Definition. We define the norm of a bounded linear operator L to be the number $\|L\| := \inf\{K \in \mathbb{R} \mid \|Lx\| \leq K\|x\| \text{ for all } x \in \mathrm{D}\}$. With this definition, we have

$$\|Lx\| \leq \|L\|\,\|x\|.$$

Fact. $\|L\| = \sup_{0 \neq x \in \mathrm{D}} \frac{\|Lx\|}{\|x\|}$. To see this, note that $\frac{\|Lx\|}{\|x\|} \leq \|L\|$ for all $x \in \mathrm{D}, x \neq 0$, and in particular $\sup_{0 \neq x \in \mathrm{D}} \frac{\|Lx\|}{\|x\|} =: M \leq \|L\|$. But, on the other hand $\|Lx\| = \frac{\|Lx\|}{\|x\|}\|x\| \leq M\|x\|$ for $0 \neq x \in \mathrm{D}$, and thus $\|L\| \leq M$. This shows that $M \leq \|L\| \leq M$, and the assertion is established. □

The norm of L can also be written in the form $\|L\| = \sup_{\|x\|=1} \|Lx\|$. It is easy to show that the mapping $\|L\|$ satisfies the properties of a norm. Moreover, the *product* $(L_1L_2)x := L_1(L_2x)$ of two linear operators L_1 and L_2 satisfies the inequality

$$\|L_1L_2\| \leq \|L_1\|\,\|L_2\|,$$

since $\|(L_1L_2)x\| \leq \|L_1\|\,\|L_2x\| \leq \|L_1\|\,\|L_2\|\,\|x\|$.

Application. We consider once again the two examples of linear operators given above.

Example 1. The integral operator $J : \mathrm{C}[a,b] \to \mathbb{R}$ on the space $(\mathrm{C}[a,b], \|\cdot\|_\infty)$ is a bounded linear operator since

$$|Jf| = |\int_a^b w(x)f(x)dx| \leq \int_a^b w(x)dx\|f\|_\infty \quad \text{for } w(x) > 0 \text{ in } (a,b),$$

and thus $\|J\| = \sup_{\|f\|_\infty=1} |Jf| \leq \int_a^b w(x)dx$. Since J is a mapping into $\mathbb{R}$, it is in fact a bounded linear functional.

We also note that for the element $f^* := 1$, the estimate $\sup_{\|f\|_\infty=1} |Jf| \geq$ $\geq |Jf^*| = \int_a^b w(x)dx$ holds, and thus $\|J\| \geq \int_a^b w(x)dx$. Combining these facts, we conclude that the norm of J is given by $\|J\| = \int_a^b w(x)dx$.

Example 2. In view of the results in 2.4.2, it follows that every finite-dimensional matrix is a bounded linear operator. Various matrix norms were calculated in 2.4.3.

1.6 Problems. 1) Show that the mapping

$$a : \mathrm{C}_1[0,1] \to \mathbb{R},\ a(f) := \Big(\int_0^1 |f'(x)|^2 w(x)dx\Big)^{\frac{1}{2}} + \sup_{x\in[0,1]} |f(x)|$$

defines a norm on $\mathrm{C}_1[0,1]$. Is this norm strict if $w(x) := 1$?

2) Let $\|\cdot\|_a$ and $\|\cdot\|_b$ be norms on the linear space V and suppose that $\|\cdot\|_a$ is strict. Show that the norm $\|v\| := \|v\|_a + \|v\|_b$ on V is also strict.

3) Show that the mapping

$$a : \mathrm{C}_m(\overline{\mathrm{G}}) \to \mathbb{R},\ a(f) := \sum_{|\gamma| \leq m} \max_{x \in \overline{\mathrm{G}}} |D^\gamma f(x)|$$

defines a norm on the linear space $\mathrm{C}_m(\overline{\mathrm{G}})$, and that $\mathrm{C}_m(\overline{\mathrm{G}})$ equipped with this norm is a Banach space.

