = 0$.

(ii) $\|\lambda v\| = |\lambda| \, \|v\|$ for any $\lambda \in \mathbb{R}$ and $v \in X$.

(iii) $\|v + w\| \leq \|v\| + \|w\|$ (the triangle inequality), for any $v, w \in X$.

Notation: $\langle X, \|\cdot\|\rangle$, space X equipped with the norm $\|\cdot\|$.

Definition 3.3. If X is a normed space, then $d(x, y) := \|x - y\|$ is a metric. This is called the *metric on X induced from/by* $\|\cdot\|$.

Example 3.1. Let $X = \mathbb{R}$ and define $d : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}_+$ as follows:

$$d(x, y) := \begin{cases} 1, & \text{if } x \neq y, \\ 0, & \text{if } x = y. \end{cases}$$

Then d is a metric.

Example 3.2. Non-homogeneous translation invariant metric on X: $d(x, y) := \frac{\|x-y\|}{1+\|x-y\|}$.

Definition 3.4. A normed space X is a *Banach space* if it is complete, *i.e.,* if every fundamental/Cauchy sequence in X converges to an element of X.

Recall: Cauchy sequence $(x_n)_n$ in a metric space: $\forall \varepsilon > 0 \exists N \in \mathbb{N}$ s.t. $\forall m, n \geq N : d(x_m, x_n) < \varepsilon$.

Examples of normed spaces (all of them but one are Banach spaces, see Exercises):

1. Euclidean n-space: all n-tuples $\overline{x} = (x_1, \ldots, x_n)$ with $\|\overline{x}\|_2 := \left(\sum_{i=1}^n x_i^2\right)^{1/2}$, *i.e.,* $X = \mathbb{R}^n$ with the norm $\|\cdot\|_2$.

2. $X = \mathbb{R}^n$ consisting of all n-tuples $\overline{x} = (x_1, \ldots, x_n)$ equipped with the norm $\|\overline{x}\|_p := (\sum_{i=1}^n |x_i|^p)^{1/p}$, where $p \geq 1$.

3. $X = C[a, b]$, the space of functions continuous on $[a, b]$ with $\|f\| := \|f\|_\infty := \max_{x\in[a,b]} |f(x)|$.

4. $X = C_2[a, b]$, the space of functions continuous on $[a, b]$ with $\|f\|_2 := \left(\int_a^b |f(t)|^2 dt\right)^{1/2}$.

5. $X = \mathbb{R}^\infty$ of all bounded sequences $\overline{x} = (x_1, x_2, \dots)$ with $\|\overline{x}\| := \sup_{i\ge1} |x_i|$.

6. $X = l_p$, $p \ge 1$, of sequences $\overline{x} = (x_1, x_2, \dots)$ with $\|\overline{x}\|_p := (\sum_{i=1}^\infty |x_i|^p)^{1/p}$.

7. $X = L_p[a,b]$, $p \ge 1$ with $\|f\|_p := \left(\int_a^b |f(x)|^p dx\right)^{1/p}$.

Exercise 3.1. *Let $\langle X, \|\cdot\|\rangle$ be a normed space. Prove that $d(x,y) := \frac{\|x-y\|}{1+\|x-y\|}$ is a metric on X.*

Exercise 3.2. *Prove that, in Examples 3.1 and 3.2, X cannot be renormalized so that the metric d is induced by the new norm. (Hint: these metrics d are non-homogeneous.)*

Exercise 3.3. *What space among the above examples of normed spaces is NOT a Banach space?*

Exercise 3.4. *Let $X = l$ be the space of all sequences $\overline{x} = (x_1, x_2, \dots)$ in $\mathbb{R}$ in which only a finite number of terms are nonzero. Define $\|\overline{x}\| := \max_{i\ge1} |x_i|$. Prove that l is not a Banach space.*

Exercise 3.5. *Let $X = l_p$ with $0 < p < 1$ be the space of sequences $\overline{x} = (x_1, x_2, \dots)$ with $\|\overline{x}\|_p := (\sum_{i=1}^\infty |x_i|^p)^{1/p}$. Is $\|\cdot\|$ a norm?*

Exercise 3.6. *Let X be a normed space. Prove that the following two statements are equivalent*

*(a) X is complete (*i.e.*, X is a Banach space).*

(b) Every absolutely convergent series in X converges to an element in X, i.e., *if $(x_n) \subset X$ and $\sum_{n=1}^\infty \|x_n\| < \infty$, then $\sum_{n=1}^\infty x_n$ converges in X.*

Seminorm: $|\cdot|_X$ (same properties as those of a norm except for the condition "$\|x\| = 0$ iff $x = 0$" which may be violated). The null-space of the seminorm $|\cdot|_X$ is $X_0 := \{x \in X \mid |x|_X = 0\}$.

Example 3.3. Let $X = C^r[a,b]$, the space of all r-times continuously differentiable functions. Then $|f|_X := \max_{a\le x\le b} |f^{(r)}(x)|$ is a seminorm and its null space is the set of all polynomials of degree $\le r-1$.

Quasi-norm: The property (iii) is replaced by $\|v+w\| \le c\,(\|v\| + \|w\|)$, for all $v, w \in X$, where c is some constant that depends on the space X.

Example 3.4. If $0 < p < 1$, then $\|\overline{x}\|_p$ and $\|f\|_p$ defined above for l_p and L_p, respectively, are quasi-norms.

Exercise 3.7. *Prove that $|a+b|^p \le |a|^p + |b|^p$, for all $a, b \in \mathbb{R}$ and $0 < p < 1$.*

Exercise 3.8. *Prove that $\|\overline{x}\|_p$ and $\|f\|_p$ are quasi-norms if $0 < p < 1$.*

4 Spaces L_p and l_p

For $0 < p \le \infty$, the space $L_p(I)$ consists of all measurable functions f for which the following is finite

$$\|f\|_p := \|f\|_{L_p(I)} := \begin{cases} \left(\int_I |f(x)|^p dx\right)^{1/p}, & 0 < p < \infty, \\ \operatorname{ess\,sup}_{x\in I} |f(x)|, & p = \infty. \end{cases}$$

The discrete analogues of the spaces L_p are the spaces l_p of infinite sequences $\overline{x} = (x_i)_{i=1}^\infty$ with

$$\|\overline{x}\|_p := \begin{cases} \left(\sum_{i=1}^\infty |x_i|^p\right)^{1/p}, & 0 < p < \infty, \\ \sup_{i\ge1} |x_i|, & p = \infty. \end{cases}$$

Remark 4.1. The finite dimensional analogs of l_p are the spaces $l_p^n \subset \mathbb{R}^n$ of n-tuples $\overline{x} = (x_i)_{i=1}^n$.

Exercise 4.1. *Show that, for $n \in \mathbb{N}$ and $x_i \in \mathbb{R}$, $1 \le i \le n$,*

$$\lim_{p\to\infty} \left(\sum_{i=1}^n |x_i|^p \right)^{1/p} = \max_{1\le i\le n} |x_i|.$$

Exercise 4.2. *Suppose that $\overline{x} = (x_i)_{i=1}^\infty \in l_q$, for some $q > 0$. Prove that $\lim_{p\to\infty} \|\overline{x}\|_p = \|\overline{x}\|_\infty$,* i.e.,

$$\lim_{p\to\infty} \left(\sum_{i=1}^\infty |x_i|^p \right)^{1/p} = \sup_{i\ge1} |x_i|.$$

(Hint: (i) $\|\overline{x}\|_p \ge \|\overline{x}\|_\infty$, for all $p > 0$; (ii) for $p > q$ and $i \ge 1$, $|x_i|^p = |x_i|^{p-q}|x_i|^q \le \|\overline{x}\|_\infty^{p-q} |x_i|^q$.)

Lemma 4.1 (Hölder's inequality). *Let $f \in L_p[a,b]$, $g \in L_q[a,b]$, where $\frac{1}{p} + \frac{1}{q} = 1$. Then*

$$\|fg\|_{L_1[a,b]} \le \|f\|_{L_p[a,b]} \|g\|_{L_q[a,b]} . \tag{4.1}$$

In the case $1 < p < \infty$, the equality sign is achieved if and only if, for some constants $\lambda_1, \lambda_2 \in \mathbb{R}$, $\lambda_1|f(x)|^p = \lambda_2|g(x)|^q$ a.e. in $[a,b]$.

Exercise 4.3. *Why is it incorrect to state that, in the case $1 < p < \infty$, the equality sign in (4.1) is achieved if and only if, for some constant $\lambda \ge 0$, $|f(x)|^p = \lambda|g(x)|^q$ a.e. in $[a,b]$?*

Exercise 4.4. *Prove that, for $\alpha, \beta \ge 0$, $1 < p < \infty$, $1/p + 1/q = 1$, the following inequality holds*

$$\alpha\beta \le \frac{\alpha^p}{p} + \frac{\beta^q}{q}, \tag{4.2}$$

with equality if and only if $\alpha^p = \beta^q$. (Hint: Consider the areas of the following two regions: (i) between the curve $y = x^{p-1}$ and the x-axis with $0 \le x \le \alpha$, and (ii) between the curve $y = x^{p-1}$ (which can be rewritten as $x = y^{q-1}$) and the y-axis with $0 \le y \le \beta$. Compare the sum of these areas to $\alpha\beta$.

Proof of Lemma 4.1. We can assume that $1 < p < \infty$ and that f and g are both not 0 a.e. in $[a,b]$ (otherwise, the result is obvious). Then $\|f\|_p \ne 0$ and $\|g\|_q \ne 0$, and, for each $x \in [a,b]$, we let $\alpha = \dfrac{|f(x)|}{\|f\|_p}$ and $\beta = \dfrac{|g(x)|}{\|g\|_q}$. Then (4.2) implies

$$\frac{|f(x)||g(x)|}{\|f\|_p \|g\|_q} \le \frac{|f(x)|^p}{p\,\|f\|_p^p} + \frac{|g(x)|^q}{q\,\|g\|_q^q}, \quad x \in [a,b].$$

Now, integrate and use the fact that $1/p + 1/q = 1$. □

Note: If $p = q = 2$, then this inequality is sometimes called the Cauchy-Bunyakovsky-Schwarz inequality (or "Schwarz inequality" or "Bunyakovsky inequality"). The inequality for sums was published by Cauchy in 1821, and the corresponding inequality for integrals was first proved by Bunyakovsky in 1859. The modern proof of the integral inequality was given by Schwarz in 1885.

Exercise 4.5. *Prove the discrete analog of Hölder's inequality: If $\overline{x} = (x_i)_{i=1}^\infty \in l_p$ and $\overline{y} = (y_i)_{i=1}^\infty \in l_q$, where $\frac{1}{p} + \frac{1}{q} = 1$, then*

$$\sum_{i=1}^\infty |x_i y_i| \le \|\overline{x}\|_p \|\overline{y}\|_q . \tag{4.3}$$

Lemma 4.2 (Minkowski's inequality). *If $1 \le p \le \infty$, and $f, g \in L_p[a,b]$, then*

$$\|f+g\|_{L_p[a,b]} \le \|f\|_{L_p[a,b]} + \|g\|_{L_p[a,b]}\,, \tag{4.4}$$

In the case $1 < p < \infty$, the equality sign is achieved if and only $\lambda_1 f(x) = \lambda_2 g(x)$ a.e. in $[a,b]$, for some constants $\lambda_1, \lambda_2 \in \mathbb{R}$ such that $\lambda_1\lambda_2 \ge 0$.

Proof. This inequality is obvious if $p = 1$ or $p = \infty$. If $1 < p < \infty$, then applying Hölder's inequality to both terms on the right-hand side of the identity

$$(|f|+|g|)^p = |f|(|f|+|g|)^{p-1} + |g|(|f|+|g|)^{p-1}$$

we get (noting that $(p-1)q = p$)

$$\begin{aligned}\int_a^b (|f|+|g|)^p &= \int_a^b |f|(|f|+|g|)^{p-1} + \int_a^b |g|(|f|+|g|)^{p-1} \\ &\le \left(\int_a^b |f|^p\right)^{1/p} \left(\int_a^b (|f|+|g|)^{(p-1)q}\right)^{1/q} \\ &\quad + \left(\int_a^b |g|^p\right)^{1/p} \left(\int_a^b (|f|+|g|)^{(p-1)q}\right)^{1/q} \\ &= (\|f\|_p + \|g\|_p)\left(\int_a^b (|f|+|g|)^p\right)^{1/q},\end{aligned} \tag{4.5}$$

which implies

$$\||f|+|g|\|_p \le \|f\|_p + \|g\|_p\,. \tag{4.6}$$

Note that we get the equality in (4.5) (and hence in (4.6)) if and only if

$$c_1|f|^p = c_2(|f|+|g|)^{(p-1)q} \quad \text{and} \quad c_3|g|^p = c_4(|f|+|g|)^{(p-1)q},$$

for some $c_1, c_2, c_3, c_4 \in \mathbb{R}$, *i.e.,* if and only if $\lambda_1|f| = \lambda_2|g|$ a.e. in $[a,b]$, for some $\lambda_1, \lambda_2 \in \mathbb{R}$ satisfying $\lambda_1\lambda_2 \ge 0$. Now, in order for the equality to be achieved in (4.4), we need $\|f+g\|_p = \||f|+|g|\|_p$ which can only happen if $|f(x)+g(x)| = |f(x)|+|g(x)|$ a.e. in $[a,b]$. Therefore, f and g have to have the same sign a.e. in $[a,b]$. This completes the proof. □

Exercise 4.6. *When is the equality in Minkowski's inequality achieved in the case $p = 1$?*

Exercise 4.7. *When is the equality in Minkowski's inequality achieved in the case $p = \infty$?*

Exercise 4.8. *Prove the discrete analog of Minkowski's inequality: If $1 \le p \le \infty$ and $\overline{x}, \overline{y} \in l_p$, then*

$$\|\overline{x}+\overline{y}\|_p \le \|\overline{x}\|_p + \|\overline{y}\|_p\,.$$

Exercise 4.9. *Prove that, for a finite interval $[a,b]$ and $p \le q$,*

$$L_q[a,b] \subset L_p[a,b], \quad \|f\|_p \le c\,\|f\|_q\,,$$

where $c = (b-a)^{1/p-1/q}$.

Note: As a corollary, we get

$$\|f\|_1 \le \sqrt{b-a}\,\|f\|_2 \le (b-a)\,\|f\|_\infty\,.$$

Remark 4.2. If a sequence of functions converges in the L_q norm, then it also converges in L_p norm with $p \le q$. However, we cannot remove the restriction $p \le q$. For example, let $f_n(x) = x^{1/n}$ on $[0,1]$, and let $f^*(x) = 1$. Then, for any $1 \le p < \infty$, $\lim_{n\to\infty} \|f_n - f^*\|_p = 0$ but $\|f_n - f^*\|_\infty = 1$, for all $n \in \mathbb{N}$.

Exercise 4.10. *Prove that, for any* $0 < p < \infty$, $\displaystyle\lim_{n\to\infty} \left(\int_0^1 |1 - x^{1/n}|^p dx \right)^{1/p} = 0.$

5 Error and element of best approximation. Borel's theorem

Definition 5.1 (Approximation in a metric space). Let $f \in X$ and $A \subset X$, where X is a metric space with metric $d : X \times X \mapsto \mathbb{R}$. The degree (error) of best approximation of f from the set A is

$$E(f, A)_X := \inf_{a \in A} d(f, a).$$

Also, we say that $a^* \in A$ is a *best approximation (or "an element of best approximation") of* f *from* A if

$$d(a^*, f) \le d(a, f), \tag{5.1}$$

for all $a \in A$.

We note that, if X is a metric space and $A \subset X$, then

$$E(f, A)_X : f \in X \mapsto \mathbb{R},$$

i.e., E is a functional defined on X.

Lemma 5.2. *Let* X *be a metric space and* $A \subset X$. *The functional* $E(f) := E(f, A)_X$ *of best approximation is uniformly continuous on* X, i.e.,

$$\forall \epsilon > 0 \exists \delta > 0 \text{ s.t. } \forall f, g \in X, d(f, g) < \delta : \ |E(f) - E(g)| < \epsilon.$$

Exercise 5.1. *Prove Lemma 5.2.*

Lemma 5.3. *If* X *is a normed linear space and* A *is a linear subspace of* X, *then*

$$E(f + g, A)_X \le E(f, A)_X + E(g, A)_X$$

and

$$E(\alpha f, A)_X = |\alpha| E(f, A)_X, \quad \alpha \in \mathbb{R}.$$

Exercise 5.2. *Prove Lemma 5.3.*

Remark 5.1. If A is not a linear subspace, then Lemma 5.3 is no longer true. For example, let $X = \mathbb{R}$, $A = [0,1]$, $f = 1$ and $g = 2$. Then

$$E(f, A) = E(1, [0,1])_{\mathbb{R}} = 0, \quad E(g, A) = 1 \quad \text{and} \quad E(f + g, A) = 2,$$

i.e., it is not true that $2 \le 0 + 1$.

Question: does an element of best approximation exist? Clearly, if a^* is a best approximation to f from A in X, then $E(f, A)_X = d(a^*, f)$.

Theorem 5.4. *If A is a compact set in a metric space X, then, for every $f \in X$, there exists an element $a^* \in A$ of best approximation to f from A (*i.e., *such that (5.1) holds for all $a \in A$).*

Proof. Let $d^* := \inf_{a\in A} d(a, f)$. If there exists $a^* \in A$ such that this infimum is achieved, then we are done. Otherwise, there is a sequence $(a_i)_i \subset A$ such that $\lim_{i\to\infty} d(a_i, f) = d^*$. By compactness, we can pick a subsequence (a_{k_n}) that converges to some point $a^* \in A$, *i.e.*, $\lim_{n\to\infty} d(a_{k_n}, a^*) = 0$.

Now, $d(a^*, f) \le d(a^*, a_{k_n}) + d(a_{k_n}, f)$ and, taking the limit as $n \to \infty$, we get $d(a^*, f) \le d^*$. Hence, $d(a^*, f) = d^*$. □

Example 5.1. This theorem is not true of A is not compact. For example, let $X = \mathbb{R}$, $A = (0, 1)$ and $f = 2$.

We now turn our attention to approximation in normed linear spaces. First of all, the definitions of the error and the element of best approximation in normed spaces are analogous to those given for approximation in metric spaces (one just has to consider the metric induced from the norm). For completeness, we give these definitions explicitly.

Definition 5.5 (Approximation in a normed linear space). Let $\langle X, \|\cdot\|_X\rangle$ be a normed linear space, and let Y be its subspace. The degree (error) of best approximation of an element $f \in X$ by (the elements of) a subspace Y is the number

$$E_Y(f)_X := E(f, Y)_X := \inf_{y\in Y} \|f - y\|_X .$$

Also, we say that $y^* \in Y$ is an element of best approximation of $f \in X$ by (the elements of) a subspace Y if

$$\|f - y^*\|_X = E_Y(f)_X.$$

Remark 5.2. If the subspace Y is one of the polynomial spaces $\mathcal{P}_n$ or $\mathcal{T}_n$, then we write n instead of $\mathcal{P}_n$ or $\mathcal{T}_n$ in the notation for the error of approximation. For example, the error of best approximation of function $f \in C[a, b]$ by polynomials of degree $\leqslant n$ is

$$E_n(f)_{C[a,b]} := \inf_{p\in\mathcal{P}_n} \|f - p\|_{C[a,b]} ,$$

and the error of best approximation of $f \in \widetilde{C}$ by trigonometric polynomials of degree $\leqslant n$ is

$$E_n(f)_{\widetilde{C}} := \inf_{\tau\in\mathcal{T}_n} \|f - \tau\|_{\widetilde{C}} ,$$

with analogous definitions of the approximation errors $E_n(f)_{L_p[a,b]}$ and $E_n(f)_{\widetilde{L}_p}$ for approximation in the spaces $L_p[a, b]$ and $\widetilde{L}_p$, respectively.

Theorem 5.6 (Borel's Theorem). *If A is a finite-dimensional subspace of a normed linear space X, then, for every $f \in X$, there exists an element of A that is a best approximation from A to f.*

Proof. Let $A_0 := \{a \in A \mid \|a\| \le 2\|f\|\}$. It is compact because it is closed and bounded subset of a finite-dimensional space. Also, $0 \in A_0$ and so $A_0 \ne \emptyset$. Therefore, by Theorem 5.4, there is a best approximant a_0^* from A_0 to f. Hence,

$$\|a - f\| \ge \|a_0^* - f\| , \quad \text{for all } a \in A_0.$$

Now, if $a \in A \setminus A_0$, then $\|a\| > 2\|f\|$ and so

$$\|a - f\| \ge \|a\| - \|f\| > \|f\| = \|f - 0\| \ge \|a_0^* - f\| .$$

Therefore, for any $a \in A$, $\|a - f\| \ge \|a_0^* - f\|$, and so a_0^* is a best approximation to f from A. □

Corollary 5.7. *For every function $f \in C[a,b]$, there exists a polynomial $p_n^* \in \mathcal{P}_n$ of its best approximation in $C[a,b]$,* i.e.,

$$\|f - p_n^*\|_{C[a,b]} = E_n(f)_{C[a,b]}.$$

Corollary 5.8. *For every function $f \in \widetilde{C}$, there exists a trigonometric polynomial $\tau_n^* \in \mathcal{T}_n$ of its best approximation in $\widetilde{C}$,* i.e.,

$$\|f - \tau_n^*\|_{\widetilde{C}} = E_n(f)_{\widetilde{C}}.$$

Remark 5.3. Analogs of Corollaries 5.7 and 5.8 hold for approximation in $L_p[a,b]$ and $\widetilde{L}_p$ for $1 \leqslant p < \infty$.

Exercise 5.3. *Prove that Theorem 5.6 is no longer true if we drop the condition that A is of finite dimension. Hint: consider $X = C[0,1]$ with the usual uniform norm (*i.e., $\|f\| = \sup_{x\in[0,1]} |f(x)|$*), and let A be the set of all continuous piecewise linear functions having a finite number of linear pieces. Is it true that, for any $f \in C[0,1]$, there exists an element of best approximation from A?*

6 The uniqueness of best approximation

Exercise 6.1. *Let $f(x) := \chi_{[1/2,1]}(x)$,* i.e., *$f(x) := 1$ if $x \in [1/2,1]$ and $f(x) := 0$ otherwise. Let $X := L_1[0,1]$ and $A := \left\{\lambda \in \mathbb{R} \;\middle|\; 0 \le x \le 1\right\}$. In other words, we approximate in the L_1 norm by constant polynomials. Is the polynomial of best approximation to f from A unique?*

Exercise 6.2. *What if we approximate the same function f by constant polynomials in the L_p norm with $1 < p < \infty$? What can be said about uniqueness?*

Exercise 6.3. *What if we approximate the same function f by constant polynomials in the L_p norm with $0 < p < 1$. What can be said about uniqueness?*

Definition 6.1. The set A of a linear space X is *convex* if, for all $a_1, a_2 \in A$,

$$\left\{\lambda a_1 + (1-\lambda)a_2 \;\middle|\; 0 < \lambda < 1\right\} \subset A.$$

The set is *strictly convex* if A if, for all $a_1, a_2 \in A$ with $a_1 \neq a_2$ and $0 < \lambda < 1$, the points $\lambda a_1 + (1-\lambda)a_2$ are interior points A.

Let $\langle X, \|\cdot\|\rangle$ be a normed space, and consider the unit ball centered at the origin, *i.e.,*

$$B := \left\{f \in X \;\middle|\; \|f\| \le 1\right\}.$$

Then, for any $f_1, f_2 \in B$ and $0 < \lambda < 1$,

$$\|\lambda f_1 + (1-\lambda)f_2\| \le \|\lambda f_1\| + \|(1-\lambda)f_2\| = |\lambda|\,\|f_1\| + |1-\lambda|\,\|f_2\| \le \lambda + (1-\lambda) = 1.$$

In other words, B is a convex set.

Definition 6.2. The norm $\|\cdot\|$ in X is called *strictly convex* if the unit ball in X centered at the origin is strictly convex. The space $\langle X, \|\cdot\|\rangle$ where $\|\cdot\|$ is strictly convex is called *strictly normed.*

Exercise 6.4. *Prove that the norm $\|\cdot\|$ is strictly convex if and only if the unit sphere in X centered at the origin $\left\{f \in X \;\middle|\; \|f\| = 1\right\}$ does not contain segments,* i.e., *if $\|f_1\| = \|f_2\| = 1$, $f_1 \neq f_2$, then $\|\lambda f_1 + (1-\lambda)f_2\| < 1$, for all $0 < \lambda < 1$.*

Theorem 6.3. *If X is strictly normed and A is a convex subset of X, then, if for $f \in X$, there exists an element of best approximation from A, then it has to be unique.*

Proof. Suppose that there are two elements $a_1, a_2 \in A$ of best approximation, *i.e.,* $\|f - a_1\| = E(f, A)$ and $\|f - a_2\| = E(f, A)$. Consider $a := (a_1 + a_2)/2$ and note that $a \in A$ since A is convex. Then

$$\|f - a\| = \|(f - a_1)/2 + (f - a_2)/2\| \le (1/2)\|f - a_1\| + (1/2)\|f - a_2\| = E(f, A).$$

Therefore, $\|f - a\| = E(f, A)$ and so "$\le$" above becomes "$=$". Now, let $x_1 := (f - a_1)/\|f - a_1\|$ and $x_2 := (f - a_2)/\|f - a_2\|$. Then $\|x_1\| = \|x_2\| = 1$, and

$$\|(x_1 + x_2)/2\| = \|(f - a_1) + (f - a_2)\|/(2E(f, A)) = 1,$$

which can happen only if $x_1 = x_2$ since the norm $\|\cdot\|$ is strictly convex. Hence, $a_1 = a_2$. ☐

Exercise 6.5. *Prove that, if $1 < p < \infty$, then the space L_p is strictly normed. (Hint: Minkowski's inequality.)*

Exercise 6.6. *Prove that the space $L_1[a, b]$ is not strictly normed. (Hint: one method is to use one of the Exercises in this section. Alternatively, prove directly that the unit ball in this space is not strictly convex.)*

Exercise 6.7. *Prove that the space $C[a, b]$ is not strictly normed. (Hint: prove directly that the unit ball in this space is not strictly convex.)*

Note that if X is not strictly normed, it does NOT automatically mean that an element of best approximation to $f \in X$ from some convex set A is not unique.

Exercise 6.8. *Let $X = C[0, 1]$ with the usual uniform norm, and let A be the set of all constant polynomials on $[0, 1]$. For each function $f \in C[0, 1]$, find an element from A of best approximation and prove that it is unique. (Hint: $\max_{x\in[0,1]} f(x)$ and $\min_{x\in[0,1]} f(x)$ are achieved.)*

Exercise 6.9. *Let $L := \{h \in C[a, b] \mid h(x) = \lambda x, x \in [a, b], \lambda \in \mathbb{R}\}$ be a one-dimensional subspace of $C[a, b]$. Is an element of best approximation of $f(x) \equiv 1$ from the subspace L always unique? Find conditions guaranteeing its uniqueness. (Hint: what happens when a and b are both positive or negative? What if $a < 0 < b$? What if a or b is 0?)*

7 Approximation operators

Let X be a normed linear space, and let A be a set of approximating functions. If, for every $f \in X$, there is a unique best approximation from A to f (denote it by $b(f)$), then we may regard b is an operator from X to A, *i.e.,* $b : X \mapsto A$ is such that, for each $f \in X$, $E(f, A)_X = \|f - b(f)\|_X$.

Remark 7.1. b is a projection operator since $b^2 = b$. This immediately follows from the observation that $b(g) = g$, for any $g \in A$.

If we approximate in $X = L_2$ (Hilbert space) by elements of a finite-dimensional space A, then the operator b is linear. In that case, one can easily find best approximants by solving systems of linear equations. However, in general, the operator b is NOT linear, and one often wants to find *linear approximation operators* that give *near-best* approximants.

Definition 7.1. We say that $g \in A$ is a near-best approximant to $f \in X$ if

$$\|f-g\|_X \le cE(f,A)_X, \quad \text{for some constant } c \in \mathbb{R}.$$

Exercise 7.1. *Suppose that an operator $U : X \mapsto A$ is such that, for every $f \in X$, $U(f)$ is a near-best approximant to f from A. Prove that U is a projection operator.*

Recall: The norm of a linear operator U which maps X to $A \subset X$ is

$$\|U\| := \|U\|_X := \sup_{\|f\|_X \le 1} \|U(f)\|_X .$$

Notation: If $X = L_p[a,b]$, then we write the norm of the operator U as $\|U\|_p$ (not to be confused with the norm $\|f\|_p$ of a function $f \in L_p[a,b]$). If $X = C[a,b]$, then we denote the norm of U by $\|U\|_\infty$.

7.1 Linear projection operators

Theorem 7.2 (Lebesgue's theorem). *Let A be a finite dimensional subspace of a normed linear space X, and let $U : X \mapsto A$ be a linear* projection *operator ($U(a) = a$, for all $a \in A$). Then, for any $f \in X$,*

$$\|f - U(f)\|_X \le (1 + \|U\|_X)E(f,A)_X.$$

Note: $\|U\|_X$ is sometimes called the Lebesgue constant *of the operator U.*

Exercise 7.2. *Prove Theorem 7.2.*

Example 7.1. Let $X = C[0,1]$, and let A be the linear space $\mathcal{P}_1$ of all algebraic polynomials of degree at most one, *i.e.*,

$$A = \mathcal{P}_1 := \left\{\lambda_0 + \lambda_1 x \;\middle|\; \lambda_1, \lambda_2 \in \mathbb{R}\right\}.$$

Consider the following linear projection operator $U : C[0,1] \mapsto \mathcal{P}_1$, $U(f)(x) := f(0)(1-x) + f(1)x$, $0 \le x \le 1$ (in other words, we interpolate f at the endpoints of $[0,1]$). What can we say about approximation properties of U in various norms?

- Equip X with the L_2 norm. Let $f \in C[0,1]$ be such that $\|f\|_{L_2[0,1]} \le 1$, *i.e.*, $\int_0^1 [f(x)]^2 dx \le 1$, and consider

  $$\|U(f)\|^2_{L_2[0,1]} = \int_0^1 [f(0)(1-x) + f(1)x]^2 dx = \left([f(0)]^2 + f(0)f(1) + [f(1)]^2\right)/3.$$

  Exercise 7.3. *Show that, in this case, $\|U\|_2 = \infty$.*

- Equip X with the C norm. Let $f \in C[0,1]$ be such that $\|f\|_{C[0,1]} \le 1$, *i.e.*, $\max_{x\in[0,1]} |f(x)| \le 1$. Then

  $$\|U(f)\|_{C[0,1]} = \max_{x\in[0,1]} |f(0)(1-x) + f(1)x| = \max\{|f(0)|, |f(1)|\} \le 1,$$

  and so $\|U\|_\infty \le 1$. Now, if f_0 is such that $f_0(x) = 1$, $0 \le x \le 1$, then $U(f_0)(x) = 1$, $0 \le x \le 1$, and so, since $\|f_0\|_{C[0,1]} = 1$,

  $$\|U\|_\infty = \sup_{\|f\|_X \le 1} \|U(f)\|_X \ge \|U(f_0)\|_{C[0,1]} = 1.$$

  Therefore, $\|U\|_\infty = 1$, and so

  $$\|f - U(f)\|_{C[0,1]} \le 2E(f,A)_{C[0,1]}, \quad \text{for any } f \in C[0,1].$$

Exercise 7.4. *For any $f \in C[-1,1]$, let $U(f)$ be the function*

$$U(f)(x) = \frac{1}{2}\int_{-1}^{1}(1+3xy)f(y)dy, \quad -1 \le x \le 1.$$

Prove that the operator U is a projection from $C[-1,1]$ to the space $\mathcal{P}_1$ of linear polynomials, and that $\|U\|_\infty = 5/3$.

Exercise 7.5. *Let $m \in \mathbb{N}$, $a \le x_0 < x_1 < \cdots < x_m \le b$, and let $L_m : C[a,b] \mapsto \mathcal{P}_m$ be such that*

$$L_m(f,x) := L_m(f,x;x_0,x_1,\ldots,x_m) := \sum_{j=0}^{m} f(x_j) \prod_{i=0, i\neq j}^{m} \frac{x-x_i}{x_j-x_i}. \tag{7.1}$$

Prove the following facts:

(i) *$L(f,x_j) = f(x_j)$, for all $0 \le j \le m$. In other words, $L(f)$ is the polynomial of degree $\le m$ interpolating f at the points $x_0, x_1, \ldots, x_m$.*

(ii) *If $P \in \mathcal{P}_m$ is such that $P(x_j) = f(x_j)$, for all $0 \le j \le m$, then $P(x) = L(f,x)$, for all $x \in [a,b]$,* i.e., *there is a unique polynomial of degree $\le m$ interpolating f at $m+1$ points.*

(iii) *L is a projection operator.*

Exercise 7.6. *Let $U : C[0,1] \mapsto \mathcal{P}_2$ be such that $U(f)(x) = L_2(f,x;0,1/2,1)$, for all $f \in C[0,1]$ and $0 \le x \le 1$ (L_2 is defined in (7.1)). Find $\|U\|_\infty$.*

7.2 Positive linear operators. Korovkin's theorem.

Definition 7.3. An operator $U : C[a,b] \mapsto C[a,b]$ is *positive* if, for every $f \in C[a,b]$ such that $f(x) \ge 0$, $x \in [a,b]$, $U(f)$ is a nonnegative continuous function on $[a,b]$, *i.e.,* $U(f)(x) \ge 0$, $x \in [a,b]$.

