ORLICZ CENTENARY VOLUME
BANACH CENTER PUBLICATIONS, VOLUME 64
INSTITUTE OF MATHEMATICS
POLISH ACADEMY OF SCIENCES
WARSZAWA 2004

DENSITY METHODS AND RESULTS IN APPROXIMATION THEORY

ALLAN PINKUS

Department of Mathematics, Technion, Haifa 32000, Israel
E-mail: pinkus@tx.technion.ac.il

Abstract. Approximation theory and functional analysis share many common problems and points of contact. One of the areas of mutual interest is that of density results. In this paper we briefly survey various methods and results in this area starting from work of Weierstrass and Riesz, and extending to more recent times.

1. Introduction. Approximation theory is that area of analysis which, at its core, is concerned with the ability to approximate functions by simpler and more easily calculated functions. It is an area which, like many other fields of analysis, has its primary roots in the mathematics of the 19th century.

At the beginning of the 19th century functions were essentially viewed via concrete formulae, and sometimes as series or solutions of equations. However largely as a consequence of the claims of Fourier and the results of Dirichlet, the modern concept of a function distinguished by its requisite properties was introduced and accepted. Once a function, and more specifically a continuous function, is defined implicitly rather than explicitly, then it was both inevitable and unavoidable that we would eventually witness the birth of both approximation theory and functional analysis.

It is in the theory of Fourier series that we find some of the first results of approximation theory. These include conditions on a function that ensure the pointwise or uniform convergence (of the partial sums) of its Fourier series, as well as the omnipresent L^2-convergence. Similar results were also developed for other orthogonal series, and for power series (analytic functions). However these results are of a rather particular form. They are concerned with conditions for when certain formulae hold. In the classical theory of Fourier series one does not ask if trigonometric polynomials can be used to approximate, or even if the information provided by the Fourier coefficients is sufficient to provide an approximation. Rather one wants to know if and how the partial sums of the Fourier series converge to the function in question.

2000 *Mathematics Subject Classification*: 41-02, 41A45.
The paper is in final form and no version of it will be published elsewhere.

The first question we ask in approximation theory concerns the *possibility of approximation*. Is the given family of functions from which we plan to approximate dense (or fundamental) in the set of functions we wish to approximate? That is, can we approximate any function in our set, arbitrarily well, using finite linear combinations of functions from our given family?

The first significant density results were those of Weierstrass who proved in 1885 (when he was 70 years old) the density of algebraic polynomials in the class of continuous real-valued functions on a finite interval, and the density of trigonometric polynomials in the class of 2π-periodic continuous real-valued functions. These theorems were, in a sense, a counterbalance to Weierstrass' famous example of 1861 on the existence of a continuous nowhere differentiable function. The existence of such functions accentuated the need for analytic rigour in mathematics, for a further understanding of the nature of the set of continuous functions, and substantially influenced the development of analysis. If this example represented for some a 'lamentable plague' (as Hermite wrote to Stieltjes on May 20, 1893, see Baillard and Bourget [1905]), then the approximation theorems were a panacea. While on the one hand the set of continuous functions contains deficient functions, on the other hand every continuous function can be approximated arbitrarily well by the ultimate in smooth functions, the polynomials.

The Weierstrass approximation theorems spawned numerous generalizations which were applied to other families of functions. They also led to the development of two general methods for determining density. These are the Stone-Weierstrass theorem generalizing the Weierstrass theorem to subalgebras of $C(X)$, X a compact space, and the Bohman-Korovkin theorem characterizing sequences of positive linear operators that approximate the identity operator, based on easily checked, simple, criteria.

A different and more modern approach to density theorems is via "soft analysis". This functional analytic approach actually dates back almost 100 years. A linear subspace M of a normed linear space E is dense in E if and only if the only continuous linear functional that vanishes on M is the identically zero functional. For the space $C[a,b]$ this result can already be found in the work of F. Riesz from 1910 and 1911 as one of the first applications of his "representation theorem" characterizing the set of all continuous linear functionals on $C[a,b]$.