4) Let $(\mathrm{V}, \|\cdot\|)$ be a normed linear space over $\mathbb{R}$. Show that the norm $\|\cdot\|$ is induced by an inner product $\langle\cdot,\cdot\rangle$ if and only if the "parallelogram law"

$$\|f+g\|^2 + \|f-g\|^2 = 2(\|f\|^2 + \|g\|^2)$$

holds for all $f, g \in \mathrm{V}$. Note that in $(\mathbb{R}^2, \|\cdot\|_2)$, the parallelogram law with $\langle x,y\rangle = 0$ reduces to the Pythagorean Theorem.

Hint: Assume $\langle f,g\rangle := \frac{1}{4}(\|f+g\|^2 - \|f-g\|^2)$.

5) By investigating the convergence of the sequence $(f_n)_{n\in\mathbb{Z}_+}$ defined on $[a,b] := [-1,+1]$ by

$$f_n(x) := \begin{cases} -1 & \text{for } x \in [-1, -\frac{1}{n}] \\ nx & \text{for } x \in [-\frac{1}{n}, +\frac{1}{n}] \\ 1 & \text{for } x \in [\frac{1}{n}, 1], \end{cases}$$

show that the linear space $\mathrm{C}[a,b]$ is not complete with respect to either the norm $\|\cdot\|_2$ or the norm $\|\cdot\|_1$.

6) Show that the mapping $Ff := \sum_1^n \alpha_\nu f(x_\nu)$, $\alpha_\nu \in \mathbb{R}$ defined for functions $f \in \mathrm{C}[a,b]$ is a bounded linear functional on the normed linear space $(\mathrm{C}[a,b], \|\cdot\|_\infty)$, and that $\|F\| = \sum_1^n |\alpha_\nu|$.

2. The Approximation Theorems of Weierstrass

We begin our discussion of approximation theory with the classical problem of approximating a function. A more general approximation problem will be treated later in this chapter. In this section we shall present several approximation theorems of Weierstrass which show how to approximate an arbitrary continuous function by simple functions.

2.1 Approximation by Polynomials. It is known from calculus that an analytic function f can be written as a power series

$$f(x) = a_0 + a_1x + \cdots + a_nx^n + \cdots$$

which uniformly converges to the function f inside a certain convergence interval.

Consider now the sequence $(\sigma_n)_{n\in\mathbb{N}}$ of partial sums of the power series defined by

$$\sigma_n(x) = a_0 + a_1x + \cdots + a_nx^n.$$

Then it is clear that for every $\varepsilon > 0$, there exists a number $N(\varepsilon) \in \mathbb{N}$ such that $\|f - \sigma_n\|_\infty < \varepsilon$ for every $n > N$. In other words, for any given interval, there always exists a polynomial which uniformly approximates the analytic function arbitrarily well.

It is now natural to ask whether a similar assertion still holds if we assume only that f is continuous. In general, such an approximation cannot be in the form of a power series, since as is well known, power series represent functions which are infinitely differentiable, whereas certainly not every continuous function f has derivatives.

We answer this question in the following section by establishing the classical approximation theorem of Weierstrass. Although we shall later discuss a more general theorem of Korovkin, it is worthwhile to first present the original Weierstrass Theorem with a direct proof. Indeed, in this way we can formulate the theorem in a simple instructive way, and moreover, we can give a constructive proof due to S. N. BERNSTEIN in 1912 which serves to motivate the later results of P. P. KOROVKIN.

KARL WEIERSTRASS (1815–1897) established his approximation theorems in the paper "Über die analytische Darstellbarkeit sogenannter willkürlicher Funktionen reeller Argumente" (Sitzg. ber. Kgl. Preuss. Akad. d. Wiss. Berlin 1885, pp. 663–639, 789–805). He gave non-constructive proofs of his theorems. Weierstrass became famous primarily for his fundamental results in analysis. He is considered to be one of the founders of modern function theory; the starting point of his work is the power series. In addition, Weierstrass fully understood the great importance of mathematics for applications to problems in physics and astronomy. For this reason he gave mathematics a leading position,"since through it alone can one obtain a truely satisfactory understanding of nature". (Quote from I. Runge ([1949], p. 29)).

Because of its potential applications, we now present S. N. Bernstein's constructive proof of the approximation theorem for continuous functions. The so-called Bernstein polynomials which appear in the proof came originally from probability theory.