Theorem 7.4 (Korovkin, 1957). *Let $\varphi_i(t) := t^i$, $i = 0,1,2$, and suppose that $(U_n)_{n\in\mathbb{N}}$ is a sequence of positive linear operators $U_n : C[0,1] \mapsto C[0,1]$ such that*

$$\lim_{n\to\infty} \|U_n(\varphi_i) - \varphi_i\|_{C[0,1]} = 0, \quad \text{for } i = 0,1,2. \tag{7.2}$$

Then, for any $f \in C[0,1]$,

$$\lim_{n\to\infty} \|U_n(f) - f\|_{C[0,1]} = 0.$$

Proof. Let $f \in C[0,1]$. Then f is uniformly continuous on $[0,1]$, *i.e.,*

$$\forall \epsilon > 0 \exists \delta > 0 \text{ s.t. } \forall x,t \in [0,1], |x-t| < \delta \ : \ |f(x) - f(t)| < \epsilon.$$

Also, f is bounded on $[0,1]$, *i.e.,* $\exists M \in \mathbb{R}$ s.t. $|f(x)| \le M$, for all $x \in [0,1]$. Hence, for all $x,t \in [0,1]$,

$$|f(x) - f(t)| \le |f(x)| + |f(t)| \le 2M.$$

Hence, if $|x-t| \ge \delta$, then

$$|f(x) - f(t)| \le 2M\frac{(x-t)^2}{\delta^2}.$$

Combining this with the above, we conclude that, for any $x,t \in [0,1]$,

$$|f(x) - f(t)| \le \epsilon + 2M\frac{(x-t)^2}{\delta^2} \quad \iff \quad -\epsilon - 2M\frac{(x-t)^2}{\delta^2} \le f(x) - f(t) \le \epsilon + 2M\frac{(x-t)^2}{\delta^2}.$$

Notice that, for positive linear operators U_n, $f \geq g$ on $[0,1]$ implies that $U_n(f) \geq U_n(g)$ on $[0,1]$.

We now consider x fixed, and apply U_n to the above inequalities. Note that

$$U_n(f(x) - f(t)) = f(x)U_n(1) - U_n(f)$$

and

$$\begin{aligned} U_n\left(\epsilon + 2M\frac{(x-t)^2}{\delta^2}\right) &= U_n\left(\epsilon + 2M\frac{x^2}{\delta^2}\right) + U_n\left(-\frac{4xM}{\delta^2}t\right) + U_n\left(\frac{2M}{\delta^2}t^2\right) \\ &= \left(\epsilon + 2M\frac{x^2}{\delta^2}\right)U_n(1) - \frac{4xM}{\delta^2}U_n(t) + \frac{2M}{\delta^2}U_n(t^2). \end{aligned}$$

Therefore, for all $x \in [0,1]$,

$$\begin{aligned} |f(x)U_n(1) - U_n(f)| &\leq \left|\left(\epsilon + 2M\frac{x^2}{\delta^2}\right)U_n(1) - \frac{4xM}{\delta^2}U_n(t) + \frac{2M}{\delta^2}U_n(t^2)\right| \\ &\leq \left(\epsilon + 2M\frac{x^2}{\delta^2}\right)|U_n(1) - 1| + \frac{4xM}{\delta^2}|U_n(t) - x| + \frac{2M}{\delta^2}\left|U_n(t^2) - x^2\right| \\ &\quad + \left|\left(\epsilon + 2M\frac{x^2}{\delta^2}\right) - \frac{4xM}{\delta^2}x + \frac{2M}{\delta^2}x^2\right| \\ &\leq \left(\epsilon + \frac{2M}{\delta^2}\right)|U_n(1) - 1| + \frac{4M}{\delta^2}|U_n(t) - x| + \frac{2M}{\delta^2}\left|U_n(t^2) - x^2\right| + \epsilon, \end{aligned}$$

and so

$$\begin{aligned} |f(x) - U_n(f)| &= |f(x) - f(x)U_n(1) + f(x)U_n(1) - U_n(f)| \\ &\leq |f(x)||1 - U_n(1)| + |f(x)U_n(1) - U_n(f)| \\ &\leq \left(M + \epsilon + \frac{2M}{\delta^2}\right)|U_n(1) - 1| + \frac{4M}{\delta^2}|U_n(t) - x| + \frac{2M}{\delta^2}\left|U_n(t^2) - x^2\right| + \epsilon. \end{aligned}$$

Now, taking supremum over all $x \in [0,1]$, we have, for any $\epsilon > 0$,

$$\begin{aligned} \|f - U_n(f)\|_{C[0,1]} &\leq \left(M + \epsilon + \frac{2M}{\delta^2}\right)\|U_n(\varphi_0) - \varphi_0\|_{C[0,1]} + \frac{4M}{\delta^2}\|U_n(\varphi_1) - \varphi_1\|_{C[0,1]} \\ &\quad + \frac{2M}{\delta^2}\|U_n(\varphi_2) - \varphi_2\|_{C[0,1]} + \epsilon, \end{aligned}$$

and the right-hand side can be made smaller than 4ϵ for sufficiently large n because of (7.2). □

7.3 Bernstein polynomials. Weierstrass theorem. Shape Preserving Approximation (SPA).

Definition 7.5 (Bernstein polynomials). For a function f defined on $[0,1]$, let

$$B_n(x,f) := B_n(x,f,[0,1]) := \sum_{k=0}^{n}\binom{n}{k}x^k(1-x)^{n-k}f\left(\frac{k}{n}\right), \quad n \in \mathbb{N}.$$

Remark 7.2. If the interval $[0,1]$ is replaced by $[a,b]$, then the Bernstein polynomials are defined as

$$B_n(x,f,[a,b]) = \frac{1}{(b-a)^n}\sum_{j=0}^{n}\binom{n}{j} f\left(a+\frac{j(b-a)}{n}\right)(b-x)^{n-j}(x-a)^j, \quad x\in[a,b].$$

In particular,

$$B_n(x,f,[-1,1]) = \frac{1}{2^n}\sum_{j=0}^{n}\binom{n}{j} f\left(\frac{2j}{n}-1\right)(1-x)^{n-j}(1+x)^j, \quad x\in[-1,1].$$

Note that the Bernstein operator B_n has the following properties:

(i) $B_n : C[0,1] \mapsto \mathcal{P}_n$, where $\mathcal{P}_n$ is the set of all algebraic polynomials of degree $\le n$.

(ii) B_n is a positive linear operator.

(iii) $B_n(x,1) = 1$, since

$$B_n(x,1) = \sum_{k=0}^{n}\binom{n}{k} x^k(1-x)^{n-k} = (x+(1-x))^n = 1.$$

(iv) $B_n(x,t) = x$, since

$$B_n(x,t) = \sum_{k=0}^{n}\binom{n}{k} x^k(1-x)^{n-k}\frac{k}{n} = \frac{x}{n}\frac{\partial}{\partial x}[x+(1-y)]^n\Big|_{y=x} = x.$$

(v) $B_n(x,t^2) = x^2 + \dfrac{x(1-x)}{n}$, since

$$\begin{aligned} B_n(x,t^2) &= \sum_{k=0}^{n}\binom{n}{k} x^k(1-x)^{n-k}\frac{k^2}{n^2} = \frac{x}{n^2}\frac{\partial}{\partial x}\left[x\frac{\partial}{\partial x}(x+(1-y))^n\right]\Big|_{y=x} \\ &= \frac{x}{n^2}\frac{\partial}{\partial x}\left[xn(x+(1-y))^{n-1}\right]\Big|_{y=x} = \frac{x}{n^2}\left(n+xn(n-1)\right) \\ &= x^2 + \frac{x(1-x)}{n}, \end{aligned}$$

and therefore

$$\lim_{n\to\infty}\left\|B_n(x,t^2)-x^2\right\|_{C[0,1]} = 0.$$

(vi) It follows from Theorem 7.4 that, for any $f \in C[0,1]$,

$$\lim_{n\to\infty}\|B_n(\cdot,f)-f\|_{C[0,1]} = 0.$$

(vii) If $f \in \mathcal{P}_m$, $m \in \mathbb{N}$, then $B_n(\cdot,f) \in \mathcal{P}_m$, for all $n \in \mathbb{N}$. (Recall that $\mathcal{P}_m$ denotes the set of all algebraic polynomials of degree at most m.)

Remark 7.3. Note that $B_n : C[0,1] \mapsto \mathcal{P}_n$ is not a projection operator if $n \ge 2$, since it is not true that $B_n(\cdot,g) = g$ for all $g \in \mathcal{P}_n$ (because $B_n(x,t^2) \ne x^2$). Therefore, Theorem 7.2 does not apply to operators B_n if $n \ge 2$.

Exercise 7.7. *Prove that*

$$B_n(x,t^3) = \frac{(n-1)(n-2)}{n^2}x^3 + \frac{3(n-1)}{n^2}x^2 + \frac{x}{n^2}.$$

Property (vi) of Bernstein polynomials immediately implies the following theorem.

Theorem 7.6 (Weierstrass, 1885). *For any $f \in C[0,1]$ and any $\epsilon > 0$, there is an algebraic polynomial P such that*

$$\|f - P\|_{C[0,1]} < \epsilon.$$

Exercise 7.8. *Let*

$$\widetilde{B}_n(x,f) := \sum_{k=0}^{n} \left\lfloor \binom{n}{k} f\left(\frac{k}{n}\right) \right\rfloor x^k(1-x)^{n-k}, \quad n \in \mathbb{N},$$

where $\lfloor x \rfloor$ is the largest integer not bigger than $x \in \mathbb{R}$. (Note that $\widetilde{B}_n(\cdot,f)$ are polynomials of degree at most n with integer coefficients.) Prove that, if $f \in C[0,1]$ is such that $f(0) = f(1) = 0$, then

$$\lim_{n\to\infty} \left\| \widetilde{B}_n(\cdot,f) - f \right\|_{C[0,1]} = 0.$$

(Hint: prove that (i) $n \le \binom{n}{k}$, $1 \le k \le n-1$, and (ii) $\sum_{k=1}^{n-1} x^k(1-x)^{n-k} \le \frac{1}{n}$.)

Exercise 7.9. *Prove that a function $f \in C[0,1]$ can be approximated by polynomials with integral coefficients if and only if $f(0)$ and $f(1)$ are integers. ("Can be approximated" means that, for any $\epsilon > 0$, there exists a polynomial P such that $\|f - P\|_{C[0,1]} < \epsilon$.)*

Exercise 7.10. *Prove that*

$$\frac{d}{dx}B_n(x,f) = n\sum_{k=0}^{n-1} \binom{n-1}{k} x^k(1-x)^{n-1-k}\Delta_{1/n}f\left(\frac{k}{n}\right), \tag{7.3}$$

where $\Delta_{1/n}f(t) := f\left(t + \frac{1}{n}\right) - f(t)$.

Exercise 7.11. *Prove property (vii) of Bernstein polynomials. (Hint: use the PMI, (7.3) and the observation that, if $f \in \mathcal{P}_m$, then $g \in \mathcal{P}_{m-1}$.)*

Exercise 7.12. *Prove that*

$$\frac{d^2}{dx^2}B_n(x,f) = n(n-1)\sum_{k=0}^{n-2} \binom{n-2}{k} x^k(1-x)^{n-2-k}\Delta^2_{1/n}f\left(\frac{k}{n}\right),$$

where $\Delta^2_{1/n}f(t) := f(t+2/n) - 2f(t+1/n) + f(t)$. (Hint: (7.3).)

Exercise 7.13. *Prove that for any non-decreasing f on $[0,1]$, the polynomial $B_n(\cdot,f)$ is non-decreasing on $[0,1]$. (Hint: Exercise 7.10.)*

Exercise 7.14. *Prove that if f is convex on $[0,1]$, then $B_n(\cdot,f)$ is convex on $[0,1]$. (Hint: Exercise 7.12.)*

Exercise 7.15. *Prove that, if* $f \in C^r[0,1]$ *and* $f^{(r)} \geqslant 0$, *then* $B_n(\cdot, f)^{(r)} \geqslant 0$. *(Hint: find a formula for* $d^r B_n/dx^r$ *using the same approach as was used in Exercises 7.10 and 7.12.)*

We remark that the above properties imply that Bernstein polynomials not only approximate functions but also they preserve their shapes (such as monotonicity, convexity, etc.). In other words, Bernstein polynomials provide *shape preserving approximation* (SPA).

Theorem 7.7 (SPA). *Every monotone continuous function on* $[0,1]$ *can be approximated arbitrarily well by monotone polynomials.*

Remark 7.4. Statements similar to Theorem 7.7 hold also for functions which are convex, 3-monotone (*i.e.,* whose derivatives are convex), etc.

8 Inner product spaces. Least squares approximation. Orthogonal polynomials.

Definition 8.1. Let X be a real linear space. The *inner product* $(\cdot,\cdot) : X \times X \mapsto \mathbb{R}$ is a map which has the following properties for all $f, g, h \in X$ and $\lambda \in \mathbb{R}$:

(i) $(f,f) \geq 0$, and $(f,f) = 0 \quad \Rightarrow \quad f = 0$,

(ii) $(f,g) = (g,f)$,

(iii) $(\lambda f, g) = \lambda(f,g)$ and $(f+h,g) = (f,g) + (h,g)$.

The space X equipped with an inner product is called an *inner product space* or *pre-Hilbert space.* If an inner product space equipped with the norm $\|f\| := \sqrt{(f,f)}$ (see Examples below) is complete, then it is called a *Hilbert space.* Two elements $f, g \in X$ are called *orthogonal* (notation: $f \perp g$) if $(f,g) = 0$. A set of elements $\{f_1, \dots, f_n\}$ is called *orthogonal* if $f_i \perp f_j$, for all $i \neq j$. An orthogonal set $\{f_1, \dots, f_n\}$ is called *orthonormal* if $\|f_i\| = 1$, $1 \leq i \leq n$.

Note: In these notes, if we speak about the norm in an inner product space X, we always mean $\|f\| := \sqrt{(f,f)}$.

Exercise 8.1. *Note that there is no requirement that* $(0,0) = 0$ *in the definition of the inner product. Does it mean that an inner product can be defined so that* $(0,0) \neq 0$*?*

Exercise 8.2. *Prove the Cauchy-Schwarz inequality: if* X *is an inner product space, then*

$$|(f,g)|^2 \leq (f,f) \cdot (g,g), \quad \text{for all } \ f, g \in X.$$

(Hint: $(f + \lambda g, f + \lambda g) \geq 0$, $\lambda \in \mathbb{R}$.*)*

Exercise 8.3. *Let* X *be an inner product space, and let* $\|f\| := \sqrt{(f,f)}$, $f \in X$. *Prove that* $\|\cdot\|$ *is a norm on* X.

Exercise 8.4. *Prove the Pythagorean theorem: if* $f_1, \dots, f_n \in X$ *are orthogonal, then* $\|\sum_{i=1}^n f_i\|^2 = \sum_{i=1}^n \|f_i\|^2$.

Exercise 8.5. *Prove that the following are inner product spaces:*

(a) $X = \mathbb{R}^n$ *equipped with* $(\overline{x}, \overline{y}) := \sum_{i=1}^n w_i x_i y_i$, $\overline{x} = (x_i)_{i=1}^n$, $\overline{y} = (y_i)_{i=1}^n$, *where* $w_i > 0$, $1 \leq i \leq n$.

(b) $X = C[a,b]$ *equipped with* $(f,g) := \int_a^b f(t)g(t)dt$, $f,g \in C[a,b]$.
Note: Recall that you have already seen this space above. This inner product space is not complete, and so is not a Hilbert space. The completion of this space with respect to the norm $\|f\|_2 := \left(\int_a^b |f(t)|^2 dt\right)^{1/2}$ *is the space* $L_2[a,b]$ *which is a Hilbert space.*

(c) $X = L_2^w[a,b] := \left\{ f : [a,b] \mapsto \mathbb{R} \;\middle|\; \int_a^b w|f|^2 < \infty \right\}$, *where* $w : [a,b] \mapsto \mathbb{R}$ *is a weight function (*w *is nonnegative and, for the purposes of these notes, we can assume that* w *can be zero only at finitely many points in* $[a,b]$*), equipped with*

$$(f,g) := \int_a^b w(t)f(t)g(t)dt, \quad f,g \in X.$$

Theorem 8.2. *Let* A *be a finite dimensional subspace of an inner product space* X*, and let* $f \in X$*. Then,* $g^* \in A$ *is the best approximation from* A *to* f *if and only if*

$$(f - g^*, g) = 0, \quad \text{for all } \; g \in A.$$

Exercise 8.6. *Prove Theorem 8.2.*

Exercise 8.7. *Let* $X = L_2[0,1]$*. Find the polynomial of best approximation to* $f(x) = x^2$ *from the space of all linear polynomials (in the* $L_2[0,1]$ *norm).*

Theorem 8.3. *Let* $\{g_1, \dots, g_n\}$ *be an orthonormal set of elements in an inner product space* X*, and let* $A = \operatorname{span}\{g_1, \dots, g_n\}$*. Then, for any* $f \in X$*,* $g := \sum_{i=1}^n (f, g_i) g_i$ *is the best approximant to* f *from* A*.*

Exercise 8.8. *Prove Theorem 8.3.*

Theorem 8.4 (Gram-Schmidt Theorem). *Let* $\{g_1, g_2, \dots\}$ *be a finite or infinite sequence of elements in an inner product space such that any finite collection of these elements is linearly independent. Then, for each* $n \in \mathbb{N}$*, it is possible to define* g_n^* *as a linear combination of* $\{g_1, \dots, g_n\}$ *so that the set* $\{g_1^*, g_2^*, \dots\}$ *is orthonormal.*

Exercise 8.9. *Prove Theorem 8.4 by showing that* $\{g_1^*, g_2^*, \dots\}$ *can be recursively constructed as follows:*

$$\begin{array}{ll}
h_1 := g_1, & g_1^* := h_1 / \|h_1\| \\
h_2 := g_2 - (g_2, g_1^*) g_1^*, & g_2^* := h_2 / \|h_2\| \\
\vdots & \vdots \\
h_{n+1} := g_{n+1} - \sum_{k=1}^n (g_{n+1}, g_k^*) g_k^*, & g_{n+1}^* := h_{n+1} / \|h_{n+1}\| \; .
\end{array}$$

Example 8.1. The powers $\{1, x, x^2, \dots\}$ are linearly independent in $C[a,b]$ equipped with the inner product $(f,g) = \int_a^b w(t)f(t)g(t)dt$. Therefore, the powers can be orthogonalized with respect to this inner product resulting in the orthonormal set $\{p_1^*, p_2^*, \dots\}$, where p_n^* are polynomials of degree n. Polynomials p_n^* have special names depending on $[a,b]$ and the weight function w:

$[a,b]$	$w(x)$	
$[-1,1]$	1	Legendre Polynomials
$[-1,1]$	$(1-x^2)^{-1/2}$	Chebyshev Polynomials (of the 1st kind)
$[-1,1]$	$(1-x^2)^{1/2}$	Chebyshev Polynomials (of the 2nd kind)
$[-1,1]$	$(1-x)^\alpha (1+x)^\beta$, $\alpha, \beta > -1$	Jacobi Polynomials
$[0,\infty)$	$x^\alpha e^{-x}$, $\alpha > -1$	Laguerre Polynomials
$(-\infty,\infty)$	e^{-x^2}, $\alpha > -1$	Hermite Polynomials

Note: sometimes, the same name is used to describe polynomials that are not of norm 1. For example, Legendre polynomials P_n are often "normalized" so that $P_n(1) = 1$, and Jacobi polynomials $P_n^{(\alpha,\beta)}$ are "normalized" so that $P_n^{(\alpha,\beta)}(1) = \binom{n+\alpha}{n}$.

Exercise 8.10. *Show that the first four Legendre polynomials (normalized so that their norms are* 1*) are:*

$$\begin{aligned} P_0(x) &= \frac{\sqrt{2}}{2} \\ P_1(x) &= \frac{\sqrt{6}}{2}x \\ P_2(x) &= \frac{\sqrt{10}}{4}\left(3x^2 - 1\right) \\ P_3(x) &= \frac{\sqrt{14}}{4}(5x^3 - 3x) \end{aligned}$$

Exercise 8.11. *Let $X = L_2[-1,1]$. Find the polynomial of best approximation to $f(x) = x^4$ from the space of polynomials of degree ≤ 3 (in the $L_2[-1,1]$ norm).*

Remark 8.1. Jacobi polynomials $P_n^{(\alpha,\beta)}$ (normalized so that $P_n^{(\alpha,\beta)}(1) = \binom{n+\alpha}{n}$) can also be computed using Rodrigues' formula:

$$P_n^{(\alpha,\beta)}(x) = \frac{(-1)^n}{2^n n!}(1-x)^{-\alpha}(1+x)^{-\beta}\frac{d^n}{dx^n}\left\{(1-x)^{\alpha+n}(1+x)^{\beta+n}\right\}.$$

If $\alpha = \beta = 0$, then we get a formula Legendre polynomials (normalized so that $P_n(1) = 1$):

$$P_n(x) = \frac{1}{2^n n!}\frac{d^n}{dx^n}\left\{(x^2-1)^n\right\}.$$

Exercise 8.12. *Find the first four Chebyshev polynomials of the 1st kind normalized so that their norms are* 1*.*

Remark 8.2. We will show later that Chebyshev polynomials (of the 1st kind) T_n normalized so that $T_n(1) = 1$ are $T_n(x) = \cos(n \arccos x)$, $x \in [-1,1]$.

9 Best approximation in the uniform norm. Chebyshev alternation theorem

Theorem 9.1 (Chebyshev theorem). *For a polynomial $p_n \in \mathcal{P}_n$ to be a polynomial of best approximation of $f \in C[a,b]$, it is necessary and sufficient that there exist $n+2$ points x_j satisfying*

(i) $a \leqslant x_1 < \cdots < x_{n+2} \leqslant b$,

(ii) $f(x_j) - p_n(x_j) = -(f(x_{j+1}) - p_n(x_{j+1}))$, $1 \le j \le n+1$,

and

(iii) $|f(x_j) - p_n(x_j)| = \|f - p_n\|_{C[a,b]}$, $1 \le j \le n+2$.

Proof. Denote
$$\|\cdot\| := \|\cdot\|_{C[a,b]}.$$
Without loss of generality, assume that $f \notin \mathcal{P}_n$, hence $\|f - p_n\| \neq 0$. Set
$$g := f - p_n.$$

Sufficiency. Assume to the contrary that p_n is not a polynomial of best approximation. Then there exists $p_n^* \in \mathcal{P}_n$ such that
$$\|f - p_n^*\| < \|f - p_n\|.$$
Denote $g^* := f - p_n^*$. Our assumption and conditions of the theorem imply that the graphs of continuous functions g and g^* intersect inside each interval (x_j, x_{j+1}), $1 \le j \le n+1$. This means that the function $g - g^*$ has at least $n+1$ zero, so the polynomial $p_n - p_n^*$ of degree $\leqslant n$ has at least $n+1$ zero, hence $p_n \equiv p_n^*$. A corollary of the fundamental theorem of algebra has been used: any non-trivial polynomial of degree $\leqslant n$ has no more than n real roots.
Necessity. Let p_n be a polynomial of best approximation of f and
$$E := E_n(f)_{C[a,b]} = \|f - p_n\| = \|g\|.$$

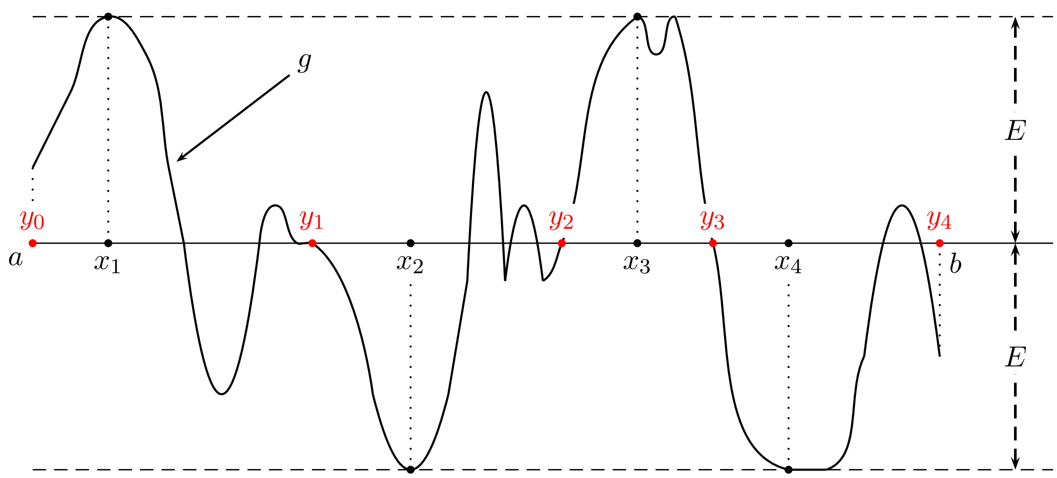

Figure 1: Chebyshev alternation $(l = 4)$

The set $\{x \in [a,b] : |g(x)| = E\}$ is closed, so we can find a point
$$x_1 := \min\{x \in [a,b] : |g(x)| = E\}.$$
We can assume that $g(x_1) > 0$, and so
$$g(x_1) = E.$$
Denote
$$x_2 := \min\{x \in [x_1, b] : g(x) = -E\}, \tag{9.1}$$
if such a point x_2 exists. We continue this process by letting
$$x_{k+1} := \min\{x \in [x_k, b] : g(x) = (-1)^k E\}, \tag{9.2}$$

if such a point x_{k+1} exists. Suppose that x_l is the last such a point that exists.

Suppose that the statement of the theorem is wrong, that is $l \le n+1$. Put $y_0 := a$, $y_l := b$,

$$y_j := \max\{x \in [x_j, x_{j+1}] : g(x) = 0\}, \quad 1 \le j \le l-1.$$

By construction, for each $0 \le j \le l-1$, we have

$$\max_{x\in[y_j,y_{j+1}]} (-1)^j g(x) = E, \tag{9.3}$$

however

$$\min_{x\in[y_j,y_{j+1}]} (-1)^j g(x) =: A_j > -E. \tag{9.4}$$

Also,

$$g(y_j) = 0, \quad 1 \le j \le l-1. \tag{9.5}$$

Now consider a polynomial

$$q(x) := \prod_{j=1}^{l-1} (y_j - x)$$

of degree $l-1 \leqslant n$ so that

$$q(y_j) = 0, \quad 1 \le j \le l-1, \tag{9.6}$$

and, for every $0 \le j \le l-1$,

$$(-1)^j q(x) > 0, \quad x \in (y_j, y_{j+1}), \tag{9.7}$$

and

$$q(a) > 0 \quad \text{and} \quad (-1)^l q(b) > 0. \tag{9.8}$$

We will show that for a small $\varepsilon > 0$ the polynomial $p_n + \varepsilon q$ of degree $\leqslant n$ approximates f "better" than the polynomial p_n, which will contradict the definition of the polynomial of best approximation. More precisely, we only need to prove the estimate

$$\|f - p_n - \varepsilon q\| < \|f - p_n\|$$

for sufficiently small $\varepsilon > 0$, or, which is the same, the inequality

$$\|g - \varepsilon q\| < \|g\| = E. \tag{9.9}$$

Indeed, if $x \in (y_j, y_{j+1})$ then the inequalities (9.3) and (9.7) imply

$$(-1)^j (g(x) - \varepsilon q(x)) \leqslant E - (-1)^j \varepsilon q(x) < E,$$

and, for all $1 \le j \le l-1$, equations (9.5) and (9.6) yield

$$(-1)^j (g(y_j) - \varepsilon q(y_j)) = 0 < E.$$

On the other hand, for $x \in [y_j, y_{j+1}]$ the inequality (9.4) implies

$$(-1)^j (\varepsilon q(x) - g(x)) \leqslant \varepsilon \|q\| - A_j = E - (E + A_j - \varepsilon \|q\|) < E,$$

where the last estimate is valid for all positive

$$\varepsilon < \frac{1}{\|q\|} \left(E + \min_{0 \le j \le l-1} A_j \right).$$

This establishes (9.9), and completes the proof of the theorem. □

Exercise 9.1. *State and prove an analog of the Chebyshev theorem for approximation of 2π-periodic continuous functions by trigonometric polynomials.*

Using the arguments of the sufficiency part of the Chebyshev theorem solve the following exercises 9.2, 9.3 and 9.4.

Exercise 9.2. *Prove the de La Vallée Poussin theorem: If for a function $f \in C[a,b]$ there exist $n+2$ points $a \leqslant x_1 < \cdots < x_{n+2} \leqslant b$ such that*

$$\operatorname{sign} f(x_j) = -\operatorname{sign} f(x_{j+1}), \quad 1 \le j \le n+1,$$

then

$$E_n(f)_{C[a,b]} \geqslant \min_{1 \le j \le n+2} |f(x_j)|.$$

Remark 9.1. The theorems of de La Vallée Poussin and Chebyshev are the foundation of the Remez algorithm, which allows to compute the polynomial of best approximation with arbitrary given accuracy (note that the polynomial of best approximation from the space $\mathcal{P}_n$ is unique, see Exercise 9.4).

Exercise 9.3. *State and prove an analog of the de La Vallée Poussin theorem for the 2π-periodic case.*

Exercise 9.4. *Given $f \in C[a,b]$, prove that its polynomial of best approximation from $\mathcal{P}_n$ is unique.*

Remark 9.2. We remark that, for approximation from arbitrary finite-dimensional subspaces of $C[a,b]$, in general, we don't have the uniqueness of the element of best approximation (see Exercise 6.9). At the same time, Exercise 9.4 states that, for approximation from $\mathcal{P}_n$, we **do** have the uniqueness. Since the question of uniqueness is of extreme importance, Haar (Chebyshev) and also Kolmogorov described all linear finite-dimensional subspaces for which the element of best approximation is unique, see the following section.

Example 9.1. The Chebyshev theorem gives an easy way to find (guess and verify) the polynomial of best approximation in certain cases. For example, it is easy to see that for $f(x) = x^2$, $x \in [-1,1]$, its polynomial of best approximation is

$$p_n^*(x) = \frac{1}{2}, \quad \text{for } n = 0 \text{ and } 1,$$

with alternation points -1, 0, and 1, and $p_n^*(x) = x^2$, for $n \ge 2$.

Similarly, for $f(x) = \sin x$, $x \in [0, 25]$, we have $p_n^* \equiv 0$, $0 \le n \le 6$, with alternation points $x_i = \frac{2i-1}{2}\pi$, $1 \le i \le 8$ (see Figure 2).

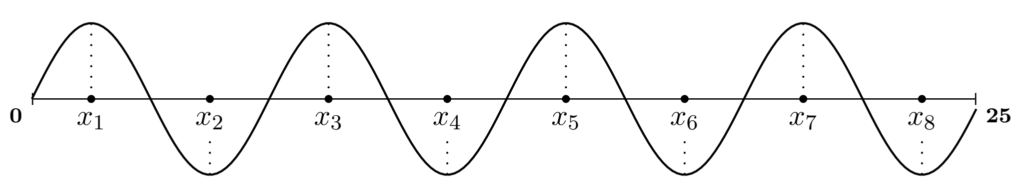

Figure 2: $f(x) = \sin x$, $0 \le x \le 25$

10 Characterization of generalized polynomials of best approximation in the uniform norm. Kolmogorov's theorem

Note: In this section, we call linear combinations of $\{\varphi_0, \dots \varphi_n\}$ *polynomials.*

Theorem 10.1 (Kolmogorov, 1948: real case). *Let $f \in C(M)$. Then P is a polynomial of best approximation if and only if, for any polynomial Q,*

$$\max_{x \in M_0} [f(x) - P(x)]Q(x) \ge 0,$$

where M_0 is the set of all points $x \in M$ such that $|f(x) - P(x)| = \|f - P\|_{C(M)}$.