Density theorems can be found almost everywhere in analysis, and not only in analysis. (For a density result equivalent to the Riemann Hypothesis see Conrey [2003, p. 345].) In this article we survey a few of the main results regarding density of linear subspaces in spaces of continuous real-valued functions endowed with the uniform norm. We only present a limited sampling of the many, many density results to be found in approximation theory and in other areas. We also restrict ourselves to real-valued functions and uniform approximation. A monograph many times the length of this paper would not suffice to include all results. In addition, we do not prove most of the results we quote. We hope, nonetheless, that you the reader will find something here of interest. A longer version of this paper is being prepared which will hopefully appear somewhere.

2. The Weierstrass approximation theorems. We first fix some notation. $C[a,b]$ will denote the class of continuous real-valued functions on the closed interval $[a,b]$, and

$\widetilde{C}[0, 2\pi]$ the class of functions in $C[0, 2\pi]$ satisfying $f(0) = f(2\pi)$. We denote by Π_n the space of algebraic polynomials of degree at most n, i.e.,

$$\Pi_n = \text{span}\{1, x, \ldots, x^n\},$$

and by T_n the space of trigonometric polynomials of degree at most n, i.e.,

$$T_n = \text{span}\{1, \sin x, \cos x, \ldots, \sin nx, \cos nx\}.$$

The paper stating and proving what we call the Weierstrass approximation theorems is Weierstrass [1885]. It seems that the importance of the paper was immediately appreciated, as the paper appeared in translation (in French) one year later in Weierstrass [1886]. Weierstrass was interested in complex function theory and in the ability to represent functions by power series. The results he obtained in this 1885 paper should be viewed from that perspective. The title of the paper emphasizes this viewpoint. The paper is titled *On the possibility of giving an analytic representation to an arbitrary function of a real variable.* It is interesting to read this paper, as Weierstrass' perception of these approximation theorems was certainly different from ours. Weierstrass' view of analytic functions was of functions that could be represented by power series. The approximation theorem, for him, was an extension of this result to continuous functions. Every continuous function could be represented by a polynomial series that converged both absolutely and uniformly. Similarly, 'nice' functions in $\widetilde{C}[0, 2\pi]$ enjoy the property that their Fourier series converges absolutely and uniformly. What Weierstrass also proved was that every function in $\widetilde{C}[0, 2\pi]$ could be represented by a trigonometric polynomial series that converged both absolutely and uniformly.

The paper Weierstrass [1885] was reprinted in Weierstrass' Mathematische Werke (collected works) with some notable additions. While this reprint appeared in 1903, there is reason to assume that Weierstrass himself edited this paper. One of these additions was a short "introduction". We quote it (verbatim in meaning if not in fact).

The main result of this paper, restricted to the one variable case, can be summarized as follows:

Let $f \in C(\mathbb{R})$. Then there exists a sequence $f_1, f_2, \ldots$ of entire functions for which

$$f(x) = \sum_{i=1}^{\infty} f_i(x)$$

for each $x \in \mathbb{R}$. In addition the convergence of the above sum is uniform on every finite interval.

Note that there is no mention of the fact that the f_i may be assumed to be polynomials.

We state the Weierstrass theorems, not as given in his paper, but as they are currently stated and understood.

Weierstrass Theorem 2.1. *For every finite $a < b$ algebraic polynomials are dense in $C[a, b]$. That is, given an f in $C[a, b]$ and an arbitrary $\varepsilon > 0$ there exists an algebraic polynomial p such that*

$$|f(x) - p(x)| < \varepsilon$$

for all x in $[a, b]$.

Weierstrass Theorem 2.2. *Trigonometric polynomials are dense in* $\widetilde{C}[0, 2\pi]$. *That is, given an* f *in* $\widetilde{C}[0, 2\pi]$ *and an arbitrary* $\varepsilon > 0$ *there exists a trigonometric polynomial* t *such that*

$$|f(x) - t(x)| < \varepsilon$$

for all x *in* $[0, 2\pi]$.

These are the first significant density theorems in analysis. They are generally paired since in fact they are equivalent. That is, each of these theorems follows from the other.