Before proceeding, we mention that there are a series of alternative proofs of these approximation theorems, for example by E. LANDAU (1908), H. LEBESGUE (1908), and others. We also mention a generalization to topological spaces due to M. H. STONE (1948).

2.2 The Approximation Theorem for Continuous Functions. In this section we prove that every continuous function on a given finite closed interval can be uniformly approximated arbitrarily well by a polynomial. This means that the polynomials are dense in the space $C[a,b]$ of continuous functions.

Let P_n denote the $(n+1)$-dimensional linear space of all polynomials of maximal degree n over the field $\mathbb{R}$, defined by

$$P_n := \{p \in C(-\infty, +\infty) \mid p(x) = \sum_{\nu=0}^{n} a_\nu x^\nu\}.$$

Approximation Theorem of Weierstrass. *Let $-\infty < a < b < +\infty$, and suppose that $f \in C[a,b]$ is an arbitrary continuous function. Then for every $\varepsilon > 0$, there exists an $n \in \mathbb{N}$ and a polynomial $p \in P_n$ such that $\|f - p\|_\infty < \varepsilon$.*

Proof. Since every interval $[a,b]$ can be mapped onto $[0,1]$ by a linear transformation, we may restrict our attention to the case $[a,b] := [0,1]$. To establish the theorem, we shall show that the sequence $(B_n f)$ of *Bernstein-Polynomials*

$$(B_n f)(x) := \sum_{\nu=0}^{n} f\Big(\frac{\nu}{n}\Big)\binom{n}{\nu} x^\nu (1-x)^{n-\nu}, \quad n = 1, 2, \cdots,$$

converges uniformly to f on $[0,1]$.

First we note that $(B_n f)(0) = f(0)$ and $(B_n f)(1) = f(1)$ for all n. Now

$$1 = [x + (1-x)]^n = \sum_{\nu=0}^{n} \binom{n}{\nu} x^\nu (1-x)^{n-\nu} =: \sum_{\nu=0}^{n} q_{n\nu}(x)$$

implies

$$(*) \qquad f(x) - (B_n f)(x) = \sum_{\nu=0}^{n} \Big[f(x) - f\Big(\frac{\nu}{n}\Big)\Big] q_{n\nu}(x),$$

and thus

$$|f(x) - (B_n f)(x)| \le \sum_{\nu=0}^{n} \Big| f(x) - f\Big(\frac{\nu}{n}\Big) \Big| \, q_{n\nu}(x)$$

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

By the (uniform) continuity of f, for every $\varepsilon > 0$ there exists a number δ, not depending on x, such that $|f(x) - f(\frac{\nu}{n})| < \frac{\varepsilon}{2}$ for all points x with $|x - \frac{\nu}{n}| < \delta$.

For every $x \in [0,1]$, consider the two sets

$$N' := \Big\{\nu \in \{0, 1, \ldots, n\} : \Big| x - \frac{\nu}{n} \Big| < \delta\Big\}$$

$$N'' := \Big\{\nu \in \{0, 1, \ldots, n\} : \Big| x - \frac{\nu}{n} \Big| \ge \delta\Big\},$$

and split the sum into two parts $\sum_{\nu=0}^{n} = \sum_{\nu\in N'} + \sum_{\nu\in N''}$. Then the first sum satisfies

$$\sum_{\nu\in N'} \left| f(x) - f\left(\frac{\nu}{n}\right) \right| q_{n\nu}(x) \le \frac{\varepsilon}{2} \sum_{\nu\in N'} q_{n\nu}(x) \le \frac{\varepsilon}{2} \sum_{\nu=0}^{n} q_{n\nu}(x) = \frac{\varepsilon}{2}.$$

Moreover, with $M := \max_{x\in[0,1]} |f(x)|$ we also have

$$\sum_{\nu\in N''} \left| f(x) - f\left(\frac{\nu}{n}\right) \right| q_{n\nu}(x) \le \sum_{\nu\in N''} \left| f(x) - f\left(\frac{\nu}{n}\right) \right| q_{n\nu}(x) \frac{(x-\frac{\nu}{n})^2}{\delta^2} \le$$

$$\le \frac{2M}{\delta^2} \sum_{\nu=0}^{n} q_{n\nu}(x) \left(x - \frac{\nu}{n}\right)^2.$$