Proof. Direction $\Rightarrow$. Suppose that P is a polynomial of best approximation, $\|f - P\| = E$, and suppose that there exists Q such that

$$\max_{x \in M_0} [f(x) - P(x)]Q(x) = -2\epsilon, \quad \text{for some } \epsilon > 0.$$

Since $f, P, Q \in C(M)$, there exists an open subset G of M such that $M_0 \subset G$ with

$$[f(x) - P(x)]Q(x) < -\epsilon, \quad \text{for all } x \in G.$$

Consider $P_1(x) := P(x) - \lambda Q(x)$, where $\lambda > 0$ is small.

Case 1: $x \in G$. Then

$$\begin{aligned} |f(x) - P_1(x)|^2 &= [f(x) - P(x) + \lambda Q(x)]^2 \\ &= |f(x) - P(x)|^2 + 2\lambda Q(x)[f(x) - P(x)] + \lambda^2 |Q(x)|^2 \\ &\le E^2 - 2\epsilon\lambda + \lambda^2 \|Q\|^2 = E^2 + \lambda \left(\lambda \|Q\|^2 - 2\epsilon\right) < E^2, \end{aligned}$$

for $\lambda < 2\epsilon / \|Q\|^2$.

Case 2: $x \notin G$. Since G is open, $F = M \setminus G$ is closed and, for any $x \in F$, $|f(x) - P(x)| < E$. Hence, there exists $\delta > 0$ such that $|f(x) - P(x)| < E - \delta$ for all $x \in F$. Therefore, we get

$$\begin{aligned} |f(x) - P_1(x)| &= |f(x) - P(x) + \lambda Q(x)| \le |f(x) - P(x)| + \lambda |Q(x)| \\ &< E - \delta + \lambda \|Q\| < E, \end{aligned}$$

if $\lambda < \delta / \|Q\|$.

Hence, for small λ, P_1 approximates f better than P. This is a contradiction.

Direction $\Leftarrow$. Suppose that the given inequality holds. Take arbitrary P_1 and consider $Q = P - P_1$. Then, there exists $x_0 \in M_0$ such that $[f(x_0) - P(x_0)]Q(x_0) \ge 0$. Hence,

$$\begin{aligned} |f(x_0) - P_1(x_0)|^2 &= |f(x_0) - P(x_0) + Q(x_0)|^2 \\ &= |f(x_0) - P(x_0)|^2 + 2Q(x_0)[f(x_0) - P(x_0)] + |Q(x_0)|^2 \\ &\ge |f(x_0) - P(x_0)|^2 = \|f - P\|^2. \end{aligned}$$

Therefore, P_1 cannot approximate f with error less than $\|f - P\|$, and P must be a polynomial of best approximation. ∎

Exercise 10.1. *Let $f(x) = x^2$, $x \in [0, 1]$. Find the polynomial of best uniform approximation (*i.e.*, in the $C[0, 1]$ norm) to f from the set of all linear polynomials.*

11 Haar/Chebyshev systems

Best approximation in $C(M)$ is not necessarily unique but it is unique for some finite dimensional subspaces X_n of $C(M)$. For example, it is unique for the space of algebraic and trigonometric polynomials.

Let $\langle X, \rho \rangle$ be a metric space, and let M be a subset of X containing at least $n+1$ element.

Definition 11.1. A collection of functions $\{\varphi_0, \dots, \varphi_n\}$ defined on M with values in $\mathbb{R}$ (or $\mathbb{C}$) is called a Haar (Chebyshev) system on M, if each non-trivial polynomial with respect to this system (*i.e.*, a function of the form $\sum_{j=0}^n \lambda_j \varphi_j$, where $\lambda_j \in \mathbb{R}$ (or from $\mathbb{C}$) not all equal to zero) has $\leqslant n$ distinct zeroes in M.

Example 11.1. $\{1, x, \dots, x^n\}$ is a Haar system on an arbitrary set $M \subset \mathbb{R}$ that has $\geqslant n+1$ points, since any non-trivial algebraic polynomial of degree $\leqslant n$ has at most n distinct zeroes.

Example 11.2. If $\tau \in \mathcal{T}_n$ is non-trivial, then it has $\leqslant 2n$ zeroes in every interval $(a, a+2\pi]$. Therefore, $\{1, \sin t, \cos t, \dots, \sin nt, \cos nt\}$ is a Haar system on $(a, a+2\pi]$.

Exercise 11.1. *Prove that* $\{\sin x, \dots, \sin nx\}$ *is a Haar system on* $(0, \pi)$.
(Hint: (i) $\sin x = (e^{ix} - e^{-ix})/(2i)$*, (ii) any nontrivial complex polynomial of degree* $\leq n$ *has at most* n *roots.)*

Exercise 11.2. *Check that* $\{\sin x, \dots, \sin nx\}$ *is not a Haar system on* $[0, \pi)$.

Exercise 11.3. *Prove that* $\{1, \cos x, \dots, \cos nx\}$ *is a Haar system on* $[0, \pi]$. *(Hint: use the previous exercise.)*

Exercise 11.4. *Prove that for any distinct* λ_j, $0 \leq j \leq n$, *the system* $\{e^{\lambda_0 x}, \dots, e^{\lambda_n x}\}$ *is a Haar system on* $\mathbb{R}$. *(Hint: (i) induction, (ii) how to make sure that differentiation reduces the number of functions in the system?)*

Exercise 11.5. *Prove that for any distinct* λ_j, $0 \leq j \leq n$, *the system* $\{x^{\lambda_0}, \dots, x^{\lambda_n}\}$ *is a Haar system on* $[a, b]$, $a > 0$.

Exercise 11.6. *Let* $M = [a, b]$, $f \in C^{n-1}[a, b]$ *and* $f^{(n)}(x) \neq 0$ *on* $[a, b]$. *Prove that*

$$\{1, x, x^2, \dots, x^{n-1}, f(x)\}$$

is a Haar system on $[a, b]$.

Theorem 11.2 (criterion for a system to be a Haar system). *A system* $\{\varphi_0, \dots, \varphi_n\}$ *is a Haar system on* M *if and only if, for any collection of distinct points* $\{x_j\}_{j=0}^n \subset M$,

$$\det \begin{pmatrix} \varphi_0(x_0) & \varphi_1(x_0) & \cdots & \varphi_n(x_0) \\ \varphi_0(x_1) & \varphi_1(x_1) & \cdots & \varphi_n(x_1) \\ \vdots & \vdots & \vdots & \vdots \\ \varphi_0(x_n) & \varphi_1(x_n) & \cdots & \varphi_n(x_n) \end{pmatrix} \neq 0. \tag{11.1}$$

Proof. By the definition of the Haar system, $\{\varphi_0, \dots, \varphi_n\}$ is a Haar system on M if and only if every non-trivial polynomial with respect to this system has $\leqslant n$ zeroes in M, *i.e.*, if and only if the following system of $n+1$ linear equations

$$\begin{cases} a_0\varphi_0(x_0) + \cdots + a_n\varphi_n(x_0) = 0 \\ a_0\varphi_0(x_1) + \cdots + a_n\varphi_n(x_1) = 0 \\ \qquad \vdots \\ a_0\varphi_0(x_n) + \cdots + a_n\varphi_n(x_n) = 0 \end{cases}$$

has the unique (trivial) solution with respect to $a_0, \dots, a_n$, for any collection of distinct points $\{x_j\}_{j=0}^n$. But this is equivalent to (11.1). □

Theorem 11.3 (interpolation theorem for Haar systems). *If $\{\varphi_0, \dots, \varphi_n\}$ is a Haar system on M, then for any collection of distinct points $\{x_j\}_{j=0}^n \subset M$ and an arbitrary set $\{y_j\}_{j=0}^n$, there exists exactly one polynomial*

$$p_n(x) = \sum_{j=0}^{n} a_j \varphi_j(x), \quad x \in M,$$

such that

$$p_n(x_j) = y_j, \quad 0 \le j \le n.$$

Proof. First, we show existence of such a polynomial. It suffices to consider

$$p_n(x) := -\frac{Y(x)}{D}, \quad x \in M,$$

where

$$Y(x) := \det \begin{pmatrix} 0 & \varphi_0(x) & \cdots & \varphi_n(x) \\ y_0 & \varphi_0(x_0) & \cdots & \varphi_n(x_0) \\ \vdots & \vdots & \vdots & \vdots \\ y_n & \varphi_0(x_n) & \cdots & \varphi_n(x_n) \end{pmatrix}, \quad x \in M,$$

and

$$D := \det \begin{pmatrix} \varphi_0(x_0) & \varphi_1(x_0) & \cdots & \varphi_n(x_0) \\ \varphi_0(x_1) & \varphi_1(x_1) & \cdots & \varphi_n(x_1) \\ \vdots & \vdots & \vdots & \vdots \\ \varphi_0(x_n) & \varphi_1(x_n) & \cdots & \varphi_n(x_n) \end{pmatrix},$$

and note that $D \neq 0$ by Theorem 11.2. It is easy to verify that $Y(x_j) = -y_j D$, $0 \le j \le n$, computing the determinant of Y using the Laplace expansion with respect to the first column.

In order to prove uniqueness we note that, if there are two such polynomials p_n and p_n^*, then $p_n - p_n^*$ is a polynomial with respect to the same system $\{\varphi_0, \dots, \varphi_n\}$ that has $n+1$ zeroes at the points x_j, $0 \le j \le n$. This can happen only if $p_n - p_n^*$ is identically equal to zero on M and, therefore, p_n and p_n^* must coincide on M. □

Lemma 11.4. *Suppose that $\{\varphi_0, \dots, \varphi_n\}$ is a Haar system on M, and M contains $\ge n+2$ points. If $P = \sum_{j=0}^n \lambda_j \varphi_j$ is a polynomial of best approximation to $f \in C(M)$, then the set M_0 of all points $x \in M$ for which*

$$|f(x) - P(x)| = \|f - P\|_{C(M)} =: E$$

contains at least $n+2$ points.

Proof. Suppose that $M_0 = \{x_1, \dots, x_s\}$, where $s \le n+1$. Since there are points x such that $|f(x) - P(x)| < E$, we must have $E > 0$. Now, there exists a polynomial Q such that $Q(x_k) = -[f(x_k) - P(x_k)]$, $1 \le k \le s$. Then,

$$\max_{x \in M_0} [f(x) - P(x)] Q(x) = \max_{1 \le k \le s} \left\{ -|f(x_k) - P(x_k)|^2 \right\} = -E^2 < 0.$$

This contradicts Kolmogorov's theorem (Theorem 10.1). □

Theorem 11.5. *For a Haar system $\Phi := \{\varphi_0, \dots, \varphi_n\}$ on M, there is a unique polynomial of best approximation for each continuous function $f \in C(M)$.*

Proof. If M has $n+1$ points, then each function on M is equal to a polynomial P, and this polynomial is unique. Suppose now that M has $\geq n+2$ points. Assume that f has two polynomials P and P_1 of best approximation, *i.e.*,

$$\|f-P\| = \|f-P_1\| = E.$$

Then $Q = (P+P_1)/2$ is also a polynomial of best approximation. By Lemma 11.4, there are at least $n+2$ points such that $|f(x)-Q(x)| = E$. At each such point x, consider $\alpha := f(x)-P(x)$ and $\beta := f(x)-P_1(x)$. Thus, $|\alpha+\beta| = 2E$, $|\alpha| \leq E$ and $|\beta| \leq E$. This is only possible if $\alpha = \beta = \pm E$. Thus, $P(x) = P_1(x)$ for at least $n+2$ points x in M. Since Φ is a Haar system, $P = P_1$. □

Theorem 11.6 (Haar-Kolmogorov theorem for $M = [a,b]$). *Let $M = [a,b]$ and $\{\varphi_j\}_{j=0}^n \subset C(M)$. A polynomial of best approximation with respect to this system is unique for every $f \in C(M)$ if and only if this system $\{\varphi_j\}_{j=0}^n$ is a Haar system on M.*

Exercise 11.7. *Prove Theorem 11.6.*

12 Chebyshev Polynomials

Definition 12.1. The Chebyshev polynomial of degree n is the polynomial T_n such that $T_n(\cos x) = \cos nx$, $x \in \mathbb{R}$, or $T_n(x) = \cos(n \arccos x)$, $x \in [-1,1]$.

Remark 12.1. Property 1 below implies that T_n is indeed a polynomial (Exercise 12.1).

Remark 12.2. By the fundamental theorem of algebra, any polynomial of degree n which is defined at $n+1$ points can be uniquely extended to the whole domain. Hence, it is sufficient to define T_n on any interval and, in particular, on $[-1,1]$ (since any polynomial is an analytic function on $\mathbb{C}$, the uniqueness theorem for analytic functions can also be used to reach the same conclusion).

Remark 12.3. Notation T_n for the Chebyshev polynomials comes from the German ("Tchebyshev", "Tschebyschow", etc.) and the French ("Tchebychef", "Tchebycheff", etc.) transliterations from the Russian ("Чебышев"). It is also common to use the notation C_n that comes from the English transliteration ("Chebyshev"), but we will use "T_n" in order to avoid confusion with the notation "C" that we use for constants.

It is easy to calculate the first two Chebyshev polynomials:

$$T_0(x) = 1 \quad \text{and} \quad T_1(x) = x.$$

For $n \geq 2$, T_n can be computed using the following recurrence relation.

Property 1. *The following recurrence relation holds:*

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), \quad x \in [-1,1], \quad n \geqslant 1. \tag{12.1}$$

Proof. We recall that

$$\cos(n+1)t + \cos(n-1)t = 2\cos nt \cos t, \quad t \in \mathbb{R},$$

and let $t := \arccos x$. □

Exercise 12.1. *Prove that the Chebyshev polynomials are indeed polynomials (hint: use (12.1) and induction).*

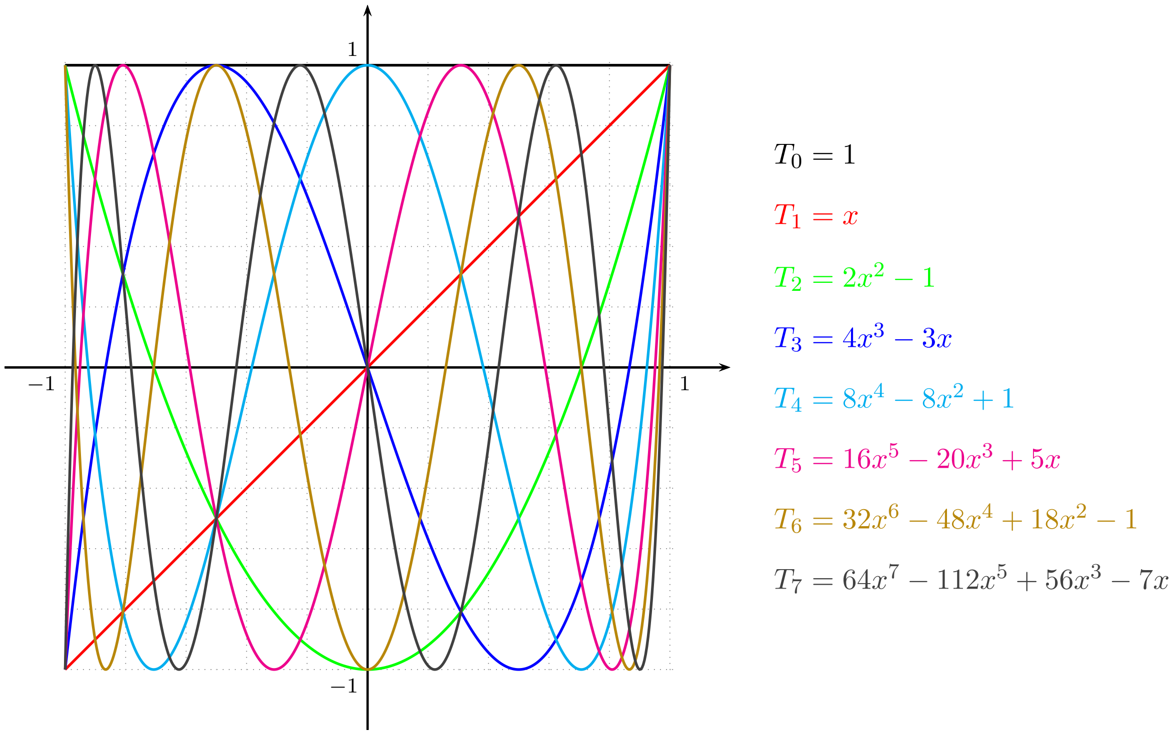

Figure 3: Chebyshev polynomials T_n, $0 \le n \le 7$

Exercise 12.2. *Prove that the recurrence relation (12.1) holds for all $x \in \mathbb{R}$.*

Exercise 12.3. *Use induction to prove that the leading coefficient of the Chebyshev polynomial T_n equals 2^{n-1}, that is*

$$T_n(x) = 2^{n-1}x^n + \dots$$

Exercise 12.4. *Use induction to prove that*

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

where $\sqrt{-|a|} = i\sqrt{|a|}$.

Exercise 12.5. *Prove that, if $n > 1$ and $|x| > 1$, then*

$$|T_n(x)| < 2^{n-1}|x|^n.$$

Property 2. *Let $f(x) := x^n$, $x \in [-1,1]$. Then the polynomial of best approximation of f of degree $\leqslant n-1$ is*

$$p_{n-1}^*(x) = x^n - \frac{1}{2^{n-1}}T_n(x), \quad x \in [-1,1],$$

and

$$E_{n-1}(f)_{C[-1,1]} = \frac{1}{2^{n-1}}.$$

Proof. The statement follows from the Chebyshev theorem (Theorem 9.1) taking into account that, for

$$g(x) := f(x) - p_{n-1}^*(x) = \frac{1}{2^{n-1}}T_n(x), \quad x \in [-1,1],$$

we have

$$\|g\|_{C[-1,1]} \le \frac{1}{2^{n-1}},$$

and the points $x_j = \cos(j\pi/n)$, $0 \le j \le n$, are the alternation points for g, since

$$g(x_j) = \frac{(-1)^j}{2^{n-1}}, \quad 0 \le j \le n.$$

□

Exercise 12.6 (Property 3)**.** *Prove that the Chebyshev polynomials satisfy the differential equation*

$$(1-x^2)T_n''(x) - xT_n'(x) + n^2T_n(x) = 0, \quad x \in \mathbb{R}.$$

Exercise 12.7 (Property 4)**.** *Prove that the Chebyshev polynomials are mutually orthogonal on* $[-1,1]$ *with the weight* $1/\sqrt{1-x^2}$, i.e.,

$$\int_{-1}^{1} \frac{T_n(x)T_m(x)}{\sqrt{1-x^2}}\,dx = 0, \quad n \ne m.$$

Exercise 12.8 (explicit form of T_n)**.** *Prove that*

$$T_n(x) = \sum_{k=0}^{[\frac{n}{2}]} (-1)^k \frac{n}{n-k}\binom{n-k}{k} 2^{n-2k-1}x^{n-2k}, \quad x \in \mathbb{R}.$$

Property 5. *Among all polynomials of degree* n *with the leading coefficient* 2^{n-1} *the Chebyshev polynomial is the one that is least deviating from zero,* i.e.,

$$\|p_n\|_{C[-1,1]} \geqslant \|T_n\|_{C[-1,1]} = 1, \quad \text{for any polynomial } p_n(x) = 2^{n-1}x^n + \dots .$$

Proof. Suppose that $\|p_n\|_{C[-1,1]} < 1$ for some polynomial $p_n(x) = 2^{n-1}x^n + \dots$, and consider the values of the difference $T_n - p_n$ at the Chebyshev knots $x_j = \cos(j\pi/n)$, $0 \le j \le n$. Since

$$T_n(x_j) = (-1)^j, \quad 0 \le j \le n,$$

then

$$\operatorname{sign}(T_n(x_j) - p_n(x_j)) = (-1)^j, \quad 0 \le j \le n,$$

and, hence, the polynomial $T_n - p_n$ of degree $\le n-1$ has at least n zeroes without being identically equal to zero. This contradicts the fundamental theorem of algebra. □

Property 5 can be restated in the following equivalent form.

Property 5′. *Among all polynomials of degree* n *with the leading coefficient* 1, *the polynomial of least deviation from zero is* $2^{1-n}T_n$.

13 Partition of Unity

Definition 13.1 (Chebyshev knots)**.** Points

$$x_j := x_{j,n} := \cos\left(\frac{j\pi}{n}\right), \quad 0 \le j \le n,$$

are called the Chebyshev knots.

Note that $T_n(x_j) = (-1)^j$, for all $0 \le j \le n$, *i.e.*, x_j's are the critical points of the Chebyshev polynomial T_n.

Let $n \in \mathbb{N}$, $n > 2$, be fixed. The Chebyshev partition of order n is the collection $\{I_j\}_{j=1}^n$ of intervals $I_j := [x_j, x_{j-1}]$. Denote $|I_j| := x_{j-1} - x_j$, $1 \le j \le n$.

Remark 13.1. In the proof of several lemmas below, we will use the inequality

$$\sin \alpha\theta < \alpha \sin\theta, \quad \text{for all } \alpha > 1 \text{ and } 0 < \theta \le \pi/\alpha, \tag{13.1}$$

and the fact that $|\sin x| \le |x|$, for all $x \in \mathbb{R}$.

Exercise 13.1. *Prove (13.1). (Hint: there are several ways to prove this inequality. For example, show that $x^{-1}\sin x$ is decreasing on $[0, \pi]$.)*

Lemma 13.2. *If $n = 2$ then $|I_1| = |I_2| = 1$, and for $n \ge 3$ the following inequalities hold*

$$|I_1| = |I_n| < |I_j| < 3|I_{j\pm1}|, \quad 2 \le j \le n-1. \tag{13.2}$$

Proof. It is easy to see that $|I_j| = |I_{n-j+1}|$, for all $1 \le j \le n$, and so it is sufficient to prove that $|I_{j+1}| < 3|I_j|$, $1 \le j \le n-1$. Using (13.1) we have

$$\frac{|I_{j+1}|}{|I_j|} = \frac{\cos\left(j\frac{\pi}{n}\right) - \cos\left((j+1)\frac{\pi}{n}\right)}{\cos\left((j-1)\frac{\pi}{n}\right) - \cos\left(j\frac{\pi}{n}\right)} = \frac{\sin\left((2j+1)\frac{\pi}{2n}\right)}{\sin\left((2j-1)\frac{\pi}{2n}\right)} < \frac{2j+1}{2j-1} \le 3.$$

Moreover, the constant 3 in the estimate (13.2) cannot be made smaller since

$$\frac{|I_2|}{|I_1|} = 4\cos^2\left(\frac{\pi}{2n}\right) - 1 \to 3, \quad n \to \infty.$$

This completes the proof of the lemma. □

Exercise 13.2. *Prove that, for any $n \in \mathbb{N}$,*

$$\frac{2}{n^2} \le |I_1| = |I_n| \le \frac{\pi^2}{2n^2}.$$

Denote by

$$x_j^0 := x_{j,n}^0 := \cos\left((j-1/2)\frac{\pi}{n}\right), \quad 1 \le j \le n,$$

the zeroes of T_n, the Chebyshev polynomial of degree n. Since

$$x_{j,n}^0 = x_{2j-1,2n} \quad \text{and} \quad x_{j,n} = x_{2j,2n},$$

by (13.2), we obtain the inequalities

$$\frac{1}{3}(x_j^0 - x_j) < x_{j-1} - x_j^0 < 3(x_j^0 - x_j),$$

and so

$$\frac{4}{3} < \frac{|I_j|}{x_j^0 - x_j} < 4 \quad \text{and} \quad \frac{4}{3} < \frac{|I_j|}{x_{j-1} - x_j^0} < 4, \quad 1 \le j \le n. \tag{13.3}$$

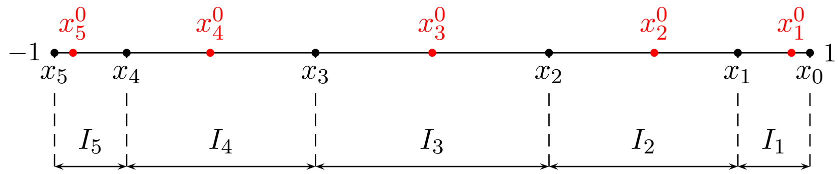

Figure 4: Chebyshev partition ($n = 5$)

Definition 13.3. For $1 \le j \le n$, let

$$Q_j(x) := \begin{cases} \dfrac{T_n(x)}{x - x_j^0}|I_j|, & x \neq x_j^0, \\ T_n'(x_j^0)|I_j|, & x = x_j^0. \end{cases}$$

(Note that, for each $1 \le j \le n$, Q_j is a polynomial of degree $n-1$.)

Clearly,

$$|Q_j(x)| \leqslant \frac{|I_j|}{|x - x_j^0|}, \quad x \in [-1,1], \quad x \neq x_j^0. \tag{13.4}$$

Lemma 13.4. *The following inequalities are valid:*

$$\frac{4}{3} < |Q_j(x)| < 4, \quad x \in I_j, \quad 1 \le j \le n. \tag{13.5}$$

Proof. Since the polynomial Q_j has $n-1$ zeros x_i^0, $1 \le i \le n$, $i \neq j$, by Rolle's theorem, its derivative Q_j' has at least one zero inside each of the intervals (x_i^0, x_{i-1}^0), where $2 \le i \le j-1$ or $j+2 \le i \le n$, and (x_{j+1}^0, x_{j-1}^0) (*i.e.*, Q_j' has at least $n-2$ zeros). Since Q_j' is a polynomial of degree $n-2$, this implies that it has to have exactly one zero in all of these $n-2$ intervals. Therefore, it has at most one zero in I_j. (Note that we use the usual convention here that if one of the endpoints of an interval is not defined, then this interval is empty, *e.g.*, $(x_2^0, x_0^0) := \emptyset$ and $(x_{n+1}^0, x_{n-1}^0) := \emptyset$.)

Note now that Q_j does not change its sign in I_j and so, if $Q_j'(x) \neq 0$, $x \in I_j$, then $|Q_j|$ is monotone on I_j. Also, if Q_j' does have a (unique) zero inside I_j, then $|Q_j|$ has a local maximum at that point. Therefore, for all $x \in I_j$,

$$|Q_j(x)| \geqslant \min\{|Q_j(x_j)|, |Q_j(x_{j-1})|\} = \min\left\{\frac{|I_j|}{|x_j - x_j^0|}, \frac{|I_j|}{|x_{j-1} - x_j^0|}\right\} > \frac{4}{3},$$

where the last inequality follows from (13.3). So, the lower bound in (13.5) is proved.

Now, we prove the upper bound in (13.5). Let

$$t := \arccos x, \quad t_j^0 := \arccos x_j^0 = \frac{(j-1/2)\pi}{n} \quad \text{and} \quad t_j := \arccos x_j = \frac{j\pi}{n}.$$

Then, for every $j \leqslant n/2, j \neq 1$, and $x \in I_j$, we have

$$\begin{aligned} |Q_j(x)| &= \left|\frac{\cos nt}{\cos t - \cos t_j^0}\right| |I_j| = \left|\frac{\cos nt - \cos nt_j^0}{\cos t - \cos t_j^0}\right| (\cos t_{j-1} - \cos t_j) \\ &= \left|\frac{\sin(\frac{n}{2}(t - t_j^0))\sin(\frac{n}{2}(t + t_j^0))}{\sin(\frac{1}{2}(t - t_j^0))\sin(\frac{1}{2}(t + t_j^0))}\right| 2\sin\frac{\pi}{2n}\sin\frac{(j-1/2)\pi}{n}. \end{aligned}$$

Now, since

$$\left|\frac{\sin(\frac{n}{2}(t-t_j^0))\sin(\frac{n}{2}(t+t_j^0))}{\sin(\frac{1}{2}(t-t_j^0))}\right| \le \left|\frac{\sin(\frac{n}{2}(t-t_j^0))}{\sin(\frac{1}{2}(t-t_j^0))}\right| \le n,$$

then using (13.1) and $\sin x \le x$, $x \ge 0$, we have

$$\begin{aligned} |Q_j(x)| &\le \frac{n}{\left|\sin(\frac{1}{2}(t+t_j^0))\right|} 2\sin\frac{\pi}{2n}\sin\frac{(j-1/2)\pi}{n} \\ &\le \frac{n}{\sin(j-\frac{3}{4})\frac{\pi}{n}} 2\sin\frac{\pi}{2n}\sin\frac{(j-1/2)\pi}{n} \\ &\leqslant \pi\frac{j-\frac{1}{2}}{j-\frac{3}{4}} \leqslant \frac{6}{5}\pi < 4. \end{aligned}$$

If $j=1$, then

$$|Q_1(x)| \leqslant Q_1(1) = \frac{|I_1|}{1-x_1^0} = \frac{|I_1|}{x_0-x_1^0} < 4,$$

where we again used (13.3). The proof for $j > n/2$ is similar. □

Lemma 13.5. *Let $m \ge 2$ be fixed. Then*

$$|I_j| \leqslant \int_{-1}^{1} |Q_j(x)|^m\,dx \leqslant c_1(m)|I_j|, \quad 1 \le j \le n,$$

where $c_1(m)$ is a constant depending only on m.

Proof. Taking into account (13.5), we have

$$\int_{-1}^{1} |Q_j(x)|^m\,dx \geqslant \int_{I_j} |Q_j(x)|^m\,dx \geqslant \int_{I_j} \left(\frac{4}{3}\right)^m dx \geqslant |I_j|.$$

We now prove the upper estimate. By (13.4), (13.5) and (13.3), we get

$$\begin{aligned} \int_{-1}^{1} |Q_j(x)|^m\,dx &= \int_{-1}^{x_j} |Q_j(x)|^m\,dx + \int_{x_{j-1}}^{1} |Q_j(x)|^m\,dx + \int_{I_j} |Q_j(x)|^m\,dx \\ &\leqslant \int_{-\infty}^{x_j} \frac{|I_j|^m}{|x-x_j^0|^m}\,dx + \int_{x_{j-1}}^{\infty} \frac{|I_j|^m}{|x-x_j^0|^m}\,dx + \int_{I_j} 4^m\,dx \\ &= \frac{1}{m-1}\left(\frac{|I_j|^m}{|x_j-x_j^0|^{m-1}} + \frac{|I_j|^m}{|x_{j-1}-x_j^0|^{m-1}}\right) + 4^m|I_j| \\ &\leqslant |I_j|\left(\frac{2}{m-1}4^{m-1} + 4^m\right) < 4^{m+1}|I_j| =: c_1(m)|I_j|. \end{aligned}$$

□

Definition 13.6. The (shifted) Heaviside step function is defined by

$$\chi_j(x) := \begin{cases} 1, & x \geqslant x_j, \\ 0, & x < x_j. \end{cases}$$

Definition 13.7. For every even number $m > 0$,

$$\widetilde{P}_j(x) := \widetilde{P}_{j,m}(x) := \frac{\int_{-1}^{x} Q_j^m(t)\,dt}{\int_{-1}^{1} Q_j^m(t)\,dt}, \quad x \in [-1,1], \quad 1 \le j \le n-1,$$

which is a polynomial of degree $m(n-1)+1 < mn$.

Since m is even, and hence $Q_j^m(t) \geqslant 0$ (and so $\widetilde{P}_j$ is nondecreasing on $[-1,1]$), then

$$0 \leqslant \widetilde{P}_j(x) \leqslant \widetilde{P}_j(1) = 1, \quad x \in [-1,1]. \tag{13.6}$$

Lemma 13.8. *For every $x \in [-1,1] \setminus \{x_j^0\}$, we have*

$$|\chi_j(x) - \widetilde{P}_j(x)| \leqslant \frac{|I_j|^{m-1}}{|x - x_j^0|^{m-1}}.$$

Proof. First, suppose that $x \in [-1, x_j)$. Then, by (13.4) and Lemma 13.5,

$$|\chi_j(x) - \widetilde{P}_j(x)| = |\widetilde{P}_j(x)| \leqslant |I_j|^{m-1} \int_{-1}^{x} \frac{1}{|t - x_j^0|^m}\,dt \leqslant |I_j|^{m-1} \int_{-\infty}^{x} \frac{1}{|t - x_j^0|^m}\,dt \le \frac{|I_j|^{m-1}}{|x - x_j^0|^{m-1}}.$$

Now, suppose that $x \in (x_{j-1}, 1]$. Then, using (13.4) and Lemma 13.5 again, we have

$$|\chi_j(x) - \widetilde{P}_j(x)| = \left| 1 - \frac{\int_{-1}^{x} Q_j^m(t)\,dt}{\int_{-1}^{1} Q_j^m(t)\,dt} \right| = \left| \frac{\int_{x}^{1} Q_j^m(t)\,dt}{\int_{-1}^{1} Q_j^m(t)\,dt} \right| \leqslant |I_j|^{m-1} \int_{x}^{\infty} \frac{1}{|t - x_j^0|^m}\,dt$$
$$\leqslant \frac{|I_j|^{m-1}}{|x - x_j^0|^{m-1}}.$$

Finally, because of (13.6), we have

$$|\chi_j(x) - \widetilde{P}_j(x)| = |1 - \widetilde{P}_j(x)| \le 1 \le \frac{|I_j|^{m-1}}{|x - x_j^0|^{m-1}},$$

for $x \in I_j$, $x \neq x_j^0$, and the proof is complete. □

Theorem 13.9 (partition of unity). *Let $m > 0$ be a fixed even number. There is a collection $\{P_{j,m}\}_{j=1}^n$ of polynomials $P_{j,m}$ of degree $< mn$ satisfying*

$$\sum_{j=1}^{n} P_{j,m}(x) \equiv 1 \tag{13.7}$$

and

$$|P_{j,m}(x)| \leqslant \min\left\{1, \frac{c_2(m)|I_j|^{m-1}}{|x - x_j|^{m-1}}\right\}, \quad x \in [-1,1], \quad 1 \le j \le n, \tag{13.8}$$

where $c_2(m)$ is a constant depending only on m.