Over the next twenty-five or so years numerous alternative proofs were given to one or the other of these results by a roster of some of the best analysts of the period. There are the proofs by Weierstrass, Picard, Fejér, Landau and de la Valleé Poussin that used singular integrals, proofs based on the idea of approximating one particular function by Runge (Phragmén), Lebesgue, Mittag-Leffler, and Lerch, proofs based on Fourier series by Lerch, Volterra and Fejér, and the wonderful proof of Bernstein. Details concerning all these proofs can be found, for example, in Pinkus [2000]. We explain, without detail, three of these proofs.

One of the more elegant and cited proofs of Weierstrass' theorem is due to Lebesgue [1898]. This was Lebesgue's first published paper. He was, at the time of publication, a 23 year old student at the École Normale Supérieure. The idea of his proof is simple and useful. Lebesgue noted that each f in $C[a, b]$ can be easily approximated by a continuous, piecewise linear curve (polygonal line). Each such polygonal line is a linear combination of translates of $|x|$. As algebraic polynomials (of any fixed degree) are translation invariant, it thus suffices to prove that one can uniformly approximate $|x|$ arbitrarily well by polynomials on any interval containing the origin. Lebesgue then does exactly that. Explicitly

$$|x| = 1 - \sum_{n=1}^{\infty} a_n (1 - x^2)^n$$

where $a_1 = 1/2$, and

$$a_n = \frac{(2n-3)!}{2^{2n-2} n! (n-1)!}, \qquad n = 2, 3, \ldots$$

This "power series" converges absolutely and uniformly to $|x|$ in a neighborhood of 0 (for all $|x| \leq 1$). Truncating this series we obtain a series of polynomial approximants to $|x|$.

When Fejér was 20 years old, he published Fejér [1900] that formed the basis for his doctoral thesis. Fejér proved more than the Weierstrass approximation theorem (for trigonometric polynomials). He proved that for any f in $\widetilde{C}[0, 2\pi]$ it is possible to uniformly approximate f based solely on the knowledge of its Fourier coefficients. He did not obtain this approximation by taking the partial sums of the Fourier series. It is well-known that these do not necessarily converge. Rather he obtained it by taking the Cesàro sums of the partial sums of the Fourier series. In other words, assume that we are given the Fourier series of f

$$f(x) \sim \sum_{k=-\infty}^{\infty} c_k e^{ikx},$$

where

$$c_k = \frac{1}{2\pi}\int_0^{2\pi} f(x)e^{-ikx}dx$$

for every $k \in \mathbb{Z}$. Define the nth partial sums of the Fourier series by

$$s_n(x) := \sum_{k=-n}^{n} c_k e^{ikx}.$$

and set

$$\sigma_n(f;x) = \frac{s_0(x) + \cdots + s_n(x)}{n+1}.$$

The σ_n is termed the nth Fejér operator. Note that $\sigma_n(f;\cdot)$ belongs to T_n for each n. What Fejér proved was that, for each f in $\widetilde{C}[0,2\pi]$, $\sigma_n(f;\cdot)$ tends uniformly to f as $n \to \infty$. This proof was also the first to construct a sequence of *linear* operators that could be used in the approximation process.

Simpler linear operators that approximate were introduced by Bernstein [1912/13]. These are the Bernstein polynomials. For f in $C[0,1]$ they are defined by

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

Bernstein proved, by probabilistic methods, that the $B_n(f\,;\cdot)$ converge uniformly to f as $n \to \infty$. A proof of this convergence is to be found in Example 4.2.

3. The functional analytic approach. The Riesz representation theorem characterizing the space of continuous linear functionals on $C[a,b]$ is contained in the 1909 paper of F. Riesz [1909]. The following year, in a rarely referenced paper, Riesz [1910] also announced the following (stated in more modern terminology).

THEOREM 3.1. *Let $u_k \in C[a,b]$, $k \in K$, where K is an index set. A necessary and sufficient condition for the existence of a continuous linear functional F on $C[a,b]$ satisfying*

$$F(u_k) = c_k, \qquad k \in K,$$

with $\|F\| \le L$ is that

$$\Big|\sum_{k\in K'} a_k c_k\Big| \le L\Big\|\sum_{k\in K'} a_k u_k\Big\|_\infty$$

holds for every finite subset K' of K, and all real a_k.