Since $(x-\frac{\nu}{n})^2 = x^2 - 2x\frac{\nu}{n} + (\frac{\nu}{n})^2$, the last sum can be separated into the following three parts:

$$(1) \sum_{\nu=0}^{n} \binom{n}{\nu} x^\nu (1-x)^{n-\nu} = 1;$$

$$(2) \sum_{\nu=0}^{n} \binom{n}{\nu} x^\nu (1-x)^{n-\nu} \frac{\nu}{n} = x \sum_{\nu=1}^{n} \binom{n-1}{\nu-1} x^{\nu-1} (1-x)^{(n-1)-(\nu-1)} = x;$$

$$(3) \sum_{\nu=0}^{n} \binom{n}{\nu} x^\nu (1-x)^{n-\nu} \left(\frac{\nu}{n}\right)^2 =$$

$$= \frac{x}{n} \sum_{\nu=1}^{n} (\nu-1) \binom{n-1}{\nu-1} x^{\nu-1} (1-x)^{(n-1)-(\nu-1)} + \frac{x}{n} =$$

$$= \frac{x^2}{n}(n-1) \sum_{\nu=2}^{n} \binom{n-2}{\nu-2} x^{\nu-2} (1-x)^{(n-2)-(\nu-2)} + \frac{x}{n} = x^2\left(1-\frac{1}{n}\right) + \frac{x}{n} =$$

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

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

$$(**) \qquad \sum_{\nu=0}^{n} q_{n\nu}(x) \left(x - \frac{\nu}{n}\right)^2 = x^2 \cdot 1 - 2x \cdot x + x^2 + \frac{x(1-x)}{n} \le \frac{1}{4n}$$

and

$$\sum_{\nu\in N''} \left| f(x) - f\left(\frac{\nu}{n}\right) \right| q_{n\nu}(x) \le \frac{2M}{\delta^2} \frac{1}{4n} < \frac{\varepsilon}{2},$$

provided that we choose $n > \frac{M}{\delta^2 \varepsilon}$. Combining these facts, we obtain the estimate

$$|f(x) - (B_n f)(x)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

for all $x \in [0,1]$, which establishes the uniform convergence of the sequence $(B_n f)$. □

Remark. We can now give an answer to the question raised in 2.1. Every analytic function can be expanded in a power series, while every continuous function can be represented as an expansion in terms of polynomials as follows:

$$f(x) = (B_1 f)(x) + [(B_2 f)(x) - (B_1 f)(x)] + \cdots + [(B_n f)(x) - (B_{n-1} f)(x)] + \cdots.$$

This series converges uniformly, but in general cannot be rearranged into a power series.

2.3 The Korovkin Approach. Examining the proof given in the previous section once again, we note that the estimation of the sums (1) – (3) is the essential part of the proof of the convergence of the sum (*). Indeed, it is clear that the convergence essentially depends on being able to establish the uniform convergence of the sums (1), (2) and (3) for the functions $e_1(x) := 1$, $e_2(x) := x$, and $e_3(x) := x^2$. This suggests that the convergence of the sequence of Bernstein polynomials to an arbitrary continuous function is already determined by the way in which the Bernstein polynomials behave for the three elements $e_1, e_2, e_3 \in C[a,b]$.

This conjecture turns out to be correct. In 1953, P. P. Korovkin established a general approximation theorem which contains this assertion. His proof depends in an essential way on the concept of a

Monotone Linear Operator. Let $f, g \in C(I)$ be given functions such that $f \le g$, where this notation means that $f(x) \le g(x)$ for all $x \in I$. Then a linear operator $L : C(I) \to C(I)$ is called *monotone* provided that $Lf \le Lg$. This property is equivalent to the property of *positivity*, i. e., $f \ge 0$ implies $Lf \ge 0$. In 2.4 we shall exploit the fact that the Bernstein operators defined there are positive.