Proof. Let $\widetilde{P}_j$, $1 \le j \le n-1$, be the polynomials from Definition 13.7. Set

$$\widetilde{P}_0(x) \equiv 0, \quad \widetilde{P}_n(x) \equiv 1 \quad \text{and} \quad P_j := P_{j,m} := \widetilde{P}_j - \widetilde{P}_{j-1}, \quad 1 \le j \le n.$$

Then (13.7) is evident, and we only need to check (13.8). Because of (13.6), we have

$$|P_j(x)| = |\widetilde{P}_j(x) - \widetilde{P}_{j-1}(x)| \le 1, \quad x \in [-1,1].$$

Now, if $x \in [x_j, x_{j-2}]$ (where $x_{-1} := x_0$), then by (13.2)

$$|x - x_j| \leqslant |I_j| + |I_{j-1}| < 4|I_j|,$$

and therefore

$$|P_j(x)| \le 1 \le \frac{4^{m-1}|I_j|^{m-1}}{|x - x_j|^{m-1}}.$$

If $x \in [-1,1] \setminus [x_j, x_{j-2}]$, then

$$|x - x_j| < 4|x - x_j^0| \quad \text{and} \quad |x - x_j| < 16|x - x_{j-1}^0|,$$

and by Lemma 13.8,

$$\begin{aligned} |P_j(x)| &\leqslant |\widetilde{P}_j(x) - \chi_j(x)| + |\widetilde{P}_{j-1}(x) - \chi_{j-1}(x)| \\ &\leqslant \frac{|I_j|^{m-1}}{|x - x_j^0|^{m-1}} + \frac{|I_{j-1}|^{m-1}}{|x - x_{j-1}^0|^{m-1}} \\ &\leqslant (4^{m-1} + 48^{m-1})\frac{|I_j|^{m-1}}{|x - x_j|^{m-1}}. \end{aligned}$$

□

Remark 13.2. Theorem 13.9 is used below to prove the estimates of errors of best approximation.

14 Modulus of Continuity

Remark 14.1. If we want to compare two continuous functions, we can say that the smoother function is the one which has more derivatives. Smoothness of non-differentiable functions can be compared using finite differences (for example, $f(x+h) - f(x)$ is the first finite difference – compare it to the derivative $f'(x) = \lim\limits_{h\to 0} \frac{f(x+h)-f(x)}{h}$).

Definition 14.1. Modulus of continuity of $f \in C[a,b]$ is the function $\omega : [0,\infty] \to [0,\infty)$ defined by

$$\omega(t) := \omega(t, f, [a,b]) := \begin{cases} \sup\limits_{h\in[0,t]} \max\limits_{x\in[a,b-h]} |f(x+h) - f(x)|, & t \leqslant b-a, \\ \max\limits_{x,x'\in[a,b]} |f(x) - f(x')|, & t > b-a. \end{cases}$$

Exercise 14.1. *Given the following functions* $f_j : [0,1] \to \mathbb{R}$, $x \in [0,1]$,

$$\begin{aligned} f_1(x) &= x^\alpha, \quad 0 < \alpha \leqslant 1; \\ f_2(x) &= 1; \\ f_3(x) &= x^2; \\ f_4(x) &= \left| x - \frac{1}{2} \right|; \\ f_5(x) &= \cos \pi x; \\ f_6(x) &= x \log \frac{e}{x}. \end{aligned}$$

prove that, for $t \in [0,1]$,

$$\begin{aligned} \omega(t, f_1, [0,1]) &= t^\alpha; \\ \omega(t, f_2, [0,1]) &= 0; \\ \omega(t, f_3, [0,1]) &= t(2-t); \\ \omega(t, f_4, [0,1]) &= \min\{1/2, t\}; \\ \omega(t, f_5, [0,1]) &= 2\sin(\pi t/2); \\ \omega(t, f_6, [0,1]) &= t \log \frac{e}{t}. \end{aligned}$$

Lemma 14.2. *Modulus of continuity* $\omega(t) := \omega(t, f, [a,b])$ *of* $f \in C[a,b]$ *has the following properties:*

1) $\omega(0) = 0$*;*

2) ω *is a non-decreasing function on* $[0, \infty)$*;*

3) ω *is a subadditive function on* $[0, \infty)$, i.e., *for arbitrary* $t_1, t_2 \geq 0$,

$$\omega(t_1 + t_2) \leqslant \omega(t_1) + \omega(t_2); \tag{14.1}$$

4) $\omega \in C[0, \infty)$.

Proof. Properties 1) and 2) are evident. In order to verify property 3), suppose that $t_1 + t_2 \leqslant b - a$. If $x, x+h \in [a,b]$, with $h = h_1 + h_2$, $h_1 \in [0, t_1]$, $h_2 \in [0, t_2]$, we obtain

$$|f(x+h) - f(x)| \leqslant |f(x + h_1 + h_2) - f(x + h_1)| + |f(x + h_1) - f(x)|,$$

which implies the required inequality. The case $t_1 + t_2 > b - a$ can be reduced to the case $t_1 + t_2 \leqslant b - a$.

We are finally ready to prove property 4). First, (right-) continuity at the point $t = 0$ follows directly from uniform continuity of f. Next, using 2) and 3), for $0 < t < t + \delta$, we get

$$\omega(t) \leqslant \omega(t + \delta) \leqslant \omega(t) + \omega(\delta),$$

which implies

$$|\omega(t + \delta) - \omega(t)| \leq \omega(\delta),$$

and so ω is uniformly continuous because $\omega(\delta) \to 0$, $\delta \to 0$. □

Exercise 14.2. *Finish the proof of subadditivity of* ω, i.e., *prove (14.1) for for* $t_1 + t_2 > b - a$.

Lemma 14.3. *Let ω be a function satisfying the conditions 1)–3) of Lemma 14.2. Then*

1) $\omega(nt) \leqslant n\omega(t)$, $n \in \mathbb{N}$, $t \geqslant 0$;

2) $\dfrac{\omega(t_2)}{t_2} \leqslant 2\dfrac{\omega(t_1)}{t_1}$, *for* $0 < t_1 < t_2$;

3) $\omega(t) \geqslant \dfrac{t}{2(b-a)}\,\omega(b-a)$, $0 < t < b-a$.

Proof. Property 1) follows from subadditivity by induction on n. To prove 2), letting $n := [t_2/t_1]$ and using 1) we have

$$\omega(t_2) \leqslant \omega((n+1)t_1) \leqslant (n+1)\omega(t_1) = \left(\left[\frac{t_2}{t_1}\right] + 1\right)\omega(t_1) \leqslant \left(\frac{t_2}{t_1} + 1\right)\omega(t_1) \leqslant 2\left(\frac{t_2}{t_1}\right)\omega(t_1).$$

Finally, property 2) with $t_1 = t$ and $t_2 = b - a$ immediately implies 3). □

Remark 14.2. If f is not a constant, then its modulus of continuity is of order $\geqslant t$, *i.e.*, $\omega(t) \geqslant ct$. Indeed, if $\omega(t) = o(t)$, then

$$\left|\frac{f(x+t) - f(x)}{t}\right| = o(t)\frac{1}{t} \to 0, \quad t \to 0,$$

which means that the derivative of f exists at any point x and equals 0, and so f is a constant function.

Lemma 14.4. *If a function $w : [0,\infty) \mapsto \mathbb{R}$ satisfies all four properties of Lemma 14.2, then it is the modulus of continuity of some continuous function (itself). In other words, for any $c > 0$,*

$$\omega(t, w, [0,c]) = w(t), \quad 0 \le t \le c.$$

Proof. Let $c > 0$ be fixed. Since w is non-decreasing and subadditive on $[0,\infty)$, we have $w(x) \le w(x+h) \le w(x) + w(h)$, for any $x, h \ge 0$. Hence, for any $t \in [0,c]$, we have

$$w(t) = w(t) - w(0) \le \omega(t, w, [0,c]) \le \sup_{0<h\le t}\max_{0\le x\le c-h} |w(x+h) - w(x)| \le \sup_{0<h\le t} w(h) = w(t).$$

Therefore, $w(t) = \omega(t, w, [0,c])$. □

Lemma 14.5. *If f is absolutely continuous on $[a,b]$ and $|f'(x)| \leqslant M$, for almost every $x \in [a,b]$ with respect to the Lebesgue measure, then*

$$\omega(t, f, [a,b]) \leqslant Mt, \quad 0 \leqslant t \leqslant b-a,$$

which means that f is a Lipschitz function (or sometimes we say that "f is a Lip1 *function"). Notation: $f \in \mathrm{Lip}_M 1$.*

Proof. By absolute continuity of f, the Lebesgue theorem implies

$$f(x) = f(a) + \int_a^x f'(t)\,dt, \quad x \in [a,b],$$

hence, for $x', x'' \in [a,b]$,

$$|f(x') - f(x'')| = \left| \int_{x'}^{x''} f'(t)\,dt \right| \leqslant \int_{x'}^{x''} |f'(t)|\,dt \leqslant M\,|x' - x''|,$$

and so $\omega(t, f, [a,b]) \leqslant Mt$. □

Lemma 14.6. *If*

$$\omega(t, f, [a,b]) \leqslant Mt, \quad 0 \leqslant t \leqslant b-a,$$

that is $f \in \mathrm{Lip}_M 1$, then f is absolutely continuous on $[a,b]$ and $|f'(x)| \leqslant M$ for almost all $x \in [a,b]$ with respect to the Lebesgue measure.

Proof. First of all, absolute continuity immediately follows from the condition $f \in \mathrm{Lip}_M 1$. Now, since f is absolutely continuous it is differentiable almost everywhere, and for those $x \in [a,b]$ where the derivative exists, we have

$$|f'(x)| = \left| \lim_{h \to 0} \frac{f(x+h) - f(x)}{h} \right| \leqslant \lim_{h \to 0} \left| \frac{Mh}{h} \right| = M.$$

□

Lemma 14.7. *If φ is such that $\varphi(0) \geq 0$ and $\varphi(t)/t$ is non-increasing on $(0, \infty)$, then φ is subadditive on $[0, \infty)$.*

Proof. Let $t_1, t_2 > 0$, then

$$\varphi(t_1 + t_2) = \frac{t_1}{t_1 + t_2}\varphi(t_1 + t_2) + \frac{t_2}{t_1 + t_2}\varphi(t_1 + t_2) \leqslant \varphi(t_1) + \varphi(t_2).$$

□

Remark 14.3. Generally speaking, the modulus of continuity $\omega(t, f, [a,b]) = \omega(t)$ is not a function such that $\omega(t)/t$ is non-increasing. However, we have the inequality

$$\omega(\lambda t) \leqslant (\lambda + 1)\omega(t), \quad \lambda, t > 0, \tag{14.2}$$

which is a corollary of property 1) in Lemma 14.3. Indeed, using this property we have

$$\omega(\lambda t) \leq \omega\left((\lfloor \lambda \rfloor + 1)t\right) \leq (\lfloor \lambda \rfloor + 1)\omega(t) \leq (\lambda + 1)\omega(t).$$

Lemma 14.8. *If φ is a concave function on $[0, \infty)$ such that $\varphi(0) \geq 0$, then φ is subadditive on $[0, \infty)$.*

Proof. We will show that $\varphi(t)/t$ is non-increasing on $(0, \infty)$ and apply Lemma 14.7. Let $0 < t_1 < t_2$, define $\lambda := t_1/t_2$ and note that $0 < \lambda < 1$. Now, since φ is concave on $[0, \infty)$, we have

$$\varphi(t_1) = \varphi(\lambda t_2) = \varphi(\lambda t_2 + (1-\lambda)\cdot 0) \geq \lambda\varphi(t_2) + (1-\lambda)\varphi(0) \geq \lambda\varphi(t_2) = (t_1/t_2)\varphi(t_2).$$

Therefore, $\varphi(t_1)/t_1 \geq \varphi(t_2)/t_2$, *i.e.*, $\varphi(t)/t$ is non-increasing on $(0, \infty)$. □

Exercise 14.3. *Show that one cannot replace $[0, \infty)$ by $(0, \infty)$ in the requirement "φ is a concave function on $[0, \infty)$" in the statement of Lemma 14.8. (Hint: $\ln t$.)*

Exercise 14.4. *Prove that a concave non-decreasing function w such that $w(0) = 0$ satisfies all four properties of Lemma 14.2.*

15 Direct Theorem for Polynomial Approximation

For convenience, denote $E_n(f) := E_n(f)_{C[-1,1]}$, and $\omega(t,f) := \omega(t,f,[-1,1])$.

Theorem 15.1 (direct theorem for polynomial approximation). *For any $f \in C[-1,1]$, the (Jackson) inequality*

$$E_n(f) \leqslant c_1\omega(1/n,f), \quad n \in \mathbb{N}, \tag{15.1}$$

holds, where c_1 is an absolute constant.

Remark 15.1. Theorem 7.6 (the Weierstrass theorem) follows from Theorem 15.1 and the continuity of the modulus of continuity at zero.

Proof of Theorem 15.1. We will use Theorem 13.9 (about partition of unity), and take $m = 6$ and $c_2 := c_2(6)$, where the absolute constant $c_2(6)$ is determined by this theorem (see (13.8)). Theorem 13.9 implies that there exists a collection $\{P_j\}_{j=1}^n$ of algebraic polynomials of degree $< mn = 6n$ satisfying $\sum_{j=1}^n P_j(x) \equiv 1$ and

$$|P_j(x)| \leqslant \min\left\{1, \frac{c_2|I_j|^5}{|x-x_j|^5}\right\}, \quad x \in [-1,1], \quad 1 \le j \le n. \tag{15.2}$$

Recall that $x_j = \cos(j\pi/n)$, $I_j = [x_j, x_{j-1}]$, $|I_j| = x_{j-1} - x_j$, and consider the polynomial

$$\mathcal{Q}_n(x) := \sum_{j=1}^n f(x_j)P_j(x), \quad x \in [-1,1],$$

of degree $< 6n$.

We now fix $x \in [-1,1]$ and $1 \le i \le n$ be such that such that $x \in I_i$. Then

$$\begin{aligned} |f(x) - \mathcal{Q}_n(x)| &= \left|\sum_{j=1}^n (f(x) - f(x_j))P_j(x)\right| \leqslant \sum_{j=1}^n \omega(|x-x_j|,f)|P_j(x)| \\ &= \sum_{j=i-1}^{i+1} \omega(|x-x_j|,f)|P_j(x)| + \sum_{j=1, j\ne i\pm1, i}^{n} \omega(|x-x_j|,f)|P_j(x)|. \end{aligned} \tag{15.3}$$

In order to estimate the first sum, taking into account

$$x - x_i \leqslant |I_i|, \quad x_{i-1} - x \leqslant |I_i|, \quad x - x_{i+1} \leqslant 4|I_i|, \quad |P_j(x)| \leqslant 1,$$

and Lemma 14.3 we get

$$\sum_{j=i-1}^{i+1} \omega(|x-x_j|,f)|P_j(x)| \leqslant 2\omega(|I_i|,f) + \omega(4|I_i|,f) \leqslant 6\omega(|I_i|,f).$$

To estimate the second sum, for $x \in I_i$ and $j \ne i, i\pm1$, we use the inequalities

$$|P_j(x)| \leqslant \frac{c_2|I_j|^5}{|x-x_j|^5} \leqslant \frac{c_3|I_j|^5}{|x-x_j||x_i-x_j|^4},$$

$$\omega(|x-x_j|,f) = \omega\left(\frac{|x-x_j|}{|I_i|}|I_i|, f\right) \leqslant \left(\frac{|x-x_j|}{|I_i|} + 1\right)\omega(|I_i|,f) \leqslant 4\frac{|x-x_j|}{|I_i|}\,\omega(|I_i|,f),$$

and obtain

$$\sum_{j=1,j\neq i\pm1,i}^{n} \omega(|x-x_j|,f)|P_j(x)| \leqslant 4c_3\omega(|I_i|,f) \sum_{j=1,j\neq i\pm1,i}^{n} \frac{|I_j|^5}{|I_i||x_i-x_j|^4} =: 4c_3\omega(|I_i|,f) \sum_{j=1,j\neq i\pm1,i}^{n} A_j.$$

Applying the inequalities (see Exercise 15.1)

$$|I_j|^2 \leqslant 9|I_i||x_i-x_j| \quad \text{and} \quad |x_i-x_{j-1}| < 4|x_i-x_j|,$$

we have that for $j \neq i-1, i, i+1$,

$$A_j = \frac{|I_j|^5}{|I_i||x_i-x_j|^4} \leqslant \frac{9^2|I_i||I_j|}{|x_i-x_j|^2} \leqslant \frac{4\cdot 9^2|I_i||I_j|}{|x_i-x_j||x_i-x_{j-1}|} = 324|I_i| \int_{I_j} \frac{dx}{(x-x_i)^2}.$$

Finally,

$$\begin{aligned}
\sum_{j=1,j\neq i\pm1,i}^{n} A_j &= \sum_{j=i+2}^{n} A_j + \sum_{j=1}^{i-2} A_j \\
&\leqslant 324|I_i| \int_{-1}^{x_{i+1}} \frac{dx}{(x-x_i)^2} + 324|I_i| \int_{x_{i-2}}^{1} \frac{dx}{(x-x_i)^2} \\
&\leqslant 324|I_i| \int_{-\infty}^{x_{i+1}} \frac{dx}{(x-x_i)^2} + 324|I_i| \int_{x_{i-2}}^{\infty} \frac{dx}{(x-x_i)^2} \\
&= 324|I_i| \left(\frac{1}{|I_{i+1}|} + \frac{1}{|I_i|+|I_{i-1}|} \right) \\
&< c_4.
\end{aligned}$$

Now, (15.3) implies

$$|f(x) - \mathcal{Q}_n(x)| \leqslant c\omega(|I_i|,f) + c\omega(|I_i|,f) \leq c\omega(|I_i|,f).$$

Since x was an arbitrary fixed number from $[-1,1]$ and since $|I_i| \leq 1/n$, we have

$$|f(x) - \mathcal{Q}_n(x)| \leqslant c\omega(|I_i|,f) \leqslant c\omega(1/n,f), \quad x \in [-1,1]. \tag{15.4}$$

Note that the polynomial $\mathcal{Q}_n$ is of degree $< 6n$, and that we need to construct a polynomial of degree $\leq n$. This construction is not difficult and we leave it as an exercise (see Exercise 15.2). The theorem is now proved. □

Exercise 15.1. *Prove that, for $j \neq i$,*

$$|I_j|^2 \leqslant 9|I_i||x_i-x_j|,$$

and

$$|x_i - x_{j-1}| < 4|x_i-x_j|,$$

Exercise 15.2. *Prove that* (15.4) *implies the statement of Theorem 15.1.*

15.1 Approximation by Bernstein polynomials: pointwise estimate

Theorem 15.2. *For any $f \in C[0,1]$ and $x \in [0,1]$, we have*

$$|f(x) - B_n(x,f)| \leqslant 2\omega\left(\sqrt{\frac{x(1-x)}{n}}, f\right). \tag{15.5}$$

Proof. First of all, since $B_n(0,f) = f(0)$ and $B_n(1,f) = f(1)$, it is sufficient to prove (15.5) for $x \in (0,1)$. Using properties of the Bernstein polynomials (see Section 7.3) and inequality (14.2) we have

$$\begin{aligned}
|f(x) - B_n(x,f)| &= \left|\sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \left[f(x) - f\left(\frac{k}{n}\right)\right]\right| \\
&\leq \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \omega\left(\left|x - \frac{k}{n}\right|\right) \\
&\leq \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \left(1 + \frac{\left|x-\frac{k}{n}\right|\sqrt{n}}{\sqrt{x(1-x)}}\right) \omega\left(\sqrt{\frac{x(1-x)}{n}}\right) \\
&= \omega\left(\sqrt{\frac{x(1-x)}{n}}\right)\left(1 + \frac{\sqrt{n}}{\sqrt{x(1-x)}} \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \left|x - \frac{k}{n}\right|\right) \\
&=: \omega\left(\sqrt{\frac{x(1-x)}{n}}\right)\left(1 + \frac{\sqrt{n}}{\sqrt{x(1-x)}} \Upsilon\right).
\end{aligned}$$

Recalling the Schwarz inequality (see (4.3) with $p = q = 2$)

$$\sum_{k=0}^{n} |x_k y_k| \leq \left(\sum_{k=0}^{n} x_k^2\right)^{1/2} \left(\sum_{k=0}^{n} y_k^2\right)^{1/2}$$

we now estimate Υ as follows

$$\begin{aligned}
\Upsilon &= \sum_{k=0}^{n} \left[\binom{n}{k} x^k (1-x)^{n-k}\right]^{1/2} \left\{\left|x - \frac{k}{n}\right| \left[\binom{n}{k} x^k (1-x)^{n-k}\right]^{1/2}\right\} \\
&\leq \left[\sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k}\right]^{1/2} \left[\sum_{k=0}^{n} \left|x - \frac{k}{n}\right|^2 \binom{n}{k} x^k (1-x)^{n-k}\right]^{1/2} \\
&= \left[\sum_{k=0}^{n} \left(x - \frac{k}{n}\right)^2 \binom{n}{k} x^k (1-x)^{n-k}\right]^{1/2} \\
&= \left[x^2 - 2x \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \left(\frac{k}{n}\right) + \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \left(\frac{k}{n}\right)^2\right]^{1/2} \\
&= \left[x^2 - 2x B_n(x,t) + B_n(x,t^2)\right]^{1/2} \\
&= \left[\frac{x(1-x)}{n}\right]^{1/2}.
\end{aligned}$$

This implies that

$$|f(x) - B_n(x,f)| \leq 2\omega\left(\sqrt{\frac{x(1-x)}{n}}\right),$$

and the proof is now complete. □

Exercise 15.3. *Prove that, for any $f \in C[0,1]$, we have*

$$\|f - B_n(\cdot, f)\|_{C[0,1]} \leqslant \frac{3}{2}\omega\left(\frac{1}{\sqrt{n}}, f\right).$$

16 Theorems of Stone, Müntz, Mergelyan

A generalization of Theorem 7.6 (the Weierstrass theorem) is the Stone (Stone-Weierstrass) theorem. Before we state it, several definitions are needed.

Definition 16.1. Let (X, ρ) be a metric space. A set A of real-valued continuous functions on X is called an *algebra* of continuous functions, if A is a linear vector space (*i.e.,* A is closed under addition and multiplication by constants), and if it is closed under multiplication (that is, if $f, g \in A$, then $fg \in A$).

Definition 16.2. An algebra A *vanishes nowhere* if not all of its elements simultaneously vanish at a point, that is, for any point $x \in X$ there is $f \in A$ satisfying $f(x) \neq 0$.

Definition 16.3. Algebra A separates points of X, if for any two different points $x_1, x_2 \in X$ there is $f \in A$ such that $f(x_1) \neq f(x_2)$.

Theorem 16.4 (Stone)**.** *Let (X, ρ) be a compact metric space. Algebra $A \subset C(X)$ is dense in $C(X)$ if and only if A vanishes nowhere in X, and A separates points of X.*

Example 16.1. Theorem 16.4 is, in general, not true for algebras A of complex valued functions. Let A be the set of all algebraic polynomials with complex coefficients in $|z| \leqslant 1$. Then A is a complex algebra, all conditions of Stone's theorem are satisfied, but A is not dense in $C(\{z : |z| \leqslant 1\})$, since the closure of this algebra is the set of analytic functions, but there are non-analytic functions in $C(\{z : |z| \leqslant 1\})$, for example, $f(z) = |z|$.

Remark 16.1. There is an analog of Stone's theorem for complex algebras (it involves the notion of antisymmetric sets).

Example 16.2. Let $X \subset \mathbb{R}^m$ be a closed bounded set, $\mathcal{P}$ be the set of all algebraic polynomials on X, that is elements of $\mathcal{P}$ are polynomials

$$p(x_1, \dots, x_m) = \sum_{j_1=0}^{l_1} \cdots \sum_{j_m=0}^{l_m} a_{j_1 \dots j_m} x_1^{j_1} \cdot \ldots \cdot x_m^{j_m}.$$

It is clear that $\mathcal{P}$ is an algebra that vanishes nowhere and separates points. Therefore, it satisfies the conditions of Stone's theorem. Hence, any continuous function on a closed bounded set can be approximated arbitrarily well by polynomials in m variables.

Example 16.3. Stone's theorem implies the Weierstrass theorem for approximation of 2π periodic functions by trigonometric polynomials. (For $[0, 2\pi)$, the points 0 and 2π are considered to be a single point on the circle e^{it}, which is a closed bounded set.)

Example 16.4. Consider the set of trigonometric polynomials of sines on $[0, \pi]$, *i.e.,*

$$\tau_n(t) := \beta_0 + \sum_{k=1}^{n} \beta_k \sin kt, \quad t \in [0, \pi].$$

This is an algebra, but it does not separate points 0 and π.

Remark 16.2. Stone's theorem does not cover all sets that are dense in the set of continuous functions. An example is given by the following Müntz theorem.

Theorem 16.5 (Müntz). *Suppose $0 = \lambda_0 < \lambda_1 < \lambda_2 < \cdots \to \infty$. The set of polynomials of the form*

$$\sum_{j=0}^{n} c_j \, x^{\lambda_j}, \quad c_j \in \mathbb{R}, \tag{16.1}$$

is dense in $C[0,1]$ if and only if

$$\sum_{j=1}^{\infty} \frac{1}{\lambda_j} = \infty.$$

Remark 16.3. Note that the condition $\lambda_0 = 0$ is needed Theorem 16.5 since x^λ vanishes at $x = 0$ if $\lambda > 0$, and so the set of polynomials of the form (16.1) is certainly not dense in $C[0,1]$ if $\lambda_j > 0$, for all $j \geq 0$.

Remark 16.4. An analog of the Weierstrass theorem for approximation of functions of complex variable was proved by Mergelyan.

Definition 16.6. Let $M \subset \mathbb{C}$. If $f : M \to \mathbb{C}$ is a complex valued function, then

$$f \in A(M) \iff \begin{cases} 1) & f \text{ is continuous on } M, \\ 2) & f \text{ is analytic at all interior points of } M. \end{cases}$$

Remark 16.5. If the interior of M is empty then $A(M)$ is just the set of all continuous functions on M (*i.e.,* the assumption about analyticity of f is vacuously satisfied).

Theorem 16.7 (Mergelyan). *Any function $f \in A(M)$ can be approximated arbitrarily well by algebraic polynomials on M if and only if M is bounded and closed set, and $\mathbb{C} \setminus M$ is a connected set.*

Remark 16.6. Note that the set M in the statement of Theorem 16.7 does not have to be connected.

17 Lagrange Polynomials. Finite and Divided Differences

Suppose that $m \in \mathbb{N}$, $\{x_0, \dots, x_m\} \subset \mathbb{R}$ is a set of $m+1$ distinct points in $\mathbb{R}$, and suppose that we know the values of a function f at these points, *i.e.,* $f(x_j)$, $0 \leq j \leq m$, are known.

17.1 Lagrange Polynomial

Definition 17.1. Lagrange polynomial $L(x, f) := L(x, f, x_0, \dots, x_m)$ of degree $\leqslant m$ interpolating f at the points $x_0, \dots, x_m$ is the algebraic polynomial of degree $\leqslant m$ such that $L(x_j, f) = f(x_j)$, $0 \leq j \leq m$.

Example 17.1. For $m = 0$, $L(x, f, x_0) \equiv f(x_0)$, and, for $m = 1$,

$$L(x, f, x_0, x_1) = \frac{x - x_0}{x_1 - x_0} f(x_1) + \frac{x - x_1}{x_0 - x_1} f(x_0) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0). \tag{17.1}$$

Definition 17.2. For a fixed m, the fundamental Lagrange polynomial is defined by

$$l_j(x) := \prod_{i=0, i \neq j}^{m} \frac{x - x_i}{x_j - x_i}, \quad 0 \leq j \leq m.$$

It is easy to see that

$$l_j(x_i) = \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases}$$

which implies that the Lagrange polynomial always exists and can be written as

$$L(x, f, x_0, \dots, x_m) = \sum_{j=0}^{m} f(x_j) l_j(x). \tag{17.2}$$

The fundamental theorem of algebra implies that the Lagrange polynomial of degree $\leq m$ interpolating a given f at the points $x_0, \dots, x_m$ is unique. Indeed, if there are two such polynomials, they would have to coincide at at least $m+1$ points x_j, $0 \leq j \leq m$, which is impossible because their difference is a non-trivial polynomial of degree $\leqslant m$ which would have $m+1$ zeros. So, for any polynomial p_m of degree $\leqslant m$, $L(x, p_m) \equiv p_m(x)$.

It is easy to see that the Lagrange polynomial depends linearly on the function that it interpolates, *i.e.*,

$$\begin{aligned} L(x, f+g) &= L(x, f) + L(x, g), \\ L(x, \alpha f) &= \alpha L(x, f), \quad \alpha \in \mathbb{R}. \end{aligned}$$

Remark 17.1. Fundamental polynomials l_j are often written in a different form. If $p(x) := (x - x_0) \dots (x - x_m)$, then

$$l_j(x) = \frac{p(x)}{(x - x_j) p'(x_j)},$$

and so (17.2) implies that

$$L(x, f, x_0, \dots, x_m) = \sum_{j=0}^{m} f(x_j) \frac{p(x)}{(x - x_j) p'(x_j)}. \tag{17.3}$$

17.2 Divided Differences

Consider the difference $f(x) - L(x, f, x_0, \dots, x_{m-1})$. We divide it by $(x - x_0) \dots (x - x_{m-1})$, and calculate the quotient at the point x_m.

Definition 17.3. The divided difference of order m of a function f at knots (points) $x_0, \dots, x_m$ is

$$[x_0, \dots, x_m; f] := \frac{f(x_m) - L(x_m, f, x_0, \dots, x_{m-1})}{(x_m - x_0) \dots (x_m - x_{m-1})}. \tag{17.4}$$

Therefore,

$$\begin{aligned} [x_0, \dots, x_m; f] \quad = \quad & \frac{f(x_0)}{(x_0 - x_1) \dots (x_0 - x_m)} + \frac{f(x_1)}{(x_1 - x_0)(x_1 - x_2) \dots (x_1 - x_m)} \\ & + \dots + \frac{f(x_m)}{(x_m - x_0) \dots (x_m - x_{m-1})}. \end{aligned} \tag{17.5}$$

Exercise 17.1. *Prove (17.5) (hint: use (17.4)).*

Remark 17.2. If $p(x) := (x - x_0) \dots (x - x_m)$, then

$$[x_0, \dots, x_m; f] = \sum_{k=0}^{m} \frac{f(x_k)}{p'(x_k)}.$$

Remark 17.3. Divided differences are symmetric with respect to the knots, *i.e.*, $[x_0, x_1; f] = [x_1, x_0; f]$, $[x_0, x_1, x_2; f] = [x_0, x_2, x_1; f] = \cdots = [x_2, x_1, x_0; f]$, and so on.

For completeness, we denote $[x_0; f] := f(x_0)$.