In this same paper Riesz also states the parallel result for $L^p[a,b]$, $1 < p < \infty$. Questions concerned with existence and uniqueness in moment problems were of major importance in the development of functional analysis. The full details of the 1910 announcement appear in Riesz [1911]. In these papers is also to be found the following result (again we switch to more modern terminology).

THEOREM 3.2. *Let M be a linear subspace of $C[a,b]$. Then $f \in C[a,b]$ is in the closure of M, i.e., f can be uniformly approximated by elements of M, if and only if every continuous linear functional on $C[a,b]$ that vanishes on M also vanishes on f.*

Riesz quotes E. Schmidt as the author of the *very interesting problem* whose solution is the above Theorem 3.2. As he writes, the question asked is: *Being given a countable system of functions $\phi_n \in C[a,b]$, $n = 1, 2, ...$, how can one know if one can approximate arbitrarily and uniformly every $f \in C[a,b]$ by the ϕ_n and their linear combinations?* (Riesz [1911, p. 51]). Schmidt, in his thesis in Schmidt [1905], had given both a necessary and a sufficient condition for the above to hold. Both were orthogonality type conditions. However neither was the correct condition. The concept of a linear functional vanishing on a set of functions is very orthogonal-like. Lerch's theorem (Lerch [1892], see also the more accessible Lerch [1903]), states that if $h \in C[0,1]$ and

$$\int_0^1 x^n h(x)\,dx = 0, \qquad n = 0, 1, \ldots,$$

then $h = 0$. This theorem was well-known and frequently quoted. So it was not unreasonable to look for conditions of the form given in Theorem 3.2. (Lerch's theorem is, in fact, a simple consequence of Weierstrass' theorem.)

As Riesz states, one consequence of the above Theorem 3.2 is that M is dense in $C[a,b]$ if and only if no nontrivial continuous linear functional vanishes on M. The proof of Theorem 3.2, contained in Riesz [1911], is just an application of Theorem 3.1.

Proof. We start with the simple direction. Assume f is in the closure of M. If F is a continuous linear functional that vanishes on M, then $F(f) = F(f-g)$ for every $g \in M$. Given $\varepsilon > 0$, there exists a $g^* \in M$ for which $\|f - g^*\|_\infty < \varepsilon$. Thus

$$|F(f)| = |F(f - g^*)| \le \|F\|\|f - g^*\|_\infty < \varepsilon\|F\|.$$

As this is valid for every $\varepsilon > 0$ we have $F(f) = 0$.

Now assume that f is not in the closure of M. Thus $\|f - g\|_\infty \ge d > 0$ for every $g \in M$. From this inequality and Theorem 3.1 there necessarily exists a continuous linear functional F on $C[a,b]$ satisfying $F(g) = 0$, for all $g \in M$, $F(f) = 1$, and $\|F\| \le L$ for any $L \ge 1/d$. This holds since we have

$$|a| \le Ld|a| \le L\|af - g\|_\infty,$$

for all $g \in M$ and all a. ■

Shortly thereafter Helly [1912] applied these results to a question concerning the range of an integral equation. He proved the following two theorems.

THEOREM 3.3. *Let $K \in C([a,b] \times [a,b])$ and $f \in C[a,b]$. Then a necessary and sufficient condition for the existence of a measure of bounded total variation ν satisfying*

$$f(x) = \int_a^b K(x,y)d\nu(y),$$

is the existence of a constant L for which

$$\Big|\sum_{k=1}^n a_k f(x_k)\Big| \le L\Big|\sum_{k=1}^n a_k K(x_k, y)\Big|$$

for all points $x_1, \ldots, x_n$ in $[a,b]$, all real values $a_1, \ldots, a_n$, all $y \in [a,b]$, and all n.