Korovkin investigated sequences $(L_n)_{n\in\mathbb{N}}$ of positive linear operators $L_n : C(I) \to C(I)$ for $I := [0,1]$ which map continuous functions $f \in C(I)$ to polynomials, as well as similar operators which map a continuous and 2π-periodic function $f \in C_{2\pi}(I)$ with $I := [-\pi,\pi]$ to trigonometric polynomials of maximal degree n. He showed that for every $f \in C([0,1])$, the sequence $(L_n f)$ converges uniformly to f provided that uniform convergence holds for the three functions $e_1(x) := 1$, $e_2(x) := x$, $e_3(x) := x^2$, and that the same holds for every $f \in C_{2\pi}([-\pi,\pi])$, provided that it holds for each of the three functions $e_1(x) := 1$, $e_2(x) := \sin(x)$, $e_3(x) := \cos(x)$.

Korovkin's proofs of these two facts are similar, but not exactly the same. Here we present a unified and generalized version of the proof due to E. Schäfer [1989]. This proof can, in fact, be further simplified if one is interested only in the two special cases of continuous functions mentioned above.

Consider the linear space $(C(I), \|\cdot\|_\infty)$. Let $Q := \{f_1, \ldots, f_k\}$, $Q \subset C(I)$, and let $e_1 \in \text{span}(Q)$. We call the set Q a *test set* provided that there exists a function $p \in C(I \times I)$ such that $p(t,x) := \sum_{\kappa=1}^{k} a_\kappa(t) f_\kappa(x)$ with $a_\kappa \in C(I)$ for $1 \leq \kappa \leq k$, $p(t,x) \geq 0$ for all $(t,x) \in I \times I$, and $p(t,t) = 0$ for all $t \in I$.

We denote by $Z(g) := \{(t,x) \in I \times I \mid g(t,x) = 0\}$ the zero set of a function $g \in C(I \times I)$, and write $d_f(t,x) := f(x) - f(t)$ for the "difference function" associated with a given $f \in C(I)$. We now have the following

Theorem. *Let $(L_n)_{n \in \mathbb{N}}$, $L_n : C(I) \to C(I)$, be a sequence of positive linear operators, and let Q be a test set with associated function p. Suppose that for every element $f \in Q$, $\lim_{n\to\infty} \|L_n f - f\|_\infty = 0$. Then it follows that $\lim_{n\to\infty} \|L_n f - f\|_\infty = 0$ for every element $f \in C(I)$ which satisfies the condition $Z(p) \subset Z(d_f)$.*

Proof. In part (a) of the proof we show that for $\lim_{n\to\infty} \|f - L_n f\|_\infty = 0$, it suffices to establish that $\lim_{n\to\infty} \max_{t \in I} |(L_n d_f(t,\cdot))(t)| = 0$. The proof that $\lim_{n\to\infty} \max_{t\in I} |(L_n d_f(t,\cdot))(t)| = 0$ for all elements $f \in C(I)$ such that $Z(p) \subset Z(d_f)$ is presented in part (b).

(a) $d_f(t,\cdot) = f - f(t)e_1$ satisfies $f - L_n f = f - f(t) L_n e_1 - L_n d_f(t,\cdot)$. From this it follows that

$$|f(t) - (L_n f)(t)| \leq \|f\|_\infty \|e_1 - L_n e_1\|_\infty + \max_{t \in I} |(L_n d_f(t,\cdot))(t)|,$$

uniformly for all $t \in I$. Since $e_1 \in \text{span}(Q)$, we get $\lim_{n\to\infty} \|e_1 - L_n e_1\|_\infty = 0$, and thus $\lim_{n\to\infty} \max_{t \in I} |(L_n d_f(t,\cdot))(t)| = 0$ gives $\lim_{n\to\infty} \|f - L_n f\|_\infty = 0$.

(b) The difference function depends continuously on the variables x and t. Hence, for every $\varepsilon > 0$, there exists an open neighborhood Ω of $Z(d_f)$, where $|d_f(t,x)| < \varepsilon$ for all $(t,x) \in \Omega$. The diagonal set D defined by $D := \{(t,x) \in I \times I \mid t = x\}$ is thus surely a subset of $Z(d_f)$. By the assumption that $Z(p) \subset Z(d_f)$, it follows that $p(t,x) > 0$ in the complement $\Omega' := I \times I \setminus \Omega$.