Theorem 17.4 (Newton's formula for the interpolating polynomial)**.** *Lagrange interpolating polynomial of degree m can be represented in the following form:*

$$\begin{aligned} L(x, f, x_0, \dots, x_m) &= [x_0; f] + [x_0, x_1; f](x - x_0) + \cdots + [x_0, \dots, x_m; f](x - x_0) \dots (x - x_{m-1}) \\ &= \sum_{k=0}^{m} [x_0, \dots, x_k; f](x - x_0) \dots (x - x_{k-1}). \end{aligned} \tag{17.6}$$

Proof. We use induction. The case $m = 1$ follows from (17.1). Assume now that (17.6) is valid for $m - 1$, and verify it for m. We need to prove that

$$L(x, f, x_0, \dots, x_m) = L(x, f, x_0, \dots, x_{m-1}) + [x_0, \dots, x_m; f](x - x_0) \dots (x - x_{m-1}). \tag{17.7}$$

On both sides of this equality we have polynomials of degree $\leqslant m$. Therefore, it is enough to check this equality at any $m + 1$ points and, in particular, at the points x_j, $0 \le j \le m$. If $j \ne m$, then, at x_j, expressions on both sides are equal to $f(x_j)$ by the definition of the Lagrange polynomial. If $j = m$, then, at x_m, this equality becomes our definition of the divided difference. □

Corollary 17.5. *Suppose that $f \in C[a, b]$ has m-th derivative on (a, b), $x_j \in [a, b]$, $0 \le j \le m$. There exists $\theta \in (a, b)$ such that*

$$[x_0, \dots, x_m; f] = \frac{f^{(m)}(\theta)}{m!}.$$

Proof. Consider $g(x) := f(x) - L(x, f, x_0, \dots, x_m)$, $x \in [a, b]$. Since $g(x_j) = 0$, $0 \le j \le m$, *i.e.*, g has at least $m+1$ zeroes on $[a, b]$, then Rolle's theorem implies that the m-th derivative of g has at least one zero on (a, b). Hence, there exists $\theta \in (a, b)$ such that $g^{(m)}(\theta) = 0$, and so using (17.6) we conclude that

$$0 = g^{(m)}(\theta) = f^{(m)}(\theta) - L^{(m)}(\theta, f, x_0, \dots, x_m) = f^{(m)}(\theta) - [x_0, \dots, x_m; f]m!.$$

□

Remark 17.4. If all the knots are "moved to one point", then the Lagrange polynomial "becomes" the Taylor polynomial for f at that point.

Corollary 17.6. *For a polynomial p_m of degree $\le m$ with the leading coefficient a_m, any divided difference of order m is equal to a_m,* i.e., *if $p_m(x) = a_m x^m + a_{m-1} x^{m-1} + \cdots + a_0$, then*

$$[x_0, \dots, x_m; p_m] = a_m.$$

Corollary 17.7. *For every polynomial p_{m-1} of degree $\leqslant m - 1$,*

$$[x_0, \dots, x_m; p_{m-1}] = 0.$$

Lemma 17.8. *The following equality holds:*

$$\frac{[x_0, \dots, x_{m-1}; f] - [x_1, \dots, x_m; f]}{x_0 - x_m} = [x_0, \dots, x_m; f]. \tag{17.8}$$

Proof. By the uniqueness of the Lagrange polynomial,

$$L(x, f, x_1, \dots, x_{m-1}, x_0, x_m) \equiv L(x, f, x_1, \dots, x_{m-1}, x_m, x_0).$$

Let $p(x) := (x - x_1) \cdot \dots \cdot (x - x_{m-1})$. By Newton's formula (Theorem 17.4),

$$\begin{aligned} L(x, f, x_1, \dots, x_{m-1}, x_0, x_m) &= L(x, f, x_1, \dots, x_{m-1}) \\ &+ [x_1, \dots, x_{m-1}, x_0; f]p(x) + [x_1, \dots, x_{m-1}, x_0, x_m; f]p(x)(x - x_0) \end{aligned} \tag{17.9}$$

and

$$\begin{aligned} L(x, f, x_1, \dots, x_{m-1}, x_m, x_0) &= L(x, f, x_1, \dots, x_{m-1}) \\ &+ [x_1, \dots, x_{m-1}, x_m; f]p(x) + [x_1, \dots, x_{m-1}, x_m, x_0; f]p(x)(x - x_m). \end{aligned} \tag{17.10}$$

Subtracting (17.9) from (17.10), and using symmetry of the divided difference, we obtain

$$0 = [x_0, \dots, x_{m-1}; f]p(x) - [x_1, \dots, x_m; f]p(x) - (x_0 - x_m)[x_0, \dots, x_m; f]p(x).$$

To finish the proof, we divide this equality by $p(x)$. □

Remark 17.5. Equality (17.8) is often used as a (recursive) definition of the divided difference.

Lemma 17.9 (representation of divided differences)**.** *If a function f is absolutely continuous on $[a, b]$ and $x_j \in [a, b]$, $0 \le j \le m$, then*

$$[x_0, \dots, x_m; f] = \int_0^1 [x_0, \dots, x_{m-1}; f_1]\, dt, \tag{17.11}$$

where

$$f_1(u) := f_1(u, t) := f'(x_m + t(u - x_m)).$$

Proof. First, we verify (17.11) for $m = 1$. Since f is absolutely continuous on $[a, b]$, Lebesgue's theorem implies

$$f(x_1) - f(x_0) = \int_{x_0}^{x_1} f'(x)\, dx.$$

After the substitution $x = x_1 + t(x_0 - x_1)$, we get

$$[x_0, x_1; f] = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{1}{x_1 - x_0} \int_{x_0}^{x_1} f'(x)\, dx = \int_0^1 f'(x_1 + t(x_0 - x_1)\, dt = \int_0^1 [x_0; f_1]\, dt.$$

Now, proceeding by induction we assume that the lemma is true for $m - 1 \geqslant 1$. Then, taking (17.8) into account we have

$$\begin{aligned} [x_0, \dots, x_m; f] &= \frac{[x_1, \dots, x_m; f] - [x_0, x_2, \dots, x_m; f]}{x_1 - x_0} \\ &= \int_0^1 \frac{[x_1, \dots, x_{m-1}; f_1] - [x_0, x_2, \dots, x_{m-1}; f_1]}{x_1 - x_0}\, dt = \int_0^1 [x_0, \dots, x_{m-1}; f_1]\, dt, \end{aligned}$$

which is (17.11) for m, which completes the proof of the lemma. □

Theorem 17.10 (representation of divided differences)**.** *If f has the $(m-1)$-st absolutely continuous derivative on $[a, b]$ and $x_j \in [a, b]$, $0 \le j \le m$, then the divided difference can be expressed as*

$$[x_0, \dots, x_m; f] = \int_0^1 \int_0^{t_1} \cdots \int_0^{t_{m-1}} f^{(m)}(x_0 + t_1(x_1 - x_0) + \dots + t_m(x_m - x_{m-1}))\, dt_m \dots dt_1.$$

Exercise 17.2. *Prove this theorem (hint: use Lemma 17.9).*

Exercise 17.3. *Prove that, if f has the $(m-1)$-st absolutely continuous derivative on $[a,b]$, $|f^{(m)}(x)| \le 1$ for almost all $x \in [a,b]$, and $x_j \in [a,b]$, $0 \le j \le m$, then*

$$|[x_0, \dots, x_m; f]| \le \frac{1}{m!}.$$

Exercise 17.4. *Prove that, if f has the $(m-1)$-st absolutely continuous derivative on $[a,b]$, $|f^{(m)}(x)| \le 1$ for almost all $x \in [a,b]$, and $x_j \in [a,b]$, $0 \le j \le m-1$, then*

$$|f(x) - L(x, f, x_0, \dots, x_{m-1})| \le \frac{1}{m!} \left|(x - x_0) \dots (x - x_{m-1})\right|.$$

17.3 Divided Differences and Lagrange-Hermite Polynomials

Everywhere below, it is convenient to denote

$$f^{(0)} := f.$$

Suppose that we have $(p+1)$ different points $y_0, \dots, y_p$, and, at every point y_j, we know not just the values of a function f, *i.e.*, the numbers $f(y_j) = f^{(0)}(y_j)$, but also values of the first q_j derivatives (we assume that f has appropriate smoothness at these points). In other words, suppose that the following values are given:

$$f^{(i)}(y_j), \quad 0 \le i \le q_j, \quad 0 \le j \le p.$$

Denote

$$m := p + \sum_{j=0}^{p} q_j.$$

Definition 17.11. The Lagrange-Hermite interpolation polynomial

$$L(x, f) := L(x, f, (y_0, q_0), \dots, (y_p, q_p))$$

is the algebraic polynomial of degree $\leqslant m$ such that

$$L^{(i)}(y_j, f) = f^{(i)}(y_j), \quad 0 \le i \le q_j, \quad 0 \le j \le p.$$

Note that the Lagrange-Hermite polynomial becomes the Lagrange polynomial if $q_j = 0$, for all j, *i.e.*, $L(x, f, (y_0, 0), \dots, (y_m, 0) = L(x, f, y_0, \dots, y_m)$.

Exercise 17.5. *Prove that the Lagrange-Hermite interpolation polynomial always exists and is unique.*
Hint: If the polynomial has the form $P(x) = \sum_{i=0}^{m} a_i x^i$, then one can find $m+1$ unknowns a_i by solving the system of $m+1$ linear equations $P^{(\nu)}(y_j) = f^{(\nu)}(y_j)$, $0 \le \nu \le q_j$, $0 \le j \le p$. If $f^{(\nu)}(y_j) = 0$, for all $0 \le \nu \le q_j$ and $0 \le j \le p$, prove that $P(x)$ has to be identically equal to zero. What can you now say about the coefficient matrix of the system?

Definition 17.12. The generalized divided difference of order m of a function f is the number

$$[(y_0, q_0), \dots, (y_p, q_p); f] := \frac{1}{q_0! \cdot \dots \cdot q_p!} \frac{\partial^{q_0 + \dots + q_p}}{\partial y_0^{q_0} \dots \partial y_p^{q_p}} [y_0, \dots, y_p; f].$$

In particular, $[(y_0, 0), \dots, (y_m, 0); f] = [y_0, \dots, y_m; f]$.

Theorem 17.13. *For $x \neq y_j$, $0 \leq j \leq p$,*

$$[(y_0,q_0),\dots,(y_p,q_p),(x,0);f] = \frac{f(x) - L(x,f,(y_0,q_0),\dots,(y_p,q_p))}{(x-y_0)^{q_0+1}\dots(x-y_p)^{q_p+1}}.$$

Exercise 17.6. *Prove this theorem.*

Remark 17.6. For the generalized divided difference, we have analogs of all relationships proved above for the regular divided differences.

17.4 Finite Differences

Suppose that $m+1$ points $x_0,\dots,x_m$ are equidistant, *i.e.*, for some $h>0$, we have $x_j = x_0 + jh$, $0 \leq j \leq m$. Denote

$$f(x_m) - L(x_m, f, x_0,\dots,x_{m-1}) = \sum_{j=0}^{m} (-1)^{m-j}\binom{m}{j} f(x_0+jh) =: \Delta_h^m(f,x_0).$$

It is also convenient to denote $\Delta_h^0(f,x_0) := f(x_0)$ and $\Delta_0^m(f,x_0) := 0$.

Definition 17.14. The expression $\Delta_h^m(f,x_0)$ is called the m-th (finite or forward) difference of the function f with the step h at the point x_0.

Lemma 17.15.

$$\Delta_h^m(f,x_0) = h^m m! [x_0,\dots,x_m,f].$$

Proof. The statement of the lemma immediately follows from the the definition of the divided difference since

$$[x_0,\dots,x_m;f] = \frac{f(x_m) - L(x_m,f,x_0,\dots,x_{m-1})}{(x_m-x_0)\dots(x_m-x_{m-1})} = \frac{\Delta_h^m(f,x_0)}{h^m m!}.$$

□

Lemma 17.15 implies that finite differences have the same properties as divided differences.

Corollary 17.16. *The following identity holds:*

$$\Delta_h^{m-1}(f,x_0+h) - \Delta_h^{m-1}(f,x_0) = \Delta_h^m(f,x_0).$$

Corollary 17.17. *Suppose that $f \in C[a,b]$ has m-th derivative on (a,b), and $[x_0,x_0+mh] \subset [a,b]$. There exists $\theta \in (x_0,x_0+mh)$ such that*

$$\Delta_h^m(f,x_0) = h^m f^{(m)}(\theta).$$

Corollary 17.18. *For a polynomial $p_m(x) = a_m x^m + a_{m-1}x^{m-1} + \dots + a_0$ and any $x_0 \in \mathbb{R}$, we have*

$$\Delta_h^m(p_m,x_0) = a_m h^m m!.$$

Corollary 17.19. *For every polynomial p_{m-1} of degree $\leqslant m-1$ and any $x_0 \in \mathbb{R}$,*

$$\Delta_h^m(p_{m-1},x_0) = 0.$$

Corollary 17.20. *If f has the $(m-1)$-st absolutely continuous derivative on $[a,b]$ and $[x_0, x_0 + mh] \subset [a,b]$, then the following identities are valid:*

$$\Delta_h^m(f, x_0) = h^m m! \int_0^1 \int_0^{t_1} \cdots \int_0^{t_{m-1}} f^{(m)}(x_0 + h(t_1 + \cdots + t_m))\, dt_m \ldots dt_1$$

and

$$\Delta_h^m(f, x_0) = \int_0^h \int_0^h \cdots \int_0^h f^{(m)}(x_0 + t_1 + \cdots + t_m)\, dt_m \ldots dt_1. \tag{17.12}$$

Exercise 17.7. *Prove identity (17.12) (hint: use induction and Corollary 17.16).*

Corollary 17.21. *If f has the $(m-1)$-st absolutely continuous derivative on $[a,b]$, $|f^{(m)}(x)| \le 1$ for almost all $x \in [a,b]$, and $[x_0, x_0 + mh] \subset [a,b]$, then*

$$|\Delta_h^m(f, x_0)| \le h^m.$$

18 Moduli of Smoothness

Definition 18.1. The (classical) modulus of smoothness of order k of a function $f \in C[a,b]$ is the function $\omega_k : [0,\infty) \to \mathbb{R}$ defined by

$$\omega_k(t, f, [a,b]) := \sup_{h\in[0,t]} \sup_{x\in[a,b-kh]} |\Delta_h^k(f,x)|,$$

for $t \in [0, (b-a)/k]$, and

$$\omega_k(t, f, [a,b]) := \omega_k((b-a)/k, f, [a,b]),$$

for $t > (b-a)/k$.

Remark 18.1. The modulus of smoothness of order 1 of a function is exactly its modulus of continuity defined in Section 14.

Exercise 18.1. *Evaluate/estimate the k-th moduli of smoothness of functions from Exercise 14.1.*

Lemma 18.2. *The kth modulus of smoothness $\omega_k(t) = \omega_k(t, f, [a,b])$ of $f \in C[a,b]$ has the following properties:*

1) $\omega_k(0) = 0$, $\omega_k(0+) := \lim_{t\to 0^+} \omega_k(t) = 0$;

2) ω_k *is non-decreasing on* $[0,\infty)$;

3) ω_k *is continuous on* $[0,\infty)$;

4) $\omega_k(nt) \leqslant n^k \omega_k(t)$, *for all* $n \in \mathbb{N}$ *and* $t \geqslant 0$.

Proof.

1) The fact that $\omega_k(0) = 0$ follows from the definition, and so we only need to show that $\omega_k(0+) = 0$. Indeed,

$$|\Delta_h^k(f,x)| = |\Delta_h^{k-1}(f, x+h) - \Delta_h^{k-1}(f,x)| \leqslant \ldots \leqslant 2^{k-1}\omega_1(h, f, [a,b]),$$

and so the fact that $\omega_k(0+) = 0$ follows from Lemma 14.2.

2) This property immediately follows from the definition.

4) This property follows from the following formula which can be proved by induction

$$\Delta_{nh}^k(f,x) = \sum_{j_1=0}^{n-1} \cdots \sum_{j_k=0}^{n-1} \Delta_h^k(f, x + h(j_1 + \cdots + j_k)). \tag{18.1}$$

□

Exercise 18.2. *Prove the identity (18.1).*

Exercise 18.3. *Prove Property 3) in Lemma 18.2,* i.e., *prove that ω_k is continuous on $[0, \infty)$.*

Exercise 18.4. *Prove that, if $f(x) := x^k$, $a \le x \le b$, then*

$$\omega_k(t, f, [a,b]) = k! \min \left\{ t^k, \left(\frac{b-a}{k} \right)^k \right\}.$$

Exercise 18.5. *Show that, in general, the modulus $\omega_k(t)$ is not a subadditive function if $k \ge 2$.*

19 Whitney's Inequality

Recall that if $f = p_{k-1}$ is a polynomial of degree $\leqslant k-1$, then $\Delta_h^k(p_{k-1}, x) \equiv 0$.

For a given function $f \in C[a,b]$, denote by p_{k-1}^* the polynomial of best approximation of f on $[a,b]$. Then

$$\begin{aligned} |\Delta_h^k(f,x)| = |\Delta_h^k(f - p_{k-1}^*, x)| &= \left| \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} (f(x+jh) - p_{k-1}^*(x+jh)) \right| \\ &\leqslant E_{k-1}(f)_{C[a,b]} \sum_{j=0}^{k} \binom{k}{j} = 2^k E_{k-1}(f)_{C[a,b]}, \end{aligned}$$

which means that

$$\omega_k((b-a)/k, f, [a,b]) \leqslant 2^k E_{k-1}(f)_{C[a,b]}.$$

H. Whitney proved the converse inequality.

Theorem 19.1. *For every $k \in \mathbb{N}$, there exists a constant $W(k)$ such that for any $f \in C[a,b]$ we have*

$$E_{k-1}(f)_{[a,b]} \leqslant W(k) \omega_k((b-a)/k, f, [a,b]).$$

Remark 19.1. It is easy to verify Theorem 19.1 for $k = 1$:

$$E_0(f)_{[a,b]} \leqslant \max_{x \in [a,b]} |f(x) - f(a)| \leqslant \omega(b-a, f, [a,b]).$$

We now prove it for $k = 2$ and $[a,b] = [0,1]$. Let l be the linear function that interpolates f at the points 0 and 1, define $g(x) := f(x) - l(x)$, $x \in [0,1]$, and note that $g(0) = g(1) = 0$. It is clear that

$$E_1(f) \leqslant \|g\|_{[0,1]}.$$

Let x^* be a point where $|g(x^*)|$ attains the largest value on $[0,1]$. Without loss of generality, we can assume that $g(x^*) \geqslant 0$ and $0 \leqslant x^* \leqslant 1/2$. Then

$$2g(x^*) = (-g(0)+2g(x^*)-g(2x^*))+g(0)+g(2x^*) \leqslant \omega_2(x^*, g, [0,1])+g(2x^*) \leqslant \omega_2(x^*, g, [0,1])+g(x^*),$$

and so $g(x^*) \leqslant \omega_2(x^*, g, [0,1])$. By the linearity of $g - f$, we now have

$$g(x^*) \leqslant \omega_2(x^*, g, [0,1]) = \omega_2(x^*, f, [0,1]) \leqslant \omega_2(1/2, f, [0,1]).$$

Lemma 19.2. *Let $f \in C[0,1]$, and let F be an antiderivative of f, $x_j = jh$, $0 \le j \le k$, $h = k^{-1}$. Then, for any $x \in [0,1]$,*

$$|F(x) - L(x, F, x_0, \dots, x_k)| \leqslant c(k)\omega_k(1/k, f, [0,1]).$$

Proof. Denote $p(x) := (x - x_0)\dots(x - x_k)$, $x \in \mathbb{R}$. Since

$$\frac{F(c) - F(d)}{c - d} = \int_0^1 f(c + (d - c)t)dt,$$

we have

$$\begin{aligned} F(x) - L(x, F, x_0, \dots, x_k) &= F(x)L(x, 1, x_0, \dots, x_k) - L(x, F, x_0, \dots, x_k) \\ &= p(x) \sum_{j=0}^{k} \frac{F(x) - F(x_j)}{(x - x_j)p'(x_j)} \\ &= p(x) \sum_{j=0}^{k} \int_0^1 f(x + (x_j - x)t)dt \cdot \frac{(-1)^{k-j}}{j!(k-j)!h^k} \\ &= \frac{p(x)}{k!h^k} \int_0^1 \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} f(x(1-t) + jht)dt \\ &= \frac{p(x)}{k!h^k} \int_0^1 \Delta_{ht}^k(f, x(1-t))dt. \end{aligned}$$

Hence, for $c(k) := \dfrac{k^k}{k!} \|p\|_{C[0,1]}$,

$$|F(x) - L(x, F, x_0, \dots, x_k)| \leqslant c(k) \int_0^1 \omega_k(t/k, f, [0,1])dt \leqslant c(k)\omega_k(1/k, f, [0,1]).$$

□

Proof of Theorem 19.1. Without loss of generality, assume that $[a, b] = [0, 1]$. Let F and L be as in Lemma 19.2, *i.e.*, let F be an antiderivative of f and $L(x) := L(x, F, 0, 1/k, \dots, (k-1)/k, 1)$.

Denoting

$$G(x) := F(x) - L(x), \quad g(x) := G'(x) = f(x) - L'(x), \quad x \in [0,1],$$

and

$$\omega_k(1/k, g, [0,1]) = \omega_k(1/k, f, [0,1]) =: \omega,$$

we have by Lemma 19.2 that

$$\|G\|_{[0,1]} \leqslant c(k)\omega_k(1/k, f, [0,1]) = c(k)\omega. \tag{19.1}$$

We need to prove $E_{k-1}(f) \leqslant W(k)\omega$. Since $E_{k-1}(f) \leqslant \|G'\|_{[0,1]}$, it is enough to check that $\|G'\|_{[0,1]} \leqslant W(k)\omega$. We will show how this follows from (19.1).

Let $x \in [0,1/2]$ (the proof is similar if $x \in (1/2,1]$). If $\delta := 1/(2k)$, then $x+\delta k \in [0,1]$, and denote

$$\int_0^1 \Delta_{\delta t}^k(g,x)dt =: I.$$

Since $0 \leqslant \delta t \leqslant 1/k$, we have $|I| \leqslant \omega$. Also,

$$\begin{aligned} I &= \sum_{j=0}^k (-1)^{k-j}\binom{k}{j}\int_0^1 g(x+j\delta t)dt \\ &= (-1)^k g(x) + \sum_{j=1}^k (-1)^{k-j}\binom{k}{j}\frac{1}{j\delta}(G(x+j\delta)-G(x)) \\ &=: (-1)^k g(x) + A. \end{aligned}$$

Inequality (19.1) implies that

$$|A| \leqslant \sum_{j=1}^k \binom{k}{j}\frac{2}{j\delta}\|G\|_{[0,1]} \leqslant 2\omega\cdot c(k)\cdot 2k\cdot 2^k = c_*(k)\omega,$$

and so

$$|G'(x)| = |g(x)| \leq |I-A| \leq (c_*(k)+1)\omega.$$

Hence, $E_{k-1}(f) \leqslant W(k)\omega$, where $W(k) = c_*(k)+1$. □

Corollary 19.3. *Let $f \in C[a,b]$, and let $x_0 \in [a,b]$ and $h>0$ be such that $[x_0, x_0+(k-1)h] \subset [a,b]$. Then for the function*

$$g(x) := f(x) - L(x,f,x_0,x_0+h,\dots,x_0+(k-1)h)$$

we have the inequality

$$|g(x)| \leqslant c(k)\omega_k(h,f,[a,b])\left(1+\frac{|x-x_0|}{h}\right)^{2k-1}, \quad \textit{for all } x \in [a,b]. \tag{19.2}$$

Remark 19.2. One can show that $2k-1$ can be replaced by k in the last inequality.

Proof of the Corollary 19.3. Denote $\omega := \omega_k(h,f,[a,b])$, and let $p^*(x,I)$ be the polynomial of best approximation to f of degree $\leqslant k-1$ on the interval $I \subset [a,b]$. Theorem 19.1 implies that

$$\|f-p^*(\cdot,I)\|_{C(I)} \leqslant c(k)\omega_k(|I|/k,f,I) \leqslant c(k)\omega_k(|I|/k,f,[a,b]).$$

(Recall that $|I|$ denotes the length of the interval I.) Part 4) of Lemma 18.2 implies that

$$\omega_k(|I|/k,f,[a,b]) \leqslant ([|I|/(kh)]+1)^k\,\omega,$$

and so

$$\|f-p^*(\cdot,I)\|_{C(I)} \leqslant c(k)\left(\max\{|I|/h,1\}\right)^k\omega. \tag{19.3}$$

We consider two cases depending on the location of the point x.

First, suppose that $x \in [x_0, x_0+(k-1)h]$, and let $q(x) := p^*(x, [x_0, x_0+(k-1)h])$. Then,

$$\|f-q\|_{C[x_0,x_0+(k-1)h]} \leqslant c\omega,$$

and therefore

$$\begin{aligned}|g(x)| &= |(f(x)-q(x)) - L(x, f-q, x_0, \dots, x_0+(k-1)h)| \\ &\leqslant c\omega + c\omega \sum_{j=0}^{k-1} \prod_{i=0, i\neq j}^{k-1} \frac{|x-x_i|}{|x_j-x_i|} \leqslant c\omega \left(1 + k\frac{((k-1)h)^{k-1}}{h^{k-1}}\right) \\ &=: c^*\omega,\end{aligned}$$

which implies the required inequality.

Suppose now that $x \in [a,b] \setminus [x_0, x_0+(k-1)h]$, for example $x > x_0+(k-1)h$ (the case for $x < x_0$ is being analogous). In this case, we let $q(y) := p^*(y, [x_0, x])$ and note that (19.3) implies that

$$\|f-q\|_{C[x_0,x]} \leqslant c(k) \left(\max\{|x-x_0|/h, 1\}\right)^k \omega.$$

Therefore,

$$\begin{aligned}|g(x)| &= |(f(x)-q(x)) - L(x, f-q, x_0, \dots, x_0+(k-1)h)| \\ &\leqslant \|f-q\|_{C[x_0,x]} + \|f-q\|_{C[x_0,x]} \sum_{j=0}^{k-1} \prod_{i=0, i\neq j}^{k-1} \frac{|x-x_i|}{|x_j-x_i|} \\ &\leqslant \|f-q\|_{C[x_0,x]} \left(1 + \frac{|x-x_0|^{k-1}}{h^{k-1}} k\right) \\ &\leqslant c(k) \left(1 + \frac{|x-x_0|}{h}\right)^{2k-1} \omega.\end{aligned}$$

□

Remark 19.3. In the proof of Theorem 15.1, we used the inequality

$$|f(x) - f(x_0)| \leqslant c\omega_1(h) \left(1 + \frac{|x-x_0|}{h}\right).$$

(Note that $f(x_0) = L(x, f, x_0)$.) If we replace this inequality (19.2) then, repeating the same arguments, one can prove a direct theorem with the estimate in terms of ω_k.

Theorem 19.4. *If $f \in C[-1,1]$, then*

$$E_n(f) \leqslant c(k)\omega_k(1/n, f, [-1,1]), \quad n \geqslant k-1.$$

An improvement of this estimate is given in the following theorem.

Theorem 19.5 (Timan ($k=1$), Dzyadyk ($k=2$), Brudnyi ($k>2$))**.** *If $f \in C[-1,1]$, then for any $n \geqslant k-1$, there exists a polynomial p_n of degree $\leqslant n$ such that*

$$|f(x) - p_n(x)| \leqslant c(k)\omega_k\left(\frac{1}{n^2} + \frac{1}{n}\sqrt{1-x^2}, f, [-1,1]\right).$$

Exercise 19.1. *Prove Theorems 19.4 and 19.5.*

Remark 19.4. Theorem 19.5 establishes a "correct" pointwise estimate for approximation by polynomials (which is better towards the endpoints of the interval) since the corresponding converse theorems have been proved. For a "correct" uniform estimate, the so-called Ditzian-Totik moduli of smoothness are used. These moduli measure smoothness differently inside an interval and near the endpoints.

Definition 19.6. For $r \in \mathbb{N}$, denote by W^r the space of continuous function on $[-1,1]$ that have $(r-1)$-st absolutely continuous derivative on $[-1,1]$ and $\|f^{(r)}\|_{L_\infty[-1,1]} < +\infty$.

Lemma 19.7. *If $f \in W^r$, then*

$$\omega_r(t, f, [-1,1]) \leqslant t^r \left\| f^{(r)} \right\|_{L_\infty[-1,1]}, \quad t \geqslant 0.$$

Lemma 19.8. *If, for some $f \in C[-1,1]$ and all $t > 0$, the estimate*

$$\omega_r(t, f, [-1,1]) \leqslant t^r,$$

holds, then $f \in W^r$, and $\|f^{(r)}\|_{L_\infty[-1,1]} \leqslant 1$.

Exercise 19.2. *Prove Lemmas 19.7 and 19.8.*

An immediate corollary of Theorem 19.4 is the following result for $f \in W^r$.

Corollary 19.9. *If $f \in W^r$, then*

$$E_n(f) \leqslant c(r) \frac{\|f^{(r)}\|_{L_\infty[-1,1]}}{n^r}, \quad n \geqslant r-1.$$

20 Trigonometric Polynomial Kernels

The following polynomial kernels are well known: Dirichlet's, Fejér's, Jackson's, de La Vallée Poussin's, Bernstein's, Rogozinski's, Poisson's, conjugate to Poisson's kernels, and others.

20.1 Dirichlet's Kernel

Definition 20.1. Dirichlet's kernel is the trigonometric polynomial

$$D_n(t) := \frac{1}{2} + \cos(t) + \cdots + \cos(nt) = \frac{\sin\left(\dfrac{2n+1}{2} t\right)}{2\sin\dfrac{t}{2}}, \quad t \in \mathbb{R}.$$

Exercise 20.1. *Verify the last identity.*

Exercise 20.2. *Prove that for a periodic $f \in L_1[0, 2\pi]$ the partial sum of the Fourier series $S_n(\cdot, f)$ can be expressed as*

$$S_n(t, f) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t-x) D_n(x)\, dx, \quad t \in \mathbb{R}.$$

20.2 Fejér's Kernel

Definition 20.2. Fejér's kernel is the arithmetic mean of the first n Dirichlet's kernels

$$\begin{aligned}F_n(t) &:= \frac{D_0(t)+\cdots+D_{n-1}(t)}{n}\\ &= \frac{1}{2}+(1-1/n)\cos(t)+(1-2/n)\cos(2t)+\cdots+1/n\cos((n-1)t)\\ &= \frac{\sin^2\dfrac{nt}{2}}{2n\sin^2\dfrac{t}{2}},\quad t\in\mathbb{R}.\end{aligned}$$

Exercise 20.3. *Prove these identities.*

Fejér proved that for a 2π-periodic $f\in C[0,2\pi]$ and

$$\sigma_n(t,f) := \frac{1}{\pi}\int_{-\pi}^{\pi} f(t-x)F_n(x)\,dx,\quad t\in\mathbb{R},$$

we always have

$$\|f-\sigma_n(\cdot,f)\|_{C[0,2\pi]}\to 0,\quad n\to\infty.$$

If $0<\alpha<1$ and $\omega_1(t,f,[0,2\pi])\leqslant t^\alpha$ for $t>0$, then one can show that

$$\|f-\sigma_n(\cdot,f)\|_{C[0,2\pi]}\leqslant c(\alpha)\frac{1}{n^\alpha},\quad n\geqslant 1.$$

For $\alpha=1$ this is not true.

20.3 Jackson's Kernels

Definition 20.3. Jackson's kernel is the trigonometric polynomial $J_n(t) := \frac{1}{\gamma_n}F_n^2(t)$, $t\in\mathbb{R}$, where F_n is the Fejér's kernel, and $\gamma_n = \frac{1}{\pi}\int_{-\pi}^{\pi} F_n^2(t)\,dt$.

Jackson proved that for

$$\mu_n(t,f) := \frac{1}{\pi}\int_{-\pi}^{\pi} J_n(t-x)f(x)\,dx,\quad t\in\mathbb{R},$$

if, for some $0<\alpha\leqslant 1$, $\omega_1(t,f)\leqslant t^\alpha$, $t>0$, then

$$\|f-\mu_n(\cdot,f)\|_{C[0,2\pi]}\leqslant c\frac{1}{n^\alpha}.$$

Moreover, if for some $0<\alpha\leqslant 2$, $\omega_2(t,f)\leqslant t^\alpha$, $t>0$, then the same estimate holds (this was proved by Zygmund).

Jackson also proved that, for $f\in C^r[0,2\pi]$,

$$\|f-\mu_n(\cdot,f)\|_{C[0,2\pi]}\leqslant c(r)\frac{\omega_1(1/n,f^{(r)})}{n^r},\quad n\geqslant r-1.$$

20.4 Stechkin's kernels

Stechkin introduced the generalized Jackson's kernels.

Definition 20.4. Let $m \in \mathbb{N}$. The generalized Jackson's kernel is defined as

$$J_{n,m}(t) := \frac{1}{\gamma_{n,m}} F_n^{2m}(t), \quad t \in \mathbb{R},$$

where

$$\gamma_{n,m} := \frac{1}{\pi} \int_{-\pi}^{\pi} F_n^{2m}(t)\, dt.$$

We will consider several properties of these kernels.