THEOREM 3.4. *Let $K \in C([a,b] \times [a,b])$. Then a necessary and sufficient condition for an $f \in C[a,b]$ to be uniformly approximated by functions of the form*

$$\int_a^b K(x,y)\phi(y)d\,y,$$

where the ϕ are piecewise continuous functions, is that for every measure μ of bounded total variation satisfying

$$\int_a^b K(x,y)d\,\mu(x) = 0$$

we also have

$$\int_a^b f(x)d\,\mu(y) = 0.$$

In 1911 the concept of a normed linear space did not exist, and the Hahn-Banach theorem had yet to be discovered (although the Helly [1912] paper contains results that compare). Banach's proof of the Hahn-Banach theorem appears in Banach [1929] (Hahn's appears in Hahn [1927]). Both the Hahn and Banach papers contain a general form of Theorem 3.1, namely the Hahn-Banach theorem. Both also essentially contain the statement that a linear subspace is dense in a normed linear space if and only if no nontrivial continuous linear functional vanishes on the subspace. Banach, in his book Banach [1932, p. 57], prefaces these next two theorems with the statement: *We are now going to establish some theorems that play in the theory of normed spaces the analogous role to that which the Weierstrass theorem on the approximation of continuous functions by polynomials plays in the theory of functions of a real variable.*

THEOREM 3.5. *Let M be a linear subspace of a real normed linear space E. Assume $f \in E$ and*

$$\|f - g\| \geq d > 0$$

for all $g \in M$. Then there exists a continuous linear functional F on E such that $F(g) = 0$ for all $g \in M$, $F(f) = 1$, and $\|F\| \leq 1/d$.

The result of Theorem 3.5 replaces Theorem 3.1 in the proof of Theorem 3.2 to give us the well-known

THEOREM 3.6. *Let M be a linear subspace of a real normed linear space E. Then $f \in E$ is in the closure of M if and only if every continuous linear functional on E that vanishes on M also vanishes on f.*

In none of these works of Hahn and Banach are the above-mentioned 1910 or 1911 papers of Riesz referenced. These Riesz papers seem to have been essentially forgotten. In fact the general method of proof of density based on this approach is to be found in the literature only after the appearance of the book of Banach and the blooming of functional analysis. The name of Riesz is often mentioned in connection with this method, but only because of the Riesz representation theorem and similar duality results bearing his name.

Today we also recognize the Hahn-Banach theorem as a separation theorem, and as such we also have the following two results.

THEOREM 3.7. *Let E be a real normed linear space, ϕ_n elements of E, $n \in I$, and $f \in E$. Then f may be approximated by finite convex linear combinations of the ϕ_n, $n \in I$, if and only if*

$$\sup\{F(\phi_n) : n \in I\} \geq F(f)$$

for every continuous linear functional F on E.

THEOREM 3.8. *Let E be a real normed linear space, ϕ_n elements of E, $n \in I$, and $f \in E$. Then f may be approximated by finite positive linear combinations of the ϕ_n, $n \in I$, if and only if for every continuous linear functional F on E satisfying $F(\phi_n) \geq 0$ for every $n \in I$ we have $F(f) \geq 0$.*

Theorem 3.8 follows from Theorem 3.7 by considering the convex cone generated by the ϕ_n.

There are numerous generalizations of these results. The book of Nachbin [1967] where these results may be found is one of the few to concentrate on density theorems. Much of the book is taken up with the Stone-Weierstrass theorem. However there are also other results such as the above Theorems 3.7 and 3.8.

4. Other density methods. The Weierstrass theorems had a significant influence on the development of density results, even though the theorems themselves simply prove the density of algebraic and trigonometric polynomials in the appropriate spaces. Various proofs of the Weierstrass theorems, for example, provided insights that led to the development of two general methods for determining density. We briefly discuss these methods in this section.

The first of these methods is given by the Stone-Weierstrass theorem. This theorem was originally proven in Stone [1937]. Stone subsequently reworked his proof in Stone [1948]. It represents, as stated by Buck [1962, p. 4], *one of the first and most striking examples of the success of the algebraic approach to analysis.* There have since been numerous modifications and extensions. See, for example, Nachbin [1967], Prolla [1993] and references therein.

We recall that an *algebra* is a linear space on which multiplication between elements has been suitably defined satisfying the usual commutative and associative type postulates. Algebraic and trigonometric polynomials in any finite number of variables are algebras. A set in $C(X)$ *separates points* if for distinct points $x, y \in X$ there exists a g in the set for which $g(x) \neq g(y)$.