Ω' is closed and hence compact, which assures that the minimum $0 < m := \min_{(t,x) \in \Omega'} p(t,x)$ exists. Thus

$$|d_f(t,x)| \leq \|d_f\|_\infty \frac{p(t,x)}{m} \quad \text{for } (t,x) \in \Omega',$$

and we have

$$|d_f(t,x)| \leq \frac{\|d_f\|_\infty}{m} p(t,x) + \varepsilon \quad \text{for } (t,x) \in I \times I.$$

Applying the positive operator L_n with respect to x for fixed t, it follows that

$$|(L_n d_f(t,\cdot))(t)| \leq \frac{\|d_f\|_\infty}{m} (L_n p(t,\cdot))(t) + \varepsilon (L_n e_1)(t) \leq$$

$$\leq \frac{\|d_f\|_\infty}{m} \max_{t \in I} (L_n p(t,\cdot))(t) + \varepsilon \|L_n e_1\|_\infty.$$

Since $p(t,t) = 0$ for all $t \in I$, we can write

$$(L_n p(t,\cdot))(t) = \sum_{\kappa=1}^{k} a_\kappa(t) [(L_n f_\kappa)(t) - f_\kappa(t)].$$

The convergence of the sequence (L_n) on span(Q) thus implies that

$$\lim_{n\to\infty} \max_{t \in I} (L_n p(t,\cdot))(t) = 0.$$

Since $\|L_n e_1\|_\infty$ is uniformly bounded in n, we finally arrive at the assertion

$$\lim_{n\to\infty} \max_{t \in I} |(L_n d_f(t,\cdot))(t)| = 0. \qquad \square$$

2.4 Applications of Theorem 2.3. In this section we apply Theorem 2.3 to obtain the classical approximation theorems of Weierstrass. Although we have already established the approximation theorem for continuous functions in 2.2, here we reprove it by showing how it follows from Theorem 2.3.

In order to apply Theorem 2.3, we must find an appropriate test set and a sequence of positive linear operators which converges on this test set. We begin by establishing the Approximation Theorem 2.2 with the help of

Bernstein-Operators. The Bernstein polynomial $B_n f$ introduced in the proof of Theorem 2.2 defines a mapping of the space of continuous functions into the linear subspace of polynomials P_n. Considering B_n as an operator $B_n : C(I) \to C(I)$, it is easy to see that it is linear und monotone. First, from the definition

$$(B_n f)(x) = \sum_{\nu=0}^{n} f\left(\frac{\nu}{n}\right) \binom{n}{\nu} x^\nu (1-x)^{n-\nu},$$

it follows immediately that $B_n(\alpha f + \beta g) = \alpha B_n f + \beta B_n g$, and thus that B_n is linear. Since $f \geq 0$ implies $B_n f \geq 0$, it follows that B_n is positive, or equivalently, monotone.

A natural choice for a set of test functions Q is the set $\{f_1, f_2, f_3\}$ with $f_1(x) := e_1(x) = 1$, $f_2(x) := e_2(x) = x$, $f_3(x) := e_3(x) = x^2$, with corresponding p defined by $p(x,t) := (t-x)^2 = t^2 - 2tx + x^2$. The condition $Z(p) \subset Z(d_f)$ holds for every $f \in C(I)$ since $p(x,t) = 0$ if and only if $x = t$.

Our choice of the elements e_1, e_2, e_3 for the set Q is motivated by the fact that in the proof of Theorem 2.2, we established the fact that $\lim_{n\to\infty} \|B_n e_\kappa - e_\kappa\|_\infty = 0$ for $\kappa = 1,2,3$. This together with Theorem 2.3 implies that $\lim_{n\to\infty} \|B_n f - f\|_\infty = 0$ for all elements $f \in C(I)$. We have obtained the Approximation Theorem 2.2 as an application of Theorem 2.3.

Periodic Functions. A natural way to approximate a 2π-periodic function as a linear combination of given elements is to use the Fourier series