Property 1. *$J_{n,m}$ is a trigonometric polynomial of order $2m(n-1)$.*

Property 2. *$J_{n,m}$ has a "steeper peak" and "faster descend" than the Fejér's kernel.*

Property 3. $\frac{1}{\pi} \int_{-\pi}^{\pi} J_{n,m}(t)\, dt = 1.$

Property 4. *The following estimates hold:*

$$c_1(m) n^{2m-1} \leqslant \gamma_{n,m} \leqslant c_2(m) n^{2m-1}.$$

Proof. We have

$$\gamma_{n,m} = \frac{2}{\pi} \int_0^{\pi} F_n^{2m}(t)\, dt = \frac{2}{\pi} \int_0^{\pi/n} F_n^{2m}(t)\, dt + \frac{2}{\pi} \int_{\pi/n}^{\pi} F_n^{2m}(t)\, dt.$$

Taking into account that $\sin u \leq u$ for $u \geqslant 0$, and that $\sin u \geqslant 2u/\pi$ for $u \in [0, \pi/2]$, we obtain

$$\begin{aligned} \gamma_{n,m} &\geqslant \frac{2}{\pi} \frac{1}{(2n)^{2m}} \int_0^{\pi/n} \left(\frac{\sin n\frac{t}{2}}{\sin \frac{t}{2}} \right)^{4m} dt \\ &\geqslant \frac{c(m)}{n^{2m}} \int_0^{\pi/n} \left(\frac{nt/\pi}{t/2} \right)^{4m} dt =: c_1(m) n^{2m-1}. \end{aligned}$$

Also,

$$\begin{aligned} \gamma_{n,m} &\leqslant \frac{c(m)}{n^{2m}} n^{4m-1} + \frac{c(m)}{n^{2m}} \int_{\pi/n}^{\pi} \left(\frac{1}{\sin t/2} \right)^{4m} dt \\ &\leqslant c(m) n^{2m-1} + \frac{c(m)}{n^{2m}} \int_{\pi/n}^{\infty} \left(\frac{\pi}{t} \right)^{4m} dt = c_2(m) n^{2m-1}. \end{aligned}$$

□

Property 5. *For each $0 \leq i \leq 4m-2$,*

$$\int_{\pi/n}^{\pi} t^i J_{n,m}(t)\, dt \leqslant c(m) \frac{1}{n^i}.$$

Exercise 20.4. *Prove Property 5.*

21 Stechkin's Theorem (Jackson – Zigmund – Akhiezer – Stechkin)

Theorem 21.1 (Stechkin). *If $k \in \mathbb{N}$ and f is a 2π-periodic continuous function, then*

$$\widetilde{E}_n(f) \leqslant c(k)\omega_k(1/n, f), \quad n \geqslant 1.$$

Remark 21.1. For $k = 1$ this is Jackson's theorem, for $k = 2$, $\omega_2(t, f) \leqslant t$, $t \geqslant 0$, this is Zygmund's theorem, for $k = 2$ this is Akhiezer's theorem, for $k > 2$ this is Stechkin's theorem.

For $f \in C^r[0, 2\pi]$ we have

$$\omega_{r+1}(t, f) \leqslant c(r)t^r\omega_1(t, f^{(r)}), \quad t \geqslant 0.$$

Therefore, as a corollary of the Stechkin's theorem we obtain the following

Theorem 21.2 (second Jackson's inequality). *If $f \in C^r[0, 2\pi]$, then*

$$\widetilde{E}_n(f) \leqslant c(r)\frac{1}{n^r}\omega_1(1/n, f^{(r)}), \quad n \geqslant 1.$$

Corollary 21.3. *If $f \in W^r$, then $\omega_r(t, f) \leqslant t^r \left\|f^{(r)}\right\|_{L_\infty[0,2\pi]}$, therefore*

$$\widetilde{E}_n(f) \leqslant \frac{c(r)}{n^r}\left\|f^{(r)}\right\|_{L_\infty[0,2\pi]}, \quad n \geqslant 1.$$

Lemma 21.4. *For $m = k$ there exists a trigonometric polynomial T_n of order $2m(n-1)$ such that*

$$\|f - T_n\|_{C(\mathbb{R})} \leqslant c(k)\omega_k(\pi/n, f), \quad n \geqslant 1.$$

Proof. Take

$$(-1)^k S(x) := \frac{1}{\pi}\int_{-\pi}^{\pi} \Delta_{-t}^k(f, x)J_{n,m}(t)dt, \quad x \in \mathbb{R}.$$

We will prove that $S = f - T_n$, and $\|S\|_{C(\mathbb{R})} \leqslant c(k)\omega_k(\pi/n, f)$. Indeed,

$$\begin{aligned}
|S(x)| &\leqslant \frac{1}{\pi}\int_{-\pi}^{\pi} \omega_k(|t|, f)J_{n,m}(t)dt = \frac{2}{\pi}\int_{0}^{\pi} \omega_k(|t|, f)J_{n,m}(t)dt \\
&= \frac{2}{\pi}\left(\int_0^{\pi/n} \omega_k(|t|, f)J_{n,m}(t)dt\,dt + \int_{\pi/n}^{\pi} \omega_k(|t|, f)J_{n,m}(t)dt\,dt\right) \\
&\leqslant \frac{2}{\pi}\omega_k(\pi/n, f)\int_0^{\pi} J_{n,m}(t)dt + \frac{2}{\pi}\int_{\pi/n}^{\pi}\left(\frac{|t|}{\pi/n} + 1\right)^k \omega_k(\pi/n, f)J_{n,m}(t)\,dt \\
&\leqslant \omega_k(\pi/n, f) + n^k\frac{2^{k+1}}{\pi^{k+1}}\omega_k(\pi/n, f)\int_{\pi/n}^{\pi} t^k J_{n,m}(t)\,dt \\
&\leqslant \omega_k(\pi/n, f)(1 + (2/\pi)^{k+1}c),
\end{aligned}$$

where $c = c(m)$ is the constant from Property 5 of $J_{n,m}$.

We have

$$(-1)^k\Delta_{-t}^k(f, x) = f(x) - \binom{k}{1}f(x-t) + \cdots + (-1)^k f(x - kt),$$

and take $(-1)^k S(x) = f(x) - g(x)$, $x \in \mathbb{R}$. It remains to show that g is a trigonometric polynomial of an appropriate order, that is

$$\int_{-\pi}^{\pi} f(x-jt)J_{n,m}(t)\,dt, \quad j \neq 0,$$

is a trigonometric polynomial of order $2m(n-1)$. Since $J_{n,m}$ is an even trigonometric polynomial and is a sum of cosines, it is sufficient to show that

$$\int_{-\pi}^{\pi} f(x-jt)\cos(lt)dt$$

is a trigonometric polynomial of the necessary order. □

Exercise 21.1. *Let f be a 2π-periodic continuous function, $j, l \in \mathbb{N}$. Prove that*

$$\int_{-\pi}^{\pi} f(x-jt)\cos(lt)dt$$

is a trigonometric polynomial of necessary order $\leqslant l$.

Exercise 21.2. *Prove that Lemma 21.4 implies the Stechkin's theorem.*

Remark 21.2. For $k=1$, in the Jackson inequality

$$\widetilde{E}_n(f) \leqslant c\omega_1(\pi/n, f), \quad n \geqslant 1,$$

Korneichuk proved that $c=1$.

22 Bernstein's Inequality

Theorem 22.1 (Bernstein's inequality)**.** *If a trigonometric polynomial T_n is of order $\leqslant n$, then*

$$\left\|T_n'\right\|_{C(\mathbb{R})} \leqslant n \left\|T_n\right\|_{C(\mathbb{R})}.$$

Remark 22.1. The Bernstein's inequality is obvious for $\cos nx$, $\sin nx$.

Proof of Theorem 22.1. Assume to the contrary that for some T_n and $\varepsilon > 0$

$$\left\|T_n'\right\|_{C(\mathbb{R})} > (n+\varepsilon)\left\|T_n\right\|_{C(\mathbb{R})},$$

i.e., for $t_n := \frac{T_n}{\|T_n\|_{C(\mathbb{R})}}(1-\varepsilon)$ we have $\|t_n\| = 1-\varepsilon$, and $\|t_n'\| > (n+\varepsilon)(1-\varepsilon) > n$. Let α_0 be such a point that $|t_n'(\alpha_0)| = \|t_n'\|_{C(\mathbb{R})}$, we will assume $t_n'(\alpha_0) > 0$. Now instead of t_n we consider a translated polynomial $\tau_n(\alpha) = t_n(\alpha - \beta_0)$, this doesn't change the norm of the polynomial and the norm of its derivative. We choose β_0 in a way that the point α_0^*, where the derivative τ_n' attains its largest value, is such that $\tau_n(\alpha_0^*) = \cos(n\alpha_0^*)$ (and also $(\cos(n\alpha_0^*))' > 0$). So, $\tau_n'(\alpha_0^*) > n$, $\cos(n\alpha_0^*)' \leqslant n$. Suppose $\alpha > \alpha_0$. Then, by Lagrange's theorem, $(\tau_n(\alpha) - \cos(n\alpha)) - (\tau_n(\alpha_0^*) - \cos(n\alpha_0^*)) = (\alpha - \alpha_0)(\tau_n'(\theta) - \cos(n\theta)) > 0$, $\alpha_0^* < \theta < \alpha$. This means that to the right of α_0^* the function τ_n is above $\cos(n\cdot)$, and to the left — below. Since $\|\tau_n\|_{C(\mathbb{R})} < 1$, the graph of τ_n will intersect every "wave" (arc) of the function $\cos(n\cdot)$, moreover one of the waves will be intersected three times. But $\cos(n\cdot)$ has only $2n$ waves, which means $\tau_n(\cdot) - \cos(n\cdot)$ will have $2n+2$ zeroes on $[-\pi, \pi]$, so a trigonometric polynomial of order n will have $2n+2$ zeroes, which is possible only if $\tau_n(x) - \cos(nx) \equiv 0$, which is not possible, beause $\|\tau_n\|_{C(\mathbb{R})} \neq \|\cos(n\cdot)\|_{C(\mathbb{R})} = 1$. □

Remark 22.2. An analog of this inequality exists in L_p, $1 \leqslant p < \infty$. If a trigonometric polynomial T_n is of order $\leqslant n$, then

$$\left\|T_n'\right\|_{L_p[0,2\pi]} \leqslant n \left\|T_n\right\|_{L_p[0,2\pi]}.$$

23 Converse Bernstein-de La Vallée Poussin Theorem

Remark 23.1. Let us recall the Jackson-Stechkin inequality:

$$\widetilde{E}_n(f) \leqslant c\omega_k(1/n, f), \quad n \geqslant 1,$$

where f is any 2π-periodic continuous function. Converse inequality is not true in general, but it is true for some important orders of the modulus of smoothness.

Theorem 23.1 (converse Bernstein-de La Vallée Poussin theorem)**.** *Suppose $0 < \alpha < k$, and f is a 2π-periodic continuous function, satisfying*

$$\widetilde{E}_n(f) \leqslant \frac{c}{n^\alpha}, \quad n \geqslant 1.$$

Then

$$\omega_k(t, f) \leqslant c(k, \alpha)t^\alpha, \quad t \geqslant 0.$$

Exercise 23.1. *Prove this theorem.*

Proof of Theorem 23.1 for $k = 1$. Expand f into a telescopic Bernstein's series, *i.e.*,

$$f = T_0 + (T_1 - T_0) + \sum_{j=1}^{\infty}(T_{2^j} - T_{2^{j-1}}),$$

where T_k is the polynomial of best approximation of f of order k. For the partial sum $S_l = T_0 + (T_1 - T_0) + \sum_{j=1}^{l}(T_{2^j} - T_{2^{j-1}}) = T_{2^l}$, implying

$$\|f - S_l\|_{C(\mathbb{R})} = \|f - T_{2^l}\|_{C(\mathbb{R})} \leqslant \frac{1}{2^{\alpha l}} \to 0, \quad l \to \infty.$$

This justifies that the telescopic Bernstein's series is indeed uniformly convergent to f.

Suppose $h > 0$, $t \in \mathbb{R}$. We need to prove that $|f(x+h) - f(x)| \leqslant c(k, \alpha)h^\alpha$. Choose j_0 in a way that $\frac{1}{2^{j_0+1}} \leqslant h < \frac{1}{2^{j_0}}$ and split the sum into two: $\sum_{j=1}^{\infty} = \sum_{j=1}^{j_0} + \sum_{j=j_0+1}^{\infty}$. We estimate the second sum:

$$\begin{aligned}\|T_{2^j} - T_{2^{j-1}}\| &\leqslant \|T_{2^j} - f\| + \|T_{2^{j-1}} - f\| = \widetilde{E}_{2^j}(f) + \widetilde{E}_{2^{j-1}}(f) \\ &\leqslant 2\widetilde{E}_{2^{j-1}}(f) \leqslant \frac{2}{2^{\alpha(j-1)}},\end{aligned}$$

hence,

$$\begin{aligned}\left\| \sum_{j=j_0+1}^{\infty} T_{2^j} - T_{2^{j-1}} \right\| &\leqslant \sum_{j=j_0+1}^{\infty} \|T_{2^j} - T_{2^{j-1}}\| \leqslant 2 \sum_{j=j_0+1}^{\infty} \frac{1}{2^{\alpha j - 1}} \\ &= \frac{2}{2^{\alpha j_0}} \cdot \frac{1}{1 - 2^{-\alpha}} = \frac{c(\alpha)}{2^{(j_0+1)\alpha}} \\ &\leqslant c(\alpha)h^\alpha.\end{aligned}$$

Applying the Lagrange's theorem and the Bernstein's inequality, we get

$$\begin{aligned}|f(x+h)-f(x)| &\leqslant |(T_1(x+h)-T_0(x+h))-(T_1(x)-T_0(x))| \\ &\quad + \sum_{j=1}^{j_0} |(T_{2^j}(x+h)-T_{2^{j-1}}(x+h))-(T_{2^j}(x)-T_{2^{j-1}}(x))| + 2ch^\alpha \\ &\leqslant h\sum_{j=0}^{j_0} \left\|T'_{2^j}-T'_{2^{j-1}}\right\| + 2ch^\alpha \leqslant h\sum_{j=0}^{j_0} 2^j \left\|T_{2^j}-T_{2^{j-1}}\right\| + 2ch^\alpha \\ &\leqslant h\sum_{j=0}^{j_0} 2^j \frac{2}{2^{\alpha(j-1)}} + 2ch^\alpha = ch\sum_{j=0}^{j_0} 2^{(1-\alpha)j} + 2ch^\alpha \\ &= ch\frac{2^{(1-\alpha)(j_0+1)}-1}{2^{1-\alpha}-1} + 2ch^\alpha \leqslant ch2^{(1-\alpha)(j_0+1)} + 2ch^\alpha \\ &= ch\frac{1}{h^{1-\alpha}} + 2ch^\alpha = ch^\alpha.\end{aligned}$$

□

24 Bernstein's Inequality for Algebraic Polynomials. Markov's Inequality

Theorem 24.1 (Bernstein's inequality for algebraic polynomials). *For every algebraic polynomial P_n of degree $\leqslant n$*

$$|P_n'(x)| \leqslant \frac{n}{\sqrt{1-x^2}} \|P_n\|_{C[-1,1]}, \quad x \in (-1,1).$$

Exercise 24.1. *Prove that this inequality follows from the Bernstein's inequality with the aid of substitution $x = \cos t$.*

Theorem 24.2 (Markov's inequality). *For every algebraic polynomial P_n of degree $\leqslant n$*

$$\left\|P_n'\right\|_{C[-1,1]} \leqslant n^2 \|P_n\|_{C[-1,1]}.$$

Proof. If $\dfrac{1}{\sqrt{1-x^2}} \leqslant n$, then this inequality follows from the Bernstein's inequality for algebraic polynomials. So, we assume that $\dfrac{1}{\sqrt{1-x^2}} > n$.

Consider the graph of the Chebyshev polynomial of degree n on $[-1,1]$. It is easy to check that x is on the right of the biggest extremum of the Chebyshev polynomial, or it is on the left of the smallest extremum of the Chebyshev polynomial. Assume that it is to the right, and let P_n be a polynomial of degree $\leqslant n$ such that $\|P_n\|_{C[-1,1]} < 1$ and $P_n'(x) > n^2$. For our Chebyshev polynomial T_n we have $\|T_n'\|_{C[-1,1]} \leqslant n^2$.

If $P_n'(x) > 0$, we compare the graph of P_n with the graph of T_n, otherwise — with the graph of $-T_n$.

If $P_n(x) \geqslant T_n(x)$, then we translate the graph of T_n to the left so that it passes through the point $(x, P_n(x))$. Let $\widetilde{T}_n$ be the translated graph. The polynomial T_n has n waves on $[-1,1]$, therefore P_n will intersect $n-1$ waves of $\widetilde{T}_n$ on $[-1,1]$. We also see that $|P_n'(x)| > n^2$, which means that P_n will have to intersect the rightmost wave of $\widetilde{T}_n$ at least 3 times, so P_n and $\widetilde{T}_n$ have $n+1$ points of intersection, which is impossible.

If $P_n(x) < T_n(x)$, then, translating T_n to the right to pass through $(x, P_n(x))$, and repeating the same arguments, we get a contradiction. $\square$

25 Dzyadyk's Inequality

Bernstein's inequality for trigonometric polynomials enabled us to prove converse theorems for approximation by trigonometric polynomials. For approximation by alegbraic polynomials on the segment, it may be natural to expect that one can successfully prove the converse theorems using the Markov-Bernstein inequality:

$$|P_n'(x)| \leqslant \frac{c}{\rho_n(x)} \|P_n\|_{C[-1,1]}, \quad x \in [-1,1],$$

where $\rho_n(x) = \frac{1}{n^2} + \frac{1}{n}\sqrt{1-x^2}$. However, this inequality is not sufficient and we need the more general Dzyadyk's inequality.

Let $s \in \mathbb{R}$, P_n be a polynomial of degree $\leqslant n$. Dzyadyk's inequality states that if

$$|P_n(x)| \leqslant \rho_n^s(x), \quad x \in [-1,1],$$

then

$$|P_n'(x)| \leqslant c(s)\rho_n^{s-1}(x), \quad x \in [-1,1].$$

Remark 25.1. The Bernstein-Markov inequality is a partial case of the Dzyuadyk's inequality for $s = 0$.

We rewrite the Dzyuadyk's inequality in the form:

$$\left\| \frac{P_n'}{\rho_n^{s-1}} \right\|_{C[-1,1]} \leqslant c(s) \left\| \frac{P_n}{\rho_n^s} \right\|_{C[-1,1]}.$$

There are several proofs of this inequality, all are quite complicated, in particular using the theory of analytic functions, conformal mappings, which allows to extend this result for general compact sets in complex plane.

If $[-1,1]$ is replaced with ∂M — the boundary of a compact set $M \subset \mathbb{C}$ with connected complement, then the Dzyuadyk's inequality holds with $c(s)$ replaced by $c(s, \partial M)$, and $\rho_n(z)$ — the distance from $z \in \partial M$ to the level curveof the set M. Let us specify what is level curve. By Riemman's theorem, there exists a conformal mapping Φ, which maps the extrior of M on the exterior of the unit disc. Let Ψ be the inverse mapping. Consider the image of the circle with the radius $1 + \frac{1}{n}$ under the mapping Ψ, which we will denote $\Gamma_{1+\frac{1}{n}}$. This image is a curve that is called a level curve of the set M.

Coming back to the segment: the Zhukovskii function $\frac{1}{2}(w + 1/w)$ maps circle into an ellipse. The distance from a point x on the segment to the ellipse will be $\rho_n(x)$ up to a constant factor.

We will prove the Dzyuadyk's inequality as a corollary of the Markov-Bernstein inequality.

Lemma 25.1. *Suppose $y \in [-1,1]$, $m \in \mathbb{N}$, P_n is a polynomial of degree $\leqslant n$. If*

$$|P_n(x)| \leqslant (|x-y| + \rho_n(y))^m, \quad x \in [-1,1],$$

then

$$|P_n'(y)| \leqslant c(m)\rho_n^{m-1}(y).$$

Proof. As usual, $x_j := \cos(j\pi/n)$, $0 \le j \le n$, $I_j := [x_j, x_{j-1}]$, $|I_j| := x_j - x_{j-1}$, $\tilde{x}_j := \cos((j-1/2)\pi/n)$ — zero of the Chebyshev polynomial that belongs to I_j. Consider the polynomial of degree $(n-1)$:

$$t_j(x) := \begin{cases} \frac{T_n(x)}{x-\tilde{x}_j}|I_j|, & x \ne \tilde{x}_j; \\ T_n'(\tilde{x}_j)|I_j|, & x = \tilde{x}_j. \end{cases}$$

It is easy to see that

$$\frac{4}{3} < t_j(x) < 4, \quad x \in I_j. \tag{25.1}$$

Consider a new polynomial:

$$q_n(x) := \frac{y - \tilde{x}_\nu}{T_n(y)} \cdot \frac{T_n(x)}{x - \tilde{x}_\nu},$$

where ν is the index of the segment: $y \in I_\nu$ (y is fixed, $T_n(y) \ne 0$).

From (25.1) we derive

$$|q_n(x)| \leqslant \frac{23\rho_n(y)}{|x-y| + \rho_n(y)}.$$

The lemma will be verified, if we prove

$$|P_n(x)| \leqslant (|x-y| + \rho_n(y))^m \cdot \frac{1}{\rho_n^m(y)}, \quad x \in [-1,1].$$

Consider the polynomial $Q_n := P_n q_n^m$. Differentiate at y:

$$P_n'(y) = P_n'(y) \cdot q_n^m(y) = Q_n'(y) - P_n(y)(q_n^m(y))'.$$

Since $\|Q\|_{C[-1,1]} \leqslant 23^m$, by Bernstein-Markov's inequality,

$$|Q_n'(y)| \leqslant \frac{c}{\rho_n(y)}.$$

Since $|P_n(y)| \leqslant 1$, $\|q_n^m\|_{C[-1,1]} \leqslant 23^m$, then by Bernstein-Markov's inequality,

$$|(q_n^m(y))'| \leqslant \frac{c}{\rho_n(y)}.$$

Finally,

$$\begin{aligned} |P_n'(y)| &\leqslant |Q_n'(y)| + |P_n(y)| \cdot |(q_n^m(y))'| \\ &\leqslant \frac{c(m)}{\rho_n(y)} + 1 \cdot \frac{c(m)}{\rho_n(y)} = \frac{c(m)}{\rho_n(y)}. \end{aligned}$$

□

Exercise 25.1. *Prove that Lemma 25.1 implies the Dzyuadyk's inequality for $s > 0$.*

26 Approximation by Rational Functions. Newman's Theorem

Consider $f(x) = |x|$, $[-1,1]$. The direct approximation theorem provides $E_n(f) \leqslant c/n$, $n \in \mathbb{N}$, for an absolute constant c. Bernstein proved that $nE_n(f) \to c^*$, $n \to \infty$, $c^* \approx 0.2801694990\ldots$. In particular, $E_n(f) \geqslant c_1/n$, $n \in \mathbb{N}$. So, $|x|$ can be approximated by algebraic polynomials with the error of order $1/n$ and not better. We will prove such estimate.

Statement 26.1. $E_n(|x|)_{C[-1,1]} \geqslant \dfrac{1}{16n}$, $n \in \mathbb{N}$.

Proof. Suppose there exists a polynomial p_n of degree $\leqslant n$ such that

$$||x| - |p_n(x)|| < \frac{1}{16n}, \quad x \in [-1,1].$$

Then $p_n^*(x) := \frac{1}{2}(p_n(x) + p_n(-x))$, $x \in [-1,1]$, is an even polynomial satisfying

$$||x| - p_n^*(x)| = \frac{1}{2}|(|x| - p_n(x)) + (|-x| - p_n(-x))| < \frac{1}{16n}, \quad x \in [-1,1].$$

Put $\hat{p}_n(x) := p_n^*(x) - p_n^*(0)$, $x \in [-1,1]$. Then

$$||x| - \hat{p}_n(x)| < \frac{1}{8n}, \quad x \in [-1,1]. \tag{26.1}$$

Since $\hat{p}_n(0) = 0$ and $\hat{p}_n$is even,

$$\tilde{p}_{n-1}(x) := \begin{cases} \dfrac{8}{9} \cdot \dfrac{1}{x}\hat{p}_n(x), & x \neq 0, \\ 0, & x = 0, \end{cases}$$

is an odd algebraic polynomial of degree $\leqslant n-1$.

For all $x \in [-1,1]$, $x \neq 0$, the estimate (26.1) gives

$$|\tilde{p}_{n-1}(x)| = \frac{8}{9} \cdot \frac{1}{|x|}|(\hat{p}_n(x) - |x|) + |x|| < \frac{1}{9nx} + \frac{8}{9},$$

in particular,

$$|\tilde{p}_{n-1}(x)| < \frac{1}{9n\sin(\frac{\pi}{2n})} + \frac{8}{9} < 1,$$

with $\sin(\frac{\pi}{2n}) \leqslant x \leqslant 1$, using the estimates

$$\frac{\pi}{2}x \leqslant \sin\left(\frac{\pi}{2}x\right) \leqslant x, \quad x \in [0, \pi/2].$$

Now we compare with the Chebyshev polynomial T_m for odd m. Let $m = n$, if n is odd, and let $m = n+1$, if n is even. For example, consider the case when $m = n$, *i.e.*, n is odd. Denote $x_j = \cos(j\pi/n)$ — the alternation points of T_n, and note that

$$x_- := x_{(n+1)/2} = -\sin\left(\frac{\pi}{2n}\right), \quad x_+ := x_{(n-1)/2} = -\sin\left(\frac{\pi}{2n}\right).$$

Remark that

$$|\tilde{p}_{n-1}(x)| < 1, \quad x \in [-1,1] \setminus [x_-, x_+].$$

Hence, because the polynomials are odd,

$$|\tilde{p}_{n-1}(x)| \leqslant |T_n(x)|, \quad x \in [x_-, x_+].$$

Take $x_0 := \sin\left(\dfrac{\pi}{6n}\right)$, $|T_n(x_0)| = \dfrac{1}{2}$. We have

$$|\hat{p}_n(x_0)| = \frac{9}{8}x_0|\tilde{p}_n(x_0)| \leqslant \frac{9}{8}x_0|T_n(x_0)| = \frac{9}{16}x_0,$$

hance

$$||x_0| - \hat{p}_n(x_0)| \geqslant |x_0| - |\hat{p}_n(x_0)| \geqslant \frac{7}{16}x_0 = \frac{7}{16}\sin\left(\frac{\pi}{6n}\right)$$
$$\geqslant \frac{7}{16}\cdot\frac{1}{3n} > \frac{6}{16\cdot 3n} = \frac{1}{8n},$$

which contradicts (26.1). □

The theory of approximation by rational functions originated with the surprising result by Newman, who proved that the function $|x|$ can be approximated by rational functions much better than by algebraic polynomials.

Definition 26.2. For $f \in C[-1,1]$ denote

$$\rho_n(f) := \inf_{r_n} \|f - r_n\|_{C[-1,1]},$$

where the infimum is taken over all $r_n = \dfrac{p_n}{q_n}$, where p_n, q_n are polynomials of degree $\leqslant n$, and q_n does not have zeroes on $[-1,1]$.

Theorem 26.3 (Newman, 1964)**.** *The following inequalities hold:*

$$e^{-\pi\sqrt{n+1}} \leqslant \rho_n(|x|) \leqslant 3e^{-\sqrt{n}}, \quad n \geqslant 5.$$

We will prove only the estimate from above, following the original proof of Newman.

Proof. For every positive integer n define $a := e^{-\frac{1}{\sqrt{n}}}$, and

$$N(x) := N_n(x) := \prod_{k=1}^{n-1}(x + a^k), \quad x \in \mathbb{R},$$

the so called Newman's polynomial.

Lemma 26.4. *For $n \geqslant 5$*

$$\left|\frac{N(-x)}{N(x)}\right| \leqslant e^{-\sqrt{n}}, \quad x \in [e^{-\sqrt{n}}, 1].$$

Proof. First we prove that

$$\prod_{j=1}^{n-1}\frac{1-a^j}{1+a^j} \leqslant e^{-\sqrt{n}}, \quad n \geqslant 5. \tag{26.2}$$

Consider the function

$$\varphi(t) := (1+t)e^{-2t} - (1-t), \quad t \in [0,\infty).$$

It is easy to check that $\varphi(0) = 0$ and $\varphi'(t) > 0$, $t > 0$. This implies

$$\frac{1-t}{1+t} \leqslant e^{-2t}, \quad t \geqslant 0.$$

But

$$2(a - a^n) \geqslant 2(e^{-\frac{1}{\sqrt{5}}} - e^{-\sqrt{5}}) > 1, \quad n \geqslant 5.$$

Also, for $t \geqslant 0$, $1 - e^{-t} \leqslant t$, hence $(1-a)^{-1} \geqslant \sqrt{n}$. So,

$$\prod_{j=1}^{n-1} \frac{1-a^j}{1+a^j} \leqslant \exp\left\{-2\sum_{j=1}^{n-1} a^j\right\} = \exp\left\{-2\frac{a-a^n}{1-a}\right\} \leqslant \exp\{-\sqrt{n}\},$$

which leads to (26.2).

Now for some j, $0 \leqslant j \leqslant n-1$, $a^{j+1} \leqslant x \leqslant a^j$. Then

$$\begin{aligned}\left|\frac{N(-x)}{N(x)}\right| &= \prod_{k=1}^{j} \frac{a^k - x}{a^k + x} \prod_{k=j+1}^{n} \frac{x - a^k}{x + a^k} \\ &\leqslant \prod_{k=1}^{j} \frac{a^k - a^n}{a^k + a^n} \prod_{k=j+1}^{n} \frac{a^j - a^k}{a^j + a^k} \\ &\leqslant \prod_{m=n-j}^{n-1} \frac{1-a^m}{1+a^m} \prod_{m=1}^{n-j-1} \frac{1-a^m}{1+a^m} \\ &= \prod_{m=1}^{n-1} \frac{1-a^m}{1+a^m} \leqslant e^{-\sqrt{n}}.\end{aligned}$$

□

The required rational function will be taken in the form

$$r_n(x) := x\,\frac{N(x) - N(-x)}{N(x) + N(-x)}, \quad x \in [-1, 1].$$

Remark that both $|x|$ and r_n are even functions, so we can check the error of approximation only for non-negative x.

If $x \in [0, a^n] = [0, e^{-\sqrt{n}}]$, then $N(x) \geqslant N(-x) \geqslant 0$, and so $0 \leqslant x - r_n(x) \leqslant x \leqslant e^{-\sqrt{n}}$.

If $x \in (e^{-\sqrt{n}}, 1]$, then by Lemma 26.4

$$\begin{aligned}|x - r_n(x)| &\leqslant 2x\left|\frac{N(-x)}{N(x)+N(-x)}\right| \leqslant 2\left[\left|\frac{N(x)}{N(-x)}\right| - 1\right]^{-1} \\ &\leqslant \frac{2}{e^{\sqrt{n}} - 1} \leqslant 3e^{-\sqrt{n}}.\end{aligned}$$

□

Vyacheslavov (1975) proved that

$$c_1 e^{-\pi\sqrt{n}} \leqslant \rho_n(|x|) \leqslant c_2 e^{-\pi\sqrt{n}}, \quad n \geqslant 1,$$

i.e., we have

$$\lim_{n\to\infty} (\rho_n(|x|))^{1/\sqrt{n}} = e^{-\pi}.$$

Then, using computer computations, performed up to $n = 80$, Varga, Ruttan and Carpenter (1991) conjectured

$$\lim_{n\to\infty} e^{\pi\sqrt{n}} \rho_n(|x|) = 8.$$

This was proved by Stahl in 1992 using logarithmic potentials and estimates of some elliptic integrals.

27 Splines

Splines are piecewise polynomial functions. They are easier-to-calculate than polynomials. In practice, the most used splines are cubic splines.

Given $n+1$ points $a = x_0 < x_1 < \cdots < x_n = b$ on $[a, b]$ or an infinite number of points x_j, $j \in \mathbb{Z}$ on $\mathbb{R}$, $x_j < x_{j+1}$ and $\lim\limits_{j \to \pm\infty} x_j = \pm\infty$.

Definition 27.1. Spline of order r and defect m ($1 \leqslant m \leqslant r$) is a function $s : [a, b] \to \mathbb{R}$ (or $s : \mathbb{R} \to \mathbb{R}$) such that $s \in C^{r-m}[a, b]$ ($C^{r-m}(\mathbb{R})$) and on every interval (x_j, x_{j+1}) this function is an algebraic polynomial of degree $\leqslant r$. The points x_j are called the knots of the spline.