STONE-WEIERSTRASS THEOREM 4.1. *Let X be a compact set and let $C(X)$ denote the space of continuous real-valued functions defined on X. Assume A is a subalgebra of $C(X)$. Then A is dense in $C(X)$ in the uniform norm if and only if A separates points and for each $x \in X$ there exists an $f \in A$ satisfying $f(x) \neq 0$.*

EXAMPLE 4.1. As we mentioned prior to the statement of the Stone-Weierstrass theorem, algebraic polynomials in any finite number of variables form an algebra. They also separate points and contain the constant function. Thus algebraic polynomials in m variables are dense in $C(X)$ where X is any compact set in $\mathbb{R}^m$. This fact first appeared in print (at least for squares) in Picard [1891] which also contains an alternative proof of Weierstrass' theorems. The paper Weierstrass [1885] as "reprinted" in Weierstrass' Mathematische Werke in 1903 contains an additional 10 pages of material including a proof of this multivariable analogue of his theorem.

Another method that can be used to prove density is based on what is called the Korovkin theorem or the Bohman-Korovkin theorem. A primitive form of this theorem was proved by Bohman in Bohman [1952]. His proof, and the main idea in his approach, was a generalization of Bernstein's proof of the Weierstrass theorem. Korovkin one year later in Korovkin [1953] proved the same theorem for integral type operators. Korovkin's original proof is in fact based on positive singular integrals and there are very obvious links to Lebesgue's work on singular operators that, in turn, was motivated by various of the proofs of the Weierstrass theorems. Korovkin was probably unaware of Bohman's result. Korovkin subsequently much extended his theory, major portions of which can be found in his book Korovkin [1960]. The theorem as presented here is taken from Korovkin's book.

A linear operator L is *positive* (monotone) if $f \geq 0$ implies $L(f) \geq 0$.

BOHMAN–KOROVKIN THEOREM 4.2. *Let* (L_n) *be a sequence of positive linear operators mapping* $C[a,b]$ *into itself. Assume that*

$$\lim_{n\to\infty} L_n(x^i) = x^i, \qquad i = 0, 1, 2,$$

and the convergence is uniform on $[a,b]$. *Then*

$$\lim_{n\to\infty} (L_n f)(x) = f(x)$$

uniformly on $[a,b]$ *for every* $f \in C[a,b]$.

A similar result holds in the periodic case $\widetilde{C}[0,2\pi]$, where "test functions" are $1, \sin x$, and $\cos x$. Numerous generalizations may be found in the book of Altomare and Campiti [1994].

How can the Bohman-Korovkin theorem be applied to obtain density results? It can, in theory, be applied easily. If the $U_n = \mathrm{span}\{u_1, \ldots, u_n\}$, $n = 1, 2, \ldots$, are a nested sequence of finite-dimensional subspaces of $C[a,b]$, and L_n is a positive linear operator mapping $C[a,b]$ into U_n that satisfies the conditions of the above theorem, then the $(u_k)_{k=1}^{\infty}$ span a dense subset of $C[a,b]$. In practice it is all too rarely applied in this manner. The importance of the Korovkin theory is primarily in that it presents conditions implying convergence, and also in that it provides calculable error bounds on the rate of approximation.

EXAMPLE 4.2. One immediate application of the Bohman-Korovkin theorem is a proof of the convergence of the Bernstein polynomials $B_n(f)$ to f for each f in $C[0,1]$. We may consider the (B_n) as a sequence of positive linear operators mapping $C[0,1]$ into Π_n, the space of algebraic polynomials of degree at most n. It is readily verified that $B_n(1\,;x) = 1$, $B_n(x\,;x) = x$ and $B_n(x^2\,;x) = x^2 + x(1-x)/n$ for all $n \geq 2$. Thus by the Bohman-Korovkin theorem $B_n(f)$ converges uniformly to f on $[0,1]$.

5. Some density results

EXAMPLE 5.1. *Müntz's Theorem.* Possibly the first generalization of consequence of the Weierstrass theorems, and certainly one of the best known, is the Müntz theorem or the Müntz-Szász theorem.