Remark 27.1. If one takes $m = 0$ in this definition, then the smoothness condition implies that s is simply a single polynomial of degree $\leqslant r$. Therefore, the splines of defect < 1 are not considered — they are polynomials. Splines of defect $m = r + 1$ are considered — these are piecewise polynomial functions which are not necessarily continuous.

Definition 27.2. With $m = 1$ the corresponding spline is called the spline of minimal defect, and with $m = r + 1$ — splines of maximal defect.

Remark 27.2. Very often in the literature a "spline" is a spline of minimal defect. In particular, a "cubic spline" is a cubic spline of minimal defect (twice continuously differentiable function).

Remark 27.3. An important set of splines is interpolatory splines, that is, those satisfying $s(x_j) = f(x_j)$, for all j, where f is a given function.

Example 27.1. Interpolatory piecewise-constant and interpolatory piecewise-linear functions.

Suppose $a = x_0 < x_1 < \cdots < x_n$, and $f : [a, b] \to \mathbb{R}$ is a given function. To construct an interpolatory cubic spline of minimal defect, consider a cubic polynomial $a_j x^3 + b_j x^2 + c_j x + d_j$ for every $[x_j, x_{j+1}]$, $0 \le j \le n-1$. In total, we have $4n$ unknown coefficients. Calculate the number of conditions that we have on these coefficients: $s \in C^2$ gives 3 conditions for every interior knot ($s(x_j-) = s(x_j+)$, $s'(x_j-) = s'(x_j+)$, $s''(x_j-) = s''(x_j+)$, $1 \le j \le n-1$), and interpolation gives one condition at every knot ($s(x_j) = f(x_j)$, $0 \le j \le n$). We have $3(n-1) + (n+1) = 4n - 2$ conditions which are linear equations on the necessary coefficients. The missing two conditions are emposed near the endpoints. For example, if an information about the behaviour of f near the endpoints is known, $s'(a) = k_a$ and $s'(b) = k_b$, where $k_a, k_b \in \mathbb{R}$.

For quadratic spline we will have $2(n-1)+(n+1) = 3n-1$ conditions on $3n$ unknown coefficients. So we need to add only one condition, but this will cause the non-symmetric choice of the condition, because we have two endpoints of the interval. Therefore, for interpolation at the knots, the splines of odd degree are preferable. For splines of even degree, one can interpolate not at the knots, but at some points between the knots, which will make the number of additional conditions even.

28 Euler's Ideal Spline. Kolomogorov's Inequality on Derivatives

Definition 28.1. Euler's ideal spline of degree r, $r \ge 0$, is a 2π-periodic spline ε_r on $\mathbb{R}$ of minimal defect with knots $k\pi$, $k \in \mathbb{Z}$, such that

1) $\varepsilon_r^{(r)} = \operatorname{sign} \sin x, \quad x \ne k\pi, k \in \mathbb{Z},$

2) $\displaystyle\int_{-\pi}^{\pi} \varepsilon_r(x) dx = 0.$

Remark 28.1. In fact, the above relations define the unique ε_r. ε_0 and ε_1 are piecewise-constant and piecewise-linear functions, respectively. For $r \geqslant 2$ the spline $\varepsilon_r(x)$ looks like $\sin(x + \frac{r\pi}{2})$.

Remark 28.2. Below in this section $\|\cdot\| := \|\cdot\|_{L_\infty(\mathbb{R})}$. Clearly, for $f \in C(\mathbb{R})$ we have $\|f\|_{L_\infty(\mathbb{R})} = \|f\|_{C(\mathbb{R})}$.

Consider properties of the Euler's splines.

Property 1. *Expanding into the Fourier series, we obtain:*

$$\varepsilon_0(x) = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{\sin(2k+1)x}{2k+1}, \quad \varepsilon_1(x) = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{\cos(2k+1)x}{(2k+1)^2}.$$

Next, since $\varepsilon_1 \in \mathrm{Lip}1$, the series for ε_1 is uniformly convergent, and we can integrate term by term. We obtain

$$\varepsilon_r(x) = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{\sin((2k+1)x + \frac{k\pi}{2})}{(2k+1)^{r+1}}.$$

Property 2. *Numbers $K_r := \|\varepsilon_r\|$ are called the Favar's constants. Remark that $K_0 = 1$, $K_1 = \frac{\pi}{2}$, and for even r : $\|\varepsilon_r\| = |\varepsilon_r(0)| = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{1}{(2k+1)^{r+1}}$, and for odd r : $\|\varepsilon_r\| = |\varepsilon_r(\frac{\pi}{2})| = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+1)^{r+1}}$.*

Property 3. *It is easy to verify the inequalities:*

$$1 = K_0 < K_2 < \cdots < \frac{4}{\pi} < \cdots < K_3 < K_1 = \frac{\pi}{2}.$$

Property 4. *The Euler's ideal spline for odd r alternatively attains the largest and the smallest value at the points $k\pi$, equal to plus or minus the norm, and for even r — at the points $k\pi + \frac{\pi}{2}$.*

Below, the part of the graph of ε_r between two consequtive extrema is called a wave.

Recall that $W^r(\mathbb{R}) = W^r_\infty(\mathbb{R})$ is the space of functions having $(r-1)$-st derivative, which is absolutely continuous and $\left\|f^{(r)}\right\| < +\infty$.

Theorem 28.2 (Kolmogorov's inequality)**.** *If $f \in W^r(\mathbb{R})$, then for arbitrary $0 \le j \le r$ we have*

$$\left\|f^{(j)}\right\| \leqslant \frac{K_{r-j}}{K_r^{1-\frac{j}{r}}}\|f\|^{1-\frac{j}{r}}\left\|f^{(r)}\right\|^{\frac{j}{r}}. \tag{28.1}$$

Remark 28.3. This inequality is sharp, since the equality is achieved for the Euler's ideal spline. Indeed, for $f = \varepsilon_r$, $f^{(j)} = \varepsilon_{r-j}$,

$$\|\varepsilon_{r-j}\| = \frac{K_{r-j}}{K_r^{1-\frac{j}{r}}}\|\varepsilon_r\|^{1-\frac{j}{r}}.$$

For other functions, we will prove this inequality using comparison with the Euler's ideal spline.

Lemma 28.3. *If $f \in W^r(\mathbb{R})$, $\left\|f^{(r)}\right\| < 1$, $\|f\| < K_r$, then $\|f'\| \leqslant K_{r-1}$.*

Proof. Consider three cases

Main case. Assume that for some $n \in \mathbb{N}$ the function f is periodicwith period $2\pi n$. Will argue as in the proof of the Bernstein's inequality. Assume to the contrary thatfor some x_0: $|f'(x_0)| > K_{r-1}$. Without loss of generality, $f'(x_0) > K_{r-1}$. Consider the translation of our function $g(\cdot) := f(\cdot - a)$ such that $g(t_0) = \varepsilon_r(t_0)$, $\varepsilon_r'(t_0) > 0$, where $t_0 = x_0 + a$. Consider the wave of ε_r, containing t_0. Since $g'(t_0) > K_{r-1} \geqslant \varepsilon_r'(t_0)$, and $\|g\| < K_r = \|\varepsilon_r\|$, the graph of g interesects this wave at at least three points. Let A be the x-coordinate of the left endpoint of the wave. Taking into account that $\|g\| < K_r = \|\varepsilon_r\|$, we obtain that the graph of g will intersect every wave of ε_r on a period $(A, A + 2\pi n]$, and one of the waves at least three times. In total, on the period $(A, A + 2\pi n]$, the difference $\varepsilon_r - g$ has $\geqslant 2n + 2$ zeroes. Periodicity and Rolle's theorem imply that the derivative of the difference also has $\geqslant 2n+2$ zeroes on the period. Therefore, $\varepsilon_r^{(r-1)} - g^{(r-1)}$ has $\geqslant 2n+2$ zereos on the period. This means that there is an interval of the length π, where there are $\geqslant 2$ zereos of this difference, and $\varepsilon_r^{(r-1)}$ is linear. Let $t_1 < t_2$ be such zeroes. Then

$$\begin{aligned} |t_1 - t_2| &= |\varepsilon_r^{(r-1)}(t_1) - \varepsilon_r^{(r-1)}(t_2)| = |g^{(r-1)}(t_1) - g^{(r-1)}(t_2)| \\ &= \left| \int_{t_1}^{t_2} g^{(r)}(x)\, dx \right| \leqslant (t_2 - t_1) \left\| g^{(r)} \right\| < t_2 - t_1, \end{aligned}$$

since $\left\|g^{(r)}\right\| = \left\|f^{(r)}\right\| < 1$, contradiction.

f is of bounded support, i.e., for some $d > 0$: $f(x) = 0$, $x \notin [-d, d]$. Then, there exists $n \in \mathbb{N}$: $2\pi n > 4d$. For $x \in [-2d, 2d + 2\pi n]$, let $g(x) := f(x)$, and extend g so that it is $2\pi n$-periodic, reducing this case to the main one.

f is arbitrary. This case can be reduced to the previous two cases using the following lemma about "smooth hat". ☐

Lemma 28.4. *For any $a > 0$, $\varepsilon > 0$, $r \in \mathbb{N}$ there exist $d > 0$ and a function s such that*

1) $s(x) = 1, \quad x \in [-a, a]$;

2) $s(x) = 0, \quad x \notin [-d, d]$;

3) $\|s\| \leqslant 1$;

4) $\left\|s^{(j)}\right\| \leqslant \varepsilon, \quad 1 \leq j \leq r.$

Exercise 28.1. *Prove Lemma 28.4, and complete the proof of the last case of Lemma 28.3.*

Proof of the theorem. Lemma 28.3 provides that if $\|f\| = \|\varepsilon_r\| = K_r$ and $\left\|f^{(r)}\right\| = \left\|\varepsilon_r^{(r)}\right\| = K_0 = 1$, then $\|f'\| \leqslant \|\varepsilon_r'\|$. Indeed, for $g_\varepsilon = \dfrac{f}{1+\varepsilon}$, $\varepsilon > 0$, we have $\|g_\varepsilon\| = \dfrac{1}{1+\varepsilon}\|f\| < \|\varepsilon_r\|$, and $\left\|g_\varepsilon^{(r)}\right\| = \dfrac{1}{1+\varepsilon}\left\|f^{(r)}\right\| < \left\|\varepsilon_r^{(r)}\right\|$, hence $\|f'\| = (1+\varepsilon)\|g_\varepsilon'\| \leqslant (1+\varepsilon)\|\varepsilon_r'\|$, and since $\varepsilon > 0$ is arbitrary, $\|f'\| \leqslant \|\varepsilon_r'\|$. Taking into account that $\|\varepsilon_r'\| = \dfrac{K_{r-1}}{K_r^{1-\frac{1}{r}}}\|\varepsilon_r\|^{1-\frac{1}{r}}\|\varepsilon_0\|^{\frac{1}{r}}$, the Kolmogorov's inequality is proved for $j = 1$ and $\|f\| = K_r$ and $\left\|f^{(r)}\right\| = K_0$. Now we reduce the general case of $j = 1$ to ours. For arbitrary f choose a and b so that the function $g(x) = af(bx)$, $x \in \mathbb{R}$, satisfies $\|g\| = a\|f\| = K_r$ and $\left\|g^{(r)}\right\| = ab^r\left\|f^{(r)}\right\| = K_0$. For $f \not\equiv 0$ and $f^{(r)} \not\equiv 0$ such a, b exist, otherwise the inequality is trivial. By the proved above, for g we obtain

$$\begin{aligned}\|f'\| = \frac{\|g'\|}{ab} &\leqslant \frac{1}{ab}\frac{K_{r-1}}{K_r^{1-\frac{1}{r}}}\|g\|^{1-\frac{1}{r}}\left\|g^{(r)}\right\|^{\frac{1}{r}}\\ &= \frac{1}{ab}\frac{K_{r-1}}{K_r^{1-\frac{1}{r}}}a^{1-\frac{1}{r}}\|f\|^{1-\frac{1}{r}}a^{\frac{1}{r}}b\left\|f^{(r)}\right\|^{\frac{1}{r}}\\ &= \frac{K_{r-1}}{K_r^{1-\frac{1}{r}}}\|f\|^{1-\frac{1}{r}}\left\|f^{(r)}\right\|^{\frac{1}{r}}.\end{aligned}$$

So, for $j=1$ the inequality is proved. For $1<j\leqslant r$, we prove by induction. Suppose that for $1<j<r$ and for $r-1$ the inequality is proved, and we will show that then for $j+1$:

$$\begin{aligned}\left\|f^{(j+1)}\right\| &\leqslant \frac{K_{r-j-1}}{K_{r-j}^{1-\frac{1}{r-j}}}\left\|f^{(j)}\right\|^{1-\frac{1}{r-j}}\left\|f^{(r)}\right\|^{\frac{1}{r-j}}\\ &\leqslant \frac{K_{r-j-1}}{K_{r-j}^{1-\frac{1}{r-j}}}\left(\frac{K_{r-j}}{K_r^{1-\frac{j}{r}}}\|f\|^{1-\frac{j}{r}}\left\|f^{(r)}\right\|^{\frac{j}{r}}\right)^{1-\frac{1}{r-j}}\left\|f^{(r)}\right\|^{\frac{1}{r-j}}\\ &= \frac{K_{r-j-1}}{K_r^{1-\frac{j+1}{r}}}\|f\|^{1-\frac{j+1}{r}}\left\|f^{(r)}\right\|^{\frac{j+1}{r}}.\end{aligned}$$

□

Remark 28.4. For $r=2$, the Kolmogorov's inequality is the Landau's inequality. Similarly, the more general inequality is investigated:

$$\left\|f^{(j)}\right\|_p \leqslant c(j,r,p,q,s)\|f\|_q^{1-\frac{j}{r}}\left\|f^{(r)}\right\|_s^{\frac{j}{r}}, \tag{28.2}$$

where sharp values of the constant $c(j,r,p,q,s)$ are of interest. This is a multiplicative Kolmogorov's inequalities, and an additive inequality (where the norms are added) is also investigated. This allows to obtain an analog for functions on the segment. Kolmogorov's inequality is true on the half-line as well.

Next remark has been brought to the authors' attention by Z. Ditzian.

Remark 28.5. If one proves (28.1), *i.e.,* the Kolmogorov's inequality for $p=q=s=\infty$, it is easy to obtain it for arbitrary $p=q=s\in[1,\infty)$, with the same constant. This was first noticed by Stein (in 1957). The idea is useful in other situations (see, for example, Remark 22.2), let us give a sketch of the proof. Let $f\in W_p^r(\mathbb{R})$. For arbitrary $g\in L_{p'}(\mathbb{R})$, $1/p+1/p'=1$, the convolution $F(x):=\int_{-\infty}^{\infty}f(x+t)g(t)\,dt$ belongs to $W_\infty^r(\mathbb{R})$ and has derivatives $F^{(j)}(x)=\int_{-\infty}^{\infty}f^{(j)}(x+t)g(t)\,dt$, $1\le j\le r$. Fix j, $0<j<r$. There exists g such that $\|g\|_{p'}=1$ ş $F^{(j)}(0)=\int_{-\infty}^{\infty}f^{(j)}(t)g(t)\,dt=\left\|f^{(j)}\right\|_p$. By (28.1), we have $|F^{(j)}(0)|\leqslant\left\|F^{(j)}\right\|_\infty\leqslant\frac{K_{r-j}}{K_r^{1-\frac{j}{r}}}\|F\|_\infty^{1-\frac{j}{r}}\left\|F^{(j)}\right\|_\infty^{\frac{j}{r}}$. By Hölder's inequality, $\|F\|_\infty\leqslant\|f\|_p\|g\|_{p'}=\|f\|_p$, and in the same way $\left\|F^{(r)}\right\|_\infty\leqslant\left\|f^{(r)}\right\|_p$, which yields (28.2) for $p=q=s\in[1,\infty)$, with $c=\frac{K_{r-j}}{K_r^{1-\frac{j}{r}}}$. Remark that this constant is sharp for $p=q=s=1$ and $p=q=s=\infty$.

29 Popoviciu's Identity

Suppose a function f is given a the points $y_0 < y_1 < \cdots < y_N$. Then, clearly $f(y_N) - f(y_0) = \sum_{j=0}^{N-1} (f(y_{j+1}) - f(y_j))$, which can be rewritten as

$$[y_0, y_N; f] = \sum_{j=0}^{N-1} (y_{j+1} - y_j)[y_j, y_{j+1}; (y_N - y_0)^{-1}].$$

This equality is the Popoviciu's identity for $m = 1$.

Given two collections of points $y_0 < y_1 < \cdots < y_N$ and $x_0 < x_1 < \cdots < x_m$, where every x_j coincides with some y_i, *i.e.,* the second collection is a subset of the first one.

Theorem 29.1 (Popoviciu's identity)**.** *One has the equality*

$$[x_0, \dots, x_m; f] = \sum_{j=0}^{N-m} (y_{j+m} - y_j)[y_j, \dots, y_{j+m}; f][x_0, \dots, x_m; \prod_{j,m}],$$

where $\prod_{j,m}(\cdot) := \prod_{s=j+1}^{j+m-1} (\cdot - y_s)_+$.

Proof. Induction on m. For $m = 1$ already proved above. Assuming that we have proved this for some $m - 1$, will prove for m. For all $0 \le i \le m$, denote

$$F_m(x_i) := \sum_{j=0}^{N-m} (y_{j+m} - y_j)[y_j, \dots, y_{j+m}; f] \prod_{j,m}(x_i).$$

By assumption,

$$[x_0, \dots, x_{m-1}; F_{m-1}] = [x_0, \dots, x_{m-1}; f]. \tag{29.1}$$

We will prove that for some polynomial P_{m-1} of degree $\leqslant m - 1$ for all $0 \le s \le n$,

$$F_m(x_s) = F_{m-1}(x_s) + P_{m-1}(x_s). \tag{29.2}$$

Indeed,

$$\begin{aligned}
F_m(x_s) &= \sum_{j=0}^{N-m} ([y_{j+1}, \dots, y_{j+m}; f] - [y_j, \dots, y_{j+m-1}; f]) \prod_{j,m}(x_s) \\
&= -[y_0, \dots, y_{m-1}; f] \prod_{0,m}(x_s) + \sum_{j=1}^{N-m} [y_j, \dots, y_{j+m-1}; f](\prod_{j-1,m}(x_s) - \prod_{j,m}(x_s)) \\
&\qquad + [y_{N-m+1}, \dots, y_N; f] \prod_{N-m,m}(x_s) \\
&= P_{m-1}(x_s) + \sum_{j=1}^{N-m} [y_j, \dots, y_{j+m-1}; f](\prod_{j-1,m}(x_s) - \prod_{j,m}(x_s)) + 0.
\end{aligned}$$

Consider the last term:

$$\prod_{j-1,m}(x_s) - \prod_{j,m}(x_s) = (x_s - x_j)_+ \cdots (x_s - x_{j+m-2})_+ - (x_s - x_{j+1})_+ \cdots (x_s - x_{j+m-1})_+$$
$$= ((x_s - x_j)_+ - (x_s - x_{j+m-1})_+) \cdot (x_s - x_{j+1})_+ \cdots (x_s - x_{j+m-1})_+$$
$$= (x_{j+m-1} - x_j) \prod_{j,m-1}(x_s),$$

which verifies (29.2). Now, by (29.1) and (29.2), we obtain

$$[x_0, \ldots, x_m; f] = \frac{[x_0, \ldots, x_{m-1}; f] - [x_1, \ldots, x_m; f]}{x_0 - x_m}$$
$$= \frac{[x_0, \ldots, x_{m-1}; F_{m-1}] - [x_1, \ldots, x_m; F_{m-1}]}{x_0 - x_m}$$
$$= [x_0, \ldots, x_m; F_{m-1}] = [x_0, \ldots, x_m; F_{m-1} - P_{m-1}]$$
$$= [x_0, \ldots, x_m; F_m].$$

□

30 Spline Bases

Theorem 30.1 (spline representation)**.** *Every spline S of minimal defect of degree r with knots $a = t_0 < t_1 < \cdots < t_N = b$ can be written as*

$$S(x) = P_r(x) + \frac{1}{r!}\sum_{j=1}^{N-1}(S^{(r)}(t_j+) - S^{(r)}(t_j-))(x - t_j)_+^r, \tag{30.1}$$

where

$$P_r(x) = S(a) + \frac{S'(0)}{1!}(x-a) + \cdots + \frac{S^{(r)}(a)}{r!}(x-a)^r.$$

Proof. Denote the right hand side of (30.1) by $V(x)$. Remark that $V - S \in C^{r-1}[a,b]$, and for $x \in (t_\nu, t_{\nu+1})$ we have $S^{(r)}(x) = S^{(r)}(t_\nu+)$ and

$$V^{(r)}(x) - S^{(r)}(x) = S^{(r)}(a) + \frac{1}{r!}\sum_{j=1}^{\nu} r!(S^{(r)}(t_j+) - S^{(r)(t_j-)}) - S^{(r)}(t_\nu+) = 0,$$

i.e., $V^{(r)}(x) - S^{(r)}(x) = 0$, for $x \in [a,b] \setminus \{t_0, \ldots, t_N\}$. So, $V^{(r)} \equiv S^{(r)}$, hence $V - S =: Q_{r-1}$ is a polynomial of degree $\leqslant r-1$, which is zero at the point a along with all the derivative, meaning $Q_{r-1} = 0$, $V = S$, as required. □

Remark 30.1. For an infinite sequence of knots $\{t_j\}_{j=-\infty}^{\infty}$, $t_j \to \pm\infty$, $j \to \pm\infty$ an analog of representation (30.1) is valid, with $P_r = 0$ and $\sum_{j=-\infty}^{\infty}$ in place of $\sum_{j=1}^{N-1}$. Proof is similar.

Definition 30.2. Suppose $T = \{t_j\}_{j=0}^{N}$. By $S_r(T)$ denote the set of all splines of degree r of minimal defect with the knots t_j.

Theorem 30.3 (on basis of spliines of minimal defect)**.** *The set $S_r(T)$ is a linear space of dimension $N + r$, with basis:*

$$(x - t_1)_+^r, \ldots, (x - t_{N-1})_+^r, 1, (x-a), \ldots, (x-a)^r.$$

Proof. In view of (30.1), it is sufficient to prove that this system of functions is linearly independent. Assume there are $c_1, \dots, c_{N+r}$ such that

$$\varphi(x) := \sum_{j=1}^{N-1} c_j (x-t_j)_+^r + \sum_{j=0}^{r} c_{N+j}(x-a)^j = 0, \quad x \in [a,b].$$

Then $\varphi(a) = \varphi'(a) = \cdots = \varphi^{(r)}(a) = 0$ provides $c_r = \cdots = c_{N+r} = 0$. Then, step by step $\varphi(t_2) = 0 \Rightarrow c_1 = 0$, $\varphi(t_3) = 0 \Rightarrow c_2 = 0$, and so on. □

Remark 30.2. Denote $T = \{t_j\}_{j=0}^N$, $\bar{k} = (k_1, \dots, k_{N-1})$, where k_j are integers $1 \leqslant k_j \leqslant r$. By $S_r(T, \bar{k})$ denote the set of splines of degree r with the knots t_j such that for every $1 \le j \le N-1$ there is a neighborhood of t_j where $S \in C^{r-k_j}$ (for splines of minimal defect all k_j are equal to 1). Then, for every $S \in S_r(T, \bar{k})$ we have the representation

$$S(x) = P_r(x) + \sum_{j=1}^{N-1} \sum_{i=1}^{k_i} \frac{1}{(r-i)!} (S^{(r-i)}(t_j+) - S^{(r-i)}(t_j-))(x-t_j)_+^{r-i}$$

with the same P_r, as in (30.1). Respectively, the dimension of the space $S(T, \bar{k})$ equals $r + 1 + k_1 + \cdots + k_{N-1}$. Similar representation is valid for the case $T = \{t_j\}_{j=-\infty}^{\infty}$.

31 B-splines

Suppose $T = \{t_j\}_{j=0}^N$ is a partition, and consider additional knots $t_{-r} < \cdots < t_{-1} < t_0$ and $t_N < t_{N+1} < \cdots < t_{N+r}$.

Definition 31.1. For $-r \le j \le N-1$, the function $B_j(x) := (t_{j+r+1} - t_j)[t_j, \dots, t_{j+r+1}; \varphi_x]$, where $\varphi_x(t) := (t-x)_+^r$, is called a B-spline of degree r. (in honor of Carl de Boor)

Theorem 31.2. *B-splines $B_{-r}, \dots, B_{N-1}$ form a basis in the space $S_r(T)$. (see also Theorem 30.3)*

Proof. It is sufficient to show that every function from the basis of Theorem 30.3 can be expressed as a linear combination of B-splines. First we prove that for every $0 \le \nu \le N+r$

$$(t_\nu - x)_+^r = \sum_{j=-r}^{N-1} B_j(x)(t_\nu - t_{j+1})_+ \cdots (t_\nu - t_{j+r})_+, \quad x \in [t_0, t_N]. \tag{31.1}$$

Indeed, by Popoviciu's identity we obtain:

$$\begin{aligned}\frac{(t_\nu - x)_+^r}{(t_\nu - t_{-r}) \cdots (t_\nu - t_0)} &= [t_{-r}, \dots, t_0, t_\nu; \varphi_x] \\ &= \sum_{j=-r}^{N-1} (t_{j+r+1} - t_j)[t_j, \dots, t_{j+r+1}; \varphi_x][t_{-r}, \dots, t_0, t_\nu; \prod_{j,r+1}] \\ &= \sum_{j=-r}^{N-1} (t_{j+r+1} - t_j)[t_j, \dots, t_{j+r+1}; \varphi_x] \frac{(t_\nu - t_{j+1})_+ \cdots (t_\nu - t_{j+r})_+}{(t_\nu - t_{-r}) \cdots (t_\nu - t_0)},\end{aligned}$$

because $\varphi_x = 0$ at the points $t_{-r}, \dots, t_0$. We also have the following analog of (31.1):

$$(t-x)^r = \sum_{j=-r}^{N-1} B_j(x)(t - t_{j+1}) \cdots (t - t_{j+r}), \quad x \in [t_0, t_N]. \tag{31.2}$$

Indeed, both sides of this equality are polynomials of degree r with respect to t. Bisedes, (31.1) implies that these polynomials coincide for $t_N \le t \le t_{N+r}$, and, hence, for all t.

By (31.2) we see that all polynomials of degree $\leqslant r$ can be expressed as a linear combination of B-splines on $[t_0, t_N]$. To apply 30.3 we only need to use (31.1), and the fact that $(x-t_j)_+^r = (t_j-x)_+^r$ for even r, and $(x-t_j)_+^r = (t_j-x)_+^r + (t_j-x)^r$ for odd r. □

Now we consider some properties of B-splines.

Property 1. $\operatorname{supp} B_j \subset [t_j, t_{j+r+1}]$, $-r \le j \le N-1$.

Proof. If $x > t_{j+r+1}$, then $\varphi_x(t) = (t-x)_+^r = 0$, for $t \in [t_j, t_{j+r+1}]$, and if $x < t_r$, then $\varphi_x(t) = (t-x)_+^r = (t-x)^r$, for $t \in [t_j, t_{j+r+1}]$. In both cases φ_x is a polynomial of degree $\leqslant r$ on $[t_j, t_{j+r+1}]$, therefore $B_j(x) = (t_{j+r+1}-t_j)[t_j, \dots, t_{j+r+1}; \varphi_x] = 0$. □

Property 2. $\displaystyle\sum_{j=-r}^{N-1} B_j(x) = 1$, $x \in [t_0, t_N]$.

Proof. Differentiate (31.2) r times with respect to t. □

Property 3. $\displaystyle\int_{-\infty}^{\infty} B_j(x)\,dx = \frac{t_{j+r+1}-t_j}{r+1}$, $-r \le j \le N-1$.

Proof. Indeed, for $j \le \nu \le j+r+1$ we have

$$\int_{t_j}^{t_{j+r+1}} (t_\nu - x)_+^r\,dx = \frac{1}{r+1}(t_\nu - t_j)^{r+1}.$$

In view of $[t_j, \dots, t_{j+r+1}; (\cdot - t_j)^{r+1}] = \dfrac{1}{(r+1)!}((\cdot - t_j)^{r+1})^{(r+1)}|_\theta = 1$, we obtain

$$\int_{-\infty}^{\infty} B_j(x)\,dx = \int_{t_j}^{t_{j+r+1}} B_j(x)\,dx = \frac{t_{j+r+1}-t_j}{r+1}[t_j, \dots, t_{j+r+1}; (\cdot - t_j)^{r+1}] = \frac{t_{j+r+1}-t_j}{r+1}.$$

□

Remark 31.1. Because of property 3, B-splines are sometimes normed in a different way, considering $M_j(x) := \dfrac{(r+1)B_j(x)}{t_{j+r+1}-t_j}$, then $\displaystyle\int_{-\infty}^{\infty} M_j(x)\,dx = 1$, which has certain probabilistic sense.

Property 4. $B_j(x) \geqslant 0$, $x \in \mathbb{R}$, $-r \le j \le N-1$.

Proof. Indeed, $(\varphi_x = \varphi_x(t) = (t-x)_+^r)$

$$\begin{aligned} B_j(x) = [t_j, \dots, t_{j+r+1}; \varphi_x] &= \frac{[t_{j+1}, \dots, t_{j+r+1}; \varphi_x] - [t_j, \dots, t_{j+r}; \varphi_x]}{t_{j+r+1}-t_j} \\ &= \frac{1}{t_{j+r+1}-t_j} \int_0^1 \cdots \int_0^{u_{r-1}} (\varphi_x^{(r)}(t_{j+1} + (t_{j+2}-t_{j+1})u_1 + \cdots + (t_{j+r+1}-t_{j+r})u_r) \\ &\qquad - \varphi_x^{(r)}(t_j + (t_{j+1}-t_j)u_1 + \cdots + (t_{j+r}-t_{j+r-1})u_r))\,du_r \dots du_1 \geqslant 0, \end{aligned}$$

since $\varphi_x^{(r)}(t) = r!(t-x)_+^0$ is a non-decreasing function. □

Property 5. *B-splines form a basis with the smallest support,* i.e., *if the support of a spline $s \in S_r(T)$ contains less than $r+2$ knots of the partition T, then $s \equiv 0$.*

Proof. Case $r = 3$. Assuming $\operatorname{supp} s \subset [t_0, t_3]$, by Rolle's theorem s'' has at least two zeroes $\theta_1 < \theta_2$ in (t_0, t_3). But s'' is a continuous piecewise-linear spline. If $\theta_1 \in (t_0, t_1]$, then $s''(x) = 0$, $x \in [t_0, t_1]$, and then $\operatorname{supp} s \subset [t_1, t_2]$, and similarly there exists a zero of s'' in, say, $[t_1, t_2]$, then $\operatorname{supp} s \subset [t_2, t_3]$, and so $s \equiv 0$. Similarly, one considers the case $\theta_2 \in [t_2, t_3)$. So, we can assume that $\theta_1, \theta_2 \in [t_1, t_2]$. But then $s''(x) = 0$, $x \in [t_1, t_2]$, and continuity and linearity of s'' on $[t_0, t_1]$ and on $[t_2, t_3]$ provides $s'' \equiv 0$ and $s \equiv 0$. □

32 Estimates for Approximation by Periodic Splines

Definition 32.1. Periodic spline (with equidistant knots) of degree r of minimal defect is a 2π-periodic function $S_{r,n} \subset C^{r-1}(\mathbb{R})$ such that $S_{r,n}$ is an algebraic polynomial of degree $\leqslant r$ on each of the intervals $[j\pi/n, (j+1)\pi/n]$, $j = 0, \pm 1, \pm 2, \ldots$.

Definition 32.2. Periodic spline $S_{r,n}$ is interpolatory for a 2π-periodic continuous function f if

$$\begin{cases} f(j\frac{\pi}{n}) = S_{r,n}(j\frac{\pi}{n}), \quad j = 0, \pm 1, \pm 2, \ldots, & \text{if } r \text{ — odd}, \\ f((j+\frac{1}{2})\frac{\pi}{n}) = S_{r,n}((j+\frac{1}{2})\frac{\pi}{n}), \quad j = 0, \pm 1, \pm 2, \ldots, & \text{if } r \text{ — even}. \end{cases}$$

Theorem 32.3 (Tikhomirov's)**.** *If $f \in W^r(\mathbb{R})$ is 2π-periodic, then for the interpolatory periodic spline $S_{r-1,n}$ the inequality*

$$\|f - S_{r-1,n}\|_{L_\infty(\mathbb{R})} \leqslant \frac{K_r}{n^r} \left\| f^{(r)} \right\|_{L_\infty(\mathbb{R})}$$

holds, where K_r is the Favar's constant.

Proof. For $r = 1$ the statement is obvious. Substitution $g(x) = f(x/n)$, $x \in \mathbb{R}$, allows to reduce the theorem to the following: for every $2\pi n$-periodic $g \in W^{(r)}(\mathbb{R})$ such that $\left\| g^{(r)} \right\|_{L_\infty(\mathbb{R})} \leqslant 1$, the interpolatory $2\pi n$-periodic spline $\sigma_{r-1,n}$ of degree $\leqslant r-1$ and minimal defect with the knots $j\pi$, $j = 0, \pm 1, \pm 2, \ldots$ (interpolatory as in the previous definition — for odd r at the knots, for even — between the knots) satisfies

$$\|g - \sigma_{r-1,n}\|_{C(\mathbb{R})} \leqslant K_r. \tag{32.1}$$

We will use the Euler's ideal splines ε_r, recall that $\|\varepsilon_r\|_{L_\infty(\mathbb{R})} = K_r$. Consider the case when r is odd, the other one is similar. We have $g(j\pi) = \sigma_{r-1,n}(j\pi)$, $\varepsilon_r(j\pi) = 0$, $j = 0, \pm 1, \pm 2, \ldots$, and if (32.1) is not true, then $\|g - \sigma_{r-1,n}\|_{C(\mathbb{R})} > K_r = \|\varepsilon_r\|_{L_\infty(\mathbb{R})}$, and there exist $x^* \in \mathbb{R}$, $x^* \neq j\pi$, $j = 0, \pm 1, \pm 2, \ldots$ and λ, $|\lambda| < 1$ such that $y := \varepsilon_r - \lambda(g - \sigma_{r-1,n})$ is zero at the point x^*. So, y has $\geqslant 2n+1$ zeroes on $(a, a+2\pi n]$. Hence, $y^{(r-2)}$ also has at least $2n+1$ zeroes on $(a, a+2\pi n]$, and is continuous. For $x \in (j\pi, (j+1)\pi)$ we compute $y^{(r)}(x) = \varepsilon_r^{(r)}(x) - \lambda g^{(r)}(x) = \pm 1 - \lambda g^{(r)}(x)$ — doesn't change the sign. So, on each interval $(j\pi, (j+1)\pi)$, the function $y^{(r)}$ is either strictly positive, or strictly negative, meaning that y^{r-2} is, respectively, strictly convex, or strictly concave, depending on the parity of j. We get a contradiction due to the next lemma (proof by induction as an exercise).

Lemma 32.4. *Let $F \in C[0, 2\pi n]$ be such that $F(0) = F(2\pi n)$, and $(-1)^j F$ is strictly convex on $(j\pi, (j+1)\pi)$, $0 \le j \le n-1$. Then F has $\leqslant 2n$ zeroes on $(0, 2\pi n]$.*

In our case $F = g^{(r-2)}$ has $\geqslant 2n+1$ zeroes. □

Remark 32.1. Subbotin proved: if a 2π-periodic $f \in C(\mathbb{R})$, then the interpolatory 2π-periodic spline $S_{r-1,n}$ (according to 32.2) satisfies

$$\|f - S_{r-1,n}\|_{L_\infty(\mathbb{R})} \leqslant c(r)\omega_r(1/n, f). \tag{32.2}$$

Moreover, if $f \in C^k$, $1 \le k \le r-1$, then for all $0 \le j \le k$

$$\left\|f^{(j)} - S_{r-1,n}^{(j)}\right\|_{L_\infty(\mathbb{R})} \leqslant \frac{c(r)}{n^j}\omega_{r-j}(1/n, f^{(j)}), \tag{32.3}$$

and we have an estimate with simultaneous approximation of the derivatives. Inequality (32.2) implies immediately the Tikhomirov's theorem, but the constant is not sharp. Using the Tikhomirov's theorem and some additional arguments, we will prove (32.2) in Section 35.

Exercise 32.1. *Prove that for any periodic spline $S_{r-1,n}$ with n equidistant knots of degree $r-1$ of minimal defect we have*

$$\left\|S_{r-1,n}^{(r)}\right\|_{L_\infty(\mathbb{R})} \leqslant \frac{n^r}{K_r}\max_{j\in\mathbb{Z}}|S_{r-1,n}(j\pi/n)| \leqslant \frac{n^r}{K_r}\|S_{r-1,n}\|_{L_\infty(\mathbb{R})}.$$

33 Widths

Let A, B be subsets of a normed space X. The quantity

$$\sup_{x\in A}\inf_{y\in B}\|x-y\|_X$$

means the *deviation* of A from B, *i.e.,* how well can one approximate elements of the set B using the elements of the set A. Most frequently we approximate be finite-dimensional subspaces, so it is natural to ask: how can we choose a finite-dimensional subspace $X_n = B$ to minimize the deviation for a fixed set A?

Definition 33.1. Let X be a normed space, $K \subset X$. Kolmogorov's width of order n is the quantity

$$d_n(K, X) := \inf_{\substack{X_n\subset X\\ \dim X_n \leqslant n}}\sup_{x\in K}\inf_{y\in X_n}\|x-y\|_X,$$

where the left infimum is taken over all linear subspaces $X_n \subset X$ of dimension $\leqslant n$.

A very popular problem in the theory of widths is to find widths of Sobolev classes. For Sobolev space $\widetilde{W}^r$ of periodic functions, the corresponding Sobolev class is the unit ball $\widetilde{B}^r$ in the seminorm of the space $\widetilde{W}^r$, namely

$$\widetilde{B}^r := \{f \in \widetilde{W}^r : \left\|f^{(r)}\right\|_{L_\infty(\mathbb{R})} \leqslant 1\}.$$

In many cases it is hard to find the exact value of the width, but it is possible to find it up to a constant factor, in particular, an asymptotic behavior is found as $n \to \infty$. For $\widetilde{B}^r$, an exact value of the width is known.

Theorem 33.2 (Tikhomirov)**.**

$$d_{2n}(\widetilde{B}^r, \widetilde{L}_\infty) = K_r n^{-r}, \quad n \ge 1.$$

Proof. The upper estimate immediately follows from the Tikhomirov's theorem on approximation by interpolatory periodic spline $S_{r-1,n}$. We will prove the lower bound. It is easy to check that since $\widetilde{C}$ is dense in $\widetilde{L}_\infty$, then one can choose subspaces X_{2n} from $\widetilde{C}$. Let $X_{2n} \subset \widetilde{C}$ be the linear span of the functions $f_1, f_2, \dots, f_{2n}$. Choose $\tau \in [0, \pi/n]$ so that

$$D(\tau) := \det[f_j(k\pi/n + \tau)]_{j,k=1}^{2n} = 0.$$

This is possible by continuity and the fact that the matrix for $D(\pi/n)$ can be obtained from the matrix for $D(0)$ by a cyclic permutation of the columns, so $D(\pi/n) = -D(0)$. Define the functional $l : \widetilde{C} \to \mathbb{R}$

$$l(f) := \sum_{j=1}^{2n} c_j f(j\pi/n + \tau),$$

where $c_j \in \mathbb{R}$, $1 \le j \le 2n$, are chosen to satisfy

$$l(f_j) = 0, \quad 1 \le j \le 2n, \quad |c_1| + \cdots + |c_{2n}| = \|l\| = 1.$$

Let S be a periodic spline of degree $r-1$ with n equidistant knots, at which $S(j\pi/n) = K_r n^{-r} \operatorname{sign} c_j$, $1 \le j \le 2n$. Put $f_0(t) := S(t - \tau) \in \widetilde{B}^r$, according to Exercise 32.1. For arbitrary $g \in X_{2n}$ we have

$$\|f_0 - g\|_{\widetilde{C}} \geqslant |l(f_0 - g)| = |l(f_0)| = \frac{K_r}{n^r},$$

as required. □

Remark 33.1. The widths of Sobolev classes are well investigated, in different integral metrics. There are modifications of the definition of the width that impose certain restrictions on approximation process, such as: linearity of approximation method, preserving properties of approximated function, etc.

34 Extremal Properties of Splines. Uniqueness

We begin with cubic splines. Let $a = x_0 < x_1 < \cdots < x_N = b$ be a fixed partition. To define an interpolatory cubic spline S nof minimal defect ($S \in C^2[a, b]$) on this partition, we need two boundary conditions (except for interpolatory and smoothness conditions).

Definition 34.1. Two boundary conditions are called admissible if for any two splines S_1, S_2, that satisfy the conditions, one has

$$\sum_{j=1}^{N} \int_{I_j} (S_1'' - S_2'') S'' \, dx = (S_1'(b) - S_2'(b)) S_2''(b) - (S_1'(a) - S_2'(a)) S_2''(a) = 0,$$

where $I_j := [x_{j-1}, x_j]$, $1 \le j \le N$.

Most common boundary conditions are the following three:

1) $S'(a) = S'(b)$ and $S''(a) = S''(b)$ (periodicity);

2) $S'(a) = A$, $S'(b) = B$;

3) $S''(a) = S''(b) = 0$.

Lemma 34.2. *Suppose a function $f \in W^2_{L_2[a,b]}$ and a cubic spline S of minimal defect and the knots $x_0, \dots, x_N$ satisfy the same two admissible boundary conditions, and $S(x_j) = f(x_j)$, $0 \le j \le N$. Then*

$$\|f''\|^2_{L_2[a,b]} = \|S''\|^2_{L_2[a,b]} + \|f'' - S''\|^2_{L_2[a,b]}.$$

Proof. The required identity is obviously equivalent to the orthogonality of f'' and $f'' - S''$ on $I := [a, b]$, *i.e.,* we need to verify that

$$\int_I (f'' - S'')S''\,dx = 0, \quad \Leftrightarrow \quad \sum_{j=1}^N \int_{I_j} (f'' - S'')S''\,dx = 0,$$

where $I_j := [x_{j-1}, x_j]$, $1 \le j \le N$. Using integration by parts, admissibility of the boundary conditions, and interpolation, we obtain

$$\begin{aligned}\int_{I_j} (f'' - S'')\,dx &= (f' - S')S''\big|_{x_{j-1}}^{x_j} - \int_{I_j} (f' - S')S'''\,dx \\ &= (f' - S')S''\big|_{x_{j-1}}^{x_j} - (f - S)S'''\big|_{x_{j-1}}^{x_j} + \int_{I_j} (f - S)S^{(IV)}\,dx \\ &= (f' - S')S''\big|_{x_{j-1}}^{x_j},\end{aligned}$$

and so

$$\sum_{j=1}^N \int_{I_j} (f'' - S'')S'' = (f'(b) - S'(b))S''(b) - (f'(a) - S'(a))S''(a) = 0,$$

as required. □

Corollary 34.3 (Holiday). *From all functions $f \in W^2_{L_2[a,b]}$ satisfying $f(x_j) = y_j$, $0 \le j \le N$, for a fixed data $y_0, \dots, y_N$, and two admissible boundary conditions (for example, one of 1)–3)), the interpolatory cubic spline of minimal defect minimizes the quantity $\|f''\|_{L_2[a,b]}$.*

Corollary 34.4 (uniqueness). *Interpolatory cubic spline of minimal defect satisfying one of the boundary conditions 1)–3) is unique.*

Proof. Assume there are two such splines: S_1 and S_2. From Lemma, $\|S_1'' - S_2''\|_{L_2[a,b]} = 0$. Since $S_1'' - S_2''$ is piecewise linear, then $S_1'' \equiv S_2''$, and so $S_1 - S_2$ is a linear function. But by interpolation the difference $S_1 - S_2$ is zero at all the knots, so at least at two points, which means $S_1 \equiv S_2$. □

Remark 34.1. All statements in this section are valid for splines of minimal defect of degree $2m - 1$, proofs are similar. The corresponding conditions become:

1)′ $S'(a) = S'(b)$, $S''(a) = S''(b)$, $\dots$, $S^{(2m-2)}(a) = S^{(2m-2)}(b)$;

2)′ $S'(a) = A_1$, $S'(b) = B_1$, $\dots$, $S^{(m-1)}(a) = A_{m-1}$, $S^{(m-1)}(b) = B_{m-1}$.

Respectively, we have the identity

$$\left\|f^{(m)}\right\|^2_{L_2[a,b]} = \left\|S^{(m)}\right\|^2_{L_2[a,b]} + \left\|f^{(m)} - S^{(m)}\right\|^2_{L_2[a,b]}.$$

35 K-functionals and Examples of Applications

One of the ways to measure smoothness of functions, besides the moduli of smoothness, are K-functionals, introduced by Peetre.

Definition 35.1. Let $X_1 \subset X_0$ be continuously embedded Banach spaces. For $f \in X_0$, K-functional is defined as

$$K(f,t;X_0,X_1) := \inf_{g\in X_1}(\|f-g\|_{X_0} + t\,\|g\|_{X_1}), \quad t \geqslant 0. \tag{35.1}$$

We measure how "well" can we replace a function f from a more "complicated" space X_0 by a function g from a more "simple" space X_1. We will limit ourselves to the simplest situations $X_0 = C(I)$, $X_1 = W^r_\infty(I)$ for an interval I, and $X_0 = \widetilde{C}$, $X_1 = \widetilde{W}^r_\infty$ for periodic function on $\mathbb{R}$. For such spaces, it is convenient to consider a modification of (35.1), where in place of $\|g\|_{X_1} = \|g\|_{W^r}$ one takes the seminorm $|g|_{W^r} = \left\|g^{(r)}\right\|_{L_\infty}$, namely

$$K(f,t;C,W^r) := \inf_{g\in W^r}(\|f-g\|_C + t\left\|g^{(r)}\right\|_{L_\infty}).$$

K-functionals and moduli of smoothness have many similar properties. Besides, there is a direct relation between K-functionals and moduli of smoothness, which is useful in certain situations.

Theorem 35.2. *Let $C = C(I)$, $W^r = W^r_\infty(I)$ for an interval I, or $C = \widetilde{C}$, $W^r = \widetilde{W}^r_\infty$ for periodic functions on $\mathbb{R}$. Then for each $f \in C$, we have*

$$c_1 K(f,t^r;C,W^r) \leqslant \omega_r(t,f) \leqslant c_2 K(f,t^r;C,W^r), \tag{35.2}$$

where c_1 and c_2 are constants that depend only on r.

Proof. We will only consider the periodic case. The right hand side inequality in (35.2) is a simple corollary of the properties of moduli of smoothness, since for arbitrary $g \in W^r$ we have

$$\omega_r(t,g) \leqslant \omega_r(t,f-g) + \omega_r(t,g) \leqslant 2^r\,\|f-g\|_C + t^r\left\|g^{(r)}\right\|_{L_\infty} \leqslant 2^r K(f,t^r;C,W^r).$$

For the proof of the left hand side inequality, we will use a modification of Steklov means, setting

$$g_r(x) := \frac{1}{t^r}\int_0^t\cdots\int_0^t \sum_{j=1}^r (-1)^{j-1}\binom{r}{j} f(x + j(t_1+\cdots+t_r)/r)\,dt_1\ldots dt_r. \tag{35.3}$$

Then

$$|f(x) - g_r(x)| = \left|\frac{1}{t^r}\int_0^t\cdots\int_0^t \Delta^r_{(t_1+\cdots+t_r)/r}(f,x)\,dt_1\ldots dt_r\right| \leqslant \omega_r(t,f).$$

The expression in (35.3) is a linear combination of

$$\begin{aligned} g_{r,\nu}(x) :&= \frac{1}{t^r}\int_0^t\cdots\int_0^t f(x+\nu(t_1+\cdots+t_r))\,dt_1\ldots dt_r \\ &= \frac{1}{t^r}\int_{\frac{x}{r\nu}}^{\frac{x}{r\nu}+t}\cdots\int_{\frac{x}{r\nu}}^{\frac{x}{r\nu}+t} f(\nu(t_1+\cdots+t_r))\,dt_1\ldots dt_r, \end{aligned}$$

for $\nu = j/r$, $1 \le j \le r$. Differentiating, we obtain

$$f'_{r,\nu}(x) = r\frac{1}{t^r}\frac{1}{t\nu}\int_{\frac{x}{r\nu}}^{\frac{x}{r\nu}+t}\cdots\int_{\frac{x}{r\nu}}^{\frac{x}{r\nu}+t}\Big(f\big(\nu(\frac{x}{r\nu}+t+t_1+\cdots+t_{r-1})\big)$$

$$-f\big(\nu(\frac{x}{r\nu}+t_1+\cdots+t_{r-1})\big)\Big)\,dt_1\ldots dt_{r-1}$$

$$= \frac{1}{t\nu}(g_{r-1,\nu}(x+t\nu)-g_{r-1,\nu}(x)),$$

which is true for $r=1$ as well, if one takes $g_{0,\nu}(x) := f(x)$. So, repeating differentiation by the above formula, we get

$$|g^{(r)}_{r,\nu}(x)| = \left|\frac{1}{(t\nu)^r}\Delta^r_{t\nu}(f,x)\right| \leqslant \frac{c(r)}{t^r}\omega_r(t,f).$$

So, we can take $g := g_r$. □

Now we will consider examples of applications of the equivalence (35.2). By $c(r)$ we will denote different constants that depend on r.

Example 1. Whitney's inequality. Let $f \in C(I)$, $I=[0,1]$. According to (35.2), $t:=1$, there exists $g \in W^r_\infty(I)$ such that

$$\|f-g\|_{C(I)} + \left\|g^{(r)}\right\|_{L_\infty(I)} \leqslant c(r)\omega_r(1,f) \leqslant c(r)\omega_r(1/r,f).$$

Let P_{r-1} be the Taylor's polynomial of degree $r-1$ for g at the point 0. Using the Taylor's formula with the remainder in the Lagrange form, we obtain

$$|g(x)-P_{r-1}(x)| \leqslant \frac{1}{r!}\left\|g^{(r)}\right\|_{L_\infty(I)}, \quad x \in I.$$

So,

$$\|f-P_{r-1}\|_{C(I)} \leqslant \|f-g\|_{C(I)} + \|g-P_{r-1}\|_{C(I)} \leqslant c(r)\omega_r(1/r,f).$$

Example 2. Subbotin's estimate (32.2) for approximation by interpolatory splines in terms of modulus of smoothness. K-functional allows to reduce the problem to the Tikhomirov's theorem. Denote by $S_{r-1,n}f$ the interpolatory spline of degree $r-1$ for function f, see Definition 32.2. If $t := \frac{\pi}{2n}$, then we can find $g \in \widetilde{W}^r$ such that

$$\|f-g\|_{\widetilde{C}} + t^r\left\|g^{(r)}\right\|_{\widetilde{L}_\infty} \leqslant c(r)\omega_r(t,f) \leqslant c(r)\omega_r(1/n,f). \tag{35.4}$$

We obtain

$$\|f-S_{r-1,n}f\|_{\widetilde{C}} \leqslant \|f-g\|_{\widetilde{C}} + \|S_{r-1,n}(f-g)\|_{\widetilde{C}} + \|g-S_{r-1,n}g\|_{\widetilde{C}}.$$

If we prove that

$$\|S_{r-1,n}(f-g)\|_{\widetilde{C}} \leqslant c(r)\,\|f-g\|_{\widetilde{C}}, \tag{35.5}$$

then by Tikhomirov's theorem for g and (35.4), we get

$$\|f-S_{r-1,n}f\|_{\widetilde{C}} \leqslant c(r)\omega_r(1/n,f) + \|g-S_{r-1,n}g\|_{\widetilde{C}}$$

$$\leqslant c(r)\omega_r(1/n,f) + \frac{K_r}{n^r}\left\|g^{(r)}\right\|_{\widetilde{L}}$$

$$\leqslant c(r)\omega_r(1/n,f).$$

It remains to show (35.5). Take $h := f - g$. An auxiliary function $\sigma \in C^\infty(\mathbb{R})$

$$\sigma(x) := \begin{cases} e^{(x^2-1)^{-1}}, & x \in (-1,1), \\ 0, & x \notin (-1,1), \end{cases}$$

satisfies $\left\|\sigma^{(r)}\right\|_{C(\mathbb{R})} \leqslant c(r)$ and $\|\sigma\|_{C(\mathbb{R})} = 1$, $\sigma(0) = 1$. Since $S_{r-1,n}$ is determined by the values at the knots $\frac{j\pi}{n}$, $0 \le j \le 2n-1$, by setting

$$\tilde{h}(x) := \sum_{j=0}^{2n-1} h(j\pi/n) \sum_{k\in\mathbb{Z}} \sigma(xt^{-1} - 2k\pi), \quad x \in \mathbb{R},$$

(recall that $t = \frac{\pi}{2n}$) we obtain $S_{r-1,n}\tilde{h} = S_{r-1,n}h$, and $\tilde{h} \in \widetilde{W}^r$, $\left\|\tilde{h}\right\|_{\widetilde{C}} \leqslant \|h\|_{\widetilde{C}}$ § $\left\|\tilde{h}^{(r)}\right\|_{\widetilde{C}} \leqslant c(r)t^{-r}\|h\|_{\widetilde{C}} = c(r)n^r\|h\|_{\widetilde{C}}$. Taking this into account, by Tikhomirov's theorem

$$\begin{aligned} \|S_{r-1,n}h\|_{\widetilde{C}} = \left\|S_{r-1,n}\tilde{h}\right\|_{\widetilde{C}} &\leqslant \left\|\tilde{h} - S_{r-1,n}\tilde{h}\right\|_{\widetilde{C}} + \left\|\tilde{h}\right\|_{\widetilde{C}} \\ &\leqslant \frac{K_r}{n^r}\left\|\tilde{h}^{(r)}\right\|_{\widetilde{C}} + \|h\|_{\widetilde{C}} \\ &\leqslant c(r)\|h\|_{\widetilde{C}}, \end{aligned}$$

as required.

36 Neural Networks

It is hard to define specifically neural networks. In general, neural networks are certain computational models that are, in fact, far from a biological motivation. Neural networks have some common characteristics. Given certain input data $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, and a process that leads to evaluation of a result $y = (y_1, \ldots, y_m) \in \mathbb{R}^m$. This process is a general function $y = G(x)$, which can be very complicated. Moreover, we cannot compute exactly this unknown function G, but we are choosing a "candidate" — function $F(x) = F(x, w)$ from certain parametric set of functions using available information about G to help in selection of the parameters w. The main advantage of neural networks is that for fixed parameters w the function $F(x, w)$, as a rule, can be quickly evaluated using a sequence of simple operations. At the same time, F can (approximately) model a very complex functional dependance G. The drawback (and main problem) is that selection of proper parameters w can be very hard. The corresponding optimization problem for selection of w is solved by "training rules", that approximately find w, using available samples of values of G. Neural networks are used in the problems of classification and recognition of images, stock market predictions, and so on, where a significant amount of data for "training" is available. Practical problem is to find efficient algorithms for selection of parameters w. Theoretically, limits of application and approximation properties of the corresponding problems are investigated.

We consider one of the simplest models, the so called *MLP* (multilayer feedforward perceptron) model. It consists of a finite number of layers, that are formed by a finite number of computing elements, or neurons. Each neuron of every level is connected to all neurons of the following layer, that is, the result of computation of a neuron of a layer is available to all neurons of the following layer as input information. We distinguish the first (input) layer, the last (output) layer, and all layers in between are called hidden. Computation rules of the neurons are as follows. The neurons of the input layer obtain information $x_{0,j}$ and transfer it to the neurons of next layer $x_{1,j} = x_{0,j}$. In

i-th hidden layer, k-th neuron receives all the information $x_{i,j}$ from the previous layer, calculates the weighted sum $\sum_j w_{i,j,k}x_{i,j}$ with certain weights $w_{i,j,k}$. Then, a constant $\theta_{i,k}$ is subtracted from the result (translation) and substituted into a fixed function (activation function) σ. This creates an output of this neuron to be transferred to the neurons of the next layer

$$x_{i+1,k} = \sigma\left(\sum_j w_{i,j,k}x_{i,j} - \theta_{i,k}\right).$$

The output layer does not use translation and activation function, only computing the weighted sum. Neural network with m neurons in the output layer can be considered as m neural networks with one neuron in the output layer, so we will consider only such situation. We will also limit our consideration to one hidden layer consisting of r neurons. Then the result of computation of our model on the input data $x = (x_1, \dots, x_n)$ can be expressed as

$$y = \sum_{i=1}^{r} c_i\sigma\left(\sum_{j=1}^{n} w_{i,j}x_j - \theta_i\right),$$

where $w_{i,j}$ and c_i are the weights of the hidden and output layers respectively, and θ_i are translations.

The most popular activation functions are:

1) step function $\sigma(t) = \begin{cases} 1, & x \geqslant 0, \\ 0, & x < 0, \end{cases}$

2) logistic sigmoid $\sigma(t) = \dfrac{1}{1+e^{-t}}$,

3) piecewise linear function of the form $\sigma(t) = \begin{cases} 0, & t \leqslant -1, \\ (t+1)/2, & -1 \leqslant t \leqslant 1, \\ 1, & 1 \leqslant t, \end{cases}$

4) Gaussian sigmoid $\sigma(t) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{t} e^{-u^2/2}\,du$,

5) arctan sigmoid $\sigma(t) = \dfrac{1}{\pi}\arctan t + \dfrac{1}{2}$.

The term of sigmoidal function is applied to the functions $\sigma : \mathbb{R} \to \mathbb{R}$ such that $\lim\limits_{t\to-\infty} \sigma(t) = 0$ and $\lim\limits_{t\to\infty} \sigma(t) = 1$, sometimes adding the condition that σ is non-decreasing.

Consider the question of density of the MLP model with one hidden layer. Denote

$$M_n(\sigma) := \bigcup_{r\in\mathbb{N}} M_{n,r}(\sigma), \quad M_{n,r}(\sigma) := \Big\{ \sum_{i=1}^{r} c_i\sigma\left(\sum_{j=1}^{n} w_{i,j}x_j - \theta_i\right) : c_i, \theta_i, w_{i,j} \in \mathbb{R} \Big\},$$

the sets of all realizations (functions) of such model. A set G is dense in $C(\mathbb{R}^n)$, if for every compact set $K \subset \mathbb{R}^n$, for any $\varepsilon > 0$, $f \in C(K)$, there exists $g \in G$ such that $\|f - g\|_{C(K)} < \varepsilon$. It turns out that $M_n(\sigma)$ is dense in $C(\mathbb{R}^n)$ under very weak conditions on σ.

Theorem 36.1. *Suppose $\sigma \in C(\mathbb{R})$. $M_n(\sigma)$ is dense in $C(\mathbb{R}^n)$ if and only if σ is not a polynomial.*

Necessity is evident — if σ is a polynomial of degree k, then each element of $M_n(\sigma)$ is a polynomial of degree $\leqslant k$, and the set of polynomials of a fixed degree is not dense in $C(\mathbb{R}^n)$.

To prove sufficiency we will need the so called plane waves or ridge functions, which allow to reduce the problem of density of $M_n(\sigma)$ to the one-dimensional problem of density of $M_1(\sigma)$. Plane waves are functions of the form

$$g(a_1x_1 + \cdots + a_nx_n), \quad a_1, \ldots, a_n \in \mathbb{R},$$

where g is a function from $\mathbb{R}$ to $\mathbb{R}$. We will introduce a notation for functions that are sums of plane waves

$$R_n := \bigcup_{r\in\mathbb{N}} R_{n,r}, \quad R_{n,r} := \Big\{ \sum_{i=1}^{r} g_i \left(\sum_{j=1}^{n} a_{i,j}x_j \right) : a_{i,j} \in \mathbb{R}, g_i \in C(\mathbb{R}) \Big\}.$$

Clearly, $R_n \supset M_n(\sigma)$ for any continuous σ.

Exercise 36.1. *Prove that R_n is dense in $C(\mathbb{R}^n)$. (One can use that polynomials are dense, and every polynomial of n variables can be expressed as a sum of plane waves, created by polynomials of one variable.)*

This allows to reduce the problem to the one-dimensional problem.

Exercise 36.2. *Prove that $M_n(\sigma)$ is dense in $C(\mathbb{R}^n)$ if $M_1(\sigma)$ is dense in $C(\mathbb{R})$.*

Proof of Theorem 36.1. By the previous exercise, it is enough to prove that if $\sigma \in C(\mathbb{R})$ is not a polynomial, then $M_1(\sigma)$ is dense in $C(\mathbb{R})$. Assume that $\sigma \in C^\infty(\mathbb{R})$. We will need the following fact, which is a nice math competition problem.

Exercise 36.3. *If $\sigma \in C^\infty(\mathbb{R})$ and for for every $x \in \mathbb{R}$ there exists $k = k(x) \in \mathbb{N} \cup \{0\}$ such that $\sigma^{(k)}(x) = 0$, then σ is a polynomial.*

Hence, we have the opposite — there is a point $\theta^* \in \mathbb{R}$ such that $\sigma^{(k)}(-\theta^*) \neq 0$, for all $k \geq 0$. For every $h \neq 0$ we have that

$$\frac{\sigma((\lambda + h)t - \theta^*) - \sigma(\lambda t - \theta^*)}{h}$$

belongs to $M_1(\sigma)$, and by the smoothness of σ we obtain that

$$\frac{d}{d\lambda}\sigma(\lambda t - \theta^*)\bigg|_{\lambda=0} = t\sigma'(-\theta^*)$$

belongs to the closure of $M_1(\sigma)$. By similar arguments,

$$\frac{d^k}{d\lambda^k}\sigma(\lambda t - \theta^*)\bigg|_{\lambda=0} = t^k\sigma^{(k)}(-\theta^*)$$

belongs to the closure of $M_1(\sigma)$, for every k. Since all derivatives of σ at the point $-\theta^*$ are non-zero, the closure of $M_1(\sigma)$ contains all polynomials, and hence, all continuous functions. The case $\sigma \in C(\mathbb{R})$ can be reduced to just proved case $\sigma \in C^\infty(\mathbb{R})$ using convolutions with appropriate smooth functions. ☐

References

[1] J. H. Ahlberg, E. N. Nilson, and J. L. Walsh, *The theory of splines and their applications*, Academic Press, New York, 1967.

[2] N. I. Akhiezer, *Theory of approximation*, Dover Publications, Inc., New York, 1992. Translated from the Russian and with a preface by Charles J. Hyman; Reprint of the 1956 English translation.

[3] E. W. Cheney, *Introduction to approximation theory*, AMS Chelsea Publishing, Providence, RI, 1998. Reprint of the second (1982) edition.

[4] R. A. DeVore and G. G. Lorentz, *Constructive approximation*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 303, Springer-Verlag, Berlin, 1993.

[5] Z. Ditzian and V. Totik, *Moduli of smoothness*, Springer Series in Computational Mathematics, vol. 9, Springer-Verlag, New York, 1987.

[6] V. K. Dzyadyk and I. A. Shevchuk, *Theory of uniform approximation of functions by polynomials*, Walter de Gruyter GmbH & Co. KG, Berlin, 2008. Translated from the Russian by Dmitry V. Malyshev, Peter V. Malyshev and Vladimir V. Gorunovich.

[7] N. Korneichuk, *Exact constants in approximation theory*, Encyclopedia of Mathematics and its Applications, vol. 38, Cambridge University Press, Cambridge, 1991. Translated from the Russian by K. Ivanov.

[8] G. G. Lorentz, *Bernstein polynomials*, Mathematical Expositions, no. 8, University of Toronto Press, Toronto, 1953.

[9] G. G. Lorentz, M. v. Golitschek, and Yu. Makovoz, *Constructive approximation*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 304, Springer-Verlag, Berlin, 1996. Advanced problems.

[10] P. P. Petrushev and V. A. Popov, *Rational approximation of real functions*, Encyclopedia of Mathematics and its Applications, vol. 28, Cambridge University Press, Cambridge, 1987.

[11] A. Pinkus, *Approximation theory of the MLP model in neural networks*, Acta numerica, 1999, Acta Numer., vol. 8, Cambridge Univ. Press, Cambridge, 1999, pp. 143–195.

[12] M. J. D. Powell, *Approximation theory and methods*, Cambridge University Press, Cambridge-New York, 1981.

[13] T. J. Rivlin, *An introduction to the approximation of functions*, Dover Publications, Inc., New York, 1981. Corrected reprint of the 1969 original; Dover Books on Advanced Mathematics.

[14] A. F. Timan, *Theory of approximation of functions of a real variable*, Dover Publications Inc., New York, 1994. Translated from the Russian by J. Berry; Translation edited and with a preface by J. Cossar; Reprint of the 1963 English translation.

K. A. Kopotun
Department of Mathematics
University of Manitoba
Winnipeg, Manitoba R3T 2N2
Canada
kirill.kopotun@umanitoba.ca

A. V. Prymak
Department of Mathematics
University of Manitoba
Winnipeg, Manitoba R3T 2N2
Canada
prymak@cc.umanitoba.ca

I. A. Shevchuk
National Taras Shevchenko University of Kyiv
Kyiv, Ukraine
shevchukh@ukr.net