It was Bernstein who in a paper in the proceedings of the 1912 International Congress of Mathematicians held at Cambridge, Bernstein [1913], and in his 1912 prize-winning essay, Bernstein [1912], asked for exact conditions on an increasing sequence of positive exponents λ_n, so that the sequence (x^{λ_n}) is fundamental in the space $C[0,1]$. Bernstein himself had obtained some partial results. In the paper in the ICM proceedings Bernstein wrote the following: *It will be interesting to know if the condition that the series $\sum 1/\lambda_n$ diverges is not necessary and sufficient for the sequence of powers (x^{λ_n}) to be fundamental; it is not certain, however, that a condition of this nature should necessarily exist.*

It was just two years later that Müntz [1914] was able to provide a solution confirming Bernstein's qualified guess. What Müntz proved is the following.

MÜNTZ'S THEOREM 5.1. *The sequence*

$$x^{\lambda_0}, x^{\lambda_1}, x^{\lambda_2}, \ldots$$

where $0 \le \lambda_0 < \lambda_1 < \lambda_2 < \cdots \to \infty$ *is fundamental in* $C[0,1]$ *if and only if* $\lambda_0 = 0$ *and*

$$\sum_{k=1}^{\infty} \frac{1}{\lambda_k} = \infty.$$

There are numerous proofs and generalizations of the Müntz theorem. It is to be found in many of the classic texts on approximation theory, see e.g. Achieser [1956, p. 43–46], Cheney [1966, p. 193–198], Borwein, Erdélyi [1995, p. 171–205]. (The last reference contains many generalizations of Müntz's theorem and also surveys the literature on this topic.) An alternative method of proof of Müntz's theorem and its numerous generalizations is via the functional analytic approach, and the possible sets of uniqueness for zeros of analytic functions, see e.g. Schwartz [1943], Rudin [1966, p. 304–307], Luxemburg, Korevaar [1971], Feinerman, Newman [1974, Chap. X], and Luxemburg [1976]. For some different approaches see, for example, Rogers [1981], Burckel, Saeki [1983], and the very elegant v. Golitschek [1983].

EXAMPLE 5.2. Combining the functional analytic approach with analytic methods has proven to be a very effective method of proving density results. As a general example, assume g is in $C(\mathbb{R})$ and has an extension as an analytic function on all of $\mathbb{C}$. Let Λ be a subset of $\mathbb{R}$ that contains a finite accumulation point, i.e., there are distinct λ_n in Λ and a finite λ^* such that $\lim_{n\to\infty} \lambda_n = \lambda^*$. Set

$$\mathcal{M}_\Lambda = \operatorname{span}\{g(\lambda x) : \lambda \in \Lambda\}.$$

We wish to determine when $\mathcal{M}_\Lambda$ is dense in $C[a,b]$. The following result holds.

THEOREM 5.2. *Let g, Λ and $\mathcal{M}_\Lambda$ be as above. Set*

$$N_g = \{n : g^{(n)}(0) \neq 0\}.$$

Then $\mathcal{M}_\Lambda$ is dense in $C[a,b]$ if and only if:

i) *for* $[a,b] \subseteq (0,\infty)$ *or* $[a,b] \subseteq (-\infty,0)$

$$\sum_{n\in N_g\setminus\{0\}} \frac{1}{n} = \infty,$$

ii) *if* $a=0$ *or* $b=0$, *then* $0\in N_g$ *and*

$$\sum_{n\in N_g\setminus\{0\}} \frac{1}{n} = \infty,$$

iii) *if* $a<0<b$, *then* $0\in N_g$ *and*

$$\sum_{\substack{n\in N_g\setminus\{0\}\\ n \text{ even}}} \frac{1}{n} = \sum_{\substack{n\in N_g\\ n \text{ odd}}} \frac{1}{n} = \infty.$$

Proof. The conditions in (i), (ii) and (iii) are exactly those conditions that determine when

$$\operatorname{span}\{x^n : n\in N_g\}$$

is dense in $C[a,b]$. This is the content of the Müntz theorem in case (ii), and easily follows from the Müntz theorem in case (iii). In case (i) it follows from the Müntz theorem that the condition therein is sufficient for density. The necessity is also true, but needs an additional argument, see e.g., Schwartz [1943].

From the Hahn-Banach and Riesz representation theorems $\mathcal{M}_\Lambda$ is not dense in $C[a,b]$ if and only if there exists a nontrivial measure μ of bounded total variation on $[a,b]$ satisfying

$$\int_a^b g(\lambda x)\,d\mu(x) = 0$$

for all $\lambda\in\Lambda$. Assume such a measure exists. As g is entire, it follows that

$$h(z) = \int_a^b g(zx)\,d\mu(x)$$

is entire. Furthermore $h(\lambda)=0$ for all $\lambda\in\Lambda$. By assumption Λ contains a finite accumulation point. Thus by the uniqueness theorem for zeros of analytic functions $h=0$. However h being identically zero does not necessarily imply that μ is the zero measure. It only proves that

$$\overline{\mathcal{M}_\Lambda} = \overline{\operatorname{span}}\{g(\lambda x) : \lambda\in\mathbb{R}\}.$$

For example, if g is a polynomial of degree m, then $\mathcal{M}_\Lambda$ is simply the space of polynomials of degree m.

As $g(\lambda x)\in\overline{\mathcal{M}_\Lambda}$ for all $\lambda\in\mathbb{R}$ it follows, differentiating by λ, that

$$x^n g^{(n)}(\lambda x) \in \overline{\mathcal{M}_\Lambda}$$

for each nonnegative integer n and all $\lambda\in\mathbb{R}$. Setting $\lambda=0$ gives us

$$x^n g^{(n)}(0) \in \overline{\mathcal{M}_\Lambda}, \qquad n=0,1,...$$

Thus

$$x^n \in \overline{\mathcal{M}_\Lambda}, \qquad n\in N_g.$$

But $\operatorname{span}\{x^n : n\in N_g\}$ is dense in $C[a,b]$, a contradiction.

On the other hand, assume the conditions in (i), (ii) or (iii) do not hold. Then $\text{span}\{x^n : n \in N_g\}$ is not dense in $C[a,b]$, and there exists a nontrivial measure μ of bounded total variation satisfying

$$\int_a^b x^n d\mu(x) = 0$$

for all $n \in N_g$. Since g is entire

$$g(x) = \sum_{n \in N_g} \frac{g^{(n)}(0)}{n!} x^n$$

and it follows that

$$\int_a^b g(\lambda x)\, d\mu(x) = 0$$

for all $\lambda \in \mathbb{R}$. Thus $\mathcal{M}_\Lambda$ is not dense in $C[a,b]$. ∎

For example, if $g(x) = e^x$ then $N_g = \mathbb{Z}_+$ so that (i), (ii) and (iii) always hold. Thus

$$\text{span}\{e^{\lambda_n x} : \lambda \in \Lambda\}$$

is always dense in $C[a,b]$ assuming Λ is a subset of $\mathbb{R}$ with a finite accumulation point. A change of variable argument implies that under this same condition on Λ the set

$$\text{span}\{x^{\lambda_n} : \lambda \in \Lambda\}$$

is dense in $C[\alpha,\beta]$ for every $0 < \alpha < \beta < \infty$.

Questions related to Müntz type problems concern the fundamentality of the functions $(e^{\lambda_n x})$, where (λ_n) is a sequence of complex numbers. This has been considered in spaces of real and complex-valued functions in $C[a,b]$, $C(\mathbb{R}_+)$, $L^p[a,b]$ and $L^p(\mathbb{R}_+)$, $1 \le p < \infty$. There has been considerable research done in this area, see, for example, Paley, Wiener [1934, Chap. VI], Levinson [1940, Chap. I and II], Schwartz [1943], Levin [1964, Appendix III], Levin [1996, Lecture 18], and the many references therein.

Example 5.3. The analysis literature is replete with results concerning the density of *translates (and dilates)* of a function in various spaces. These might be arbitrary, integer, or sequence translates (or dilates). Many of these results are generalizations, in a sense, of the Müntz and/or Paley-Wiener theorems (see the previous example).

There is a characterization of those $f \in C(\mathbb{R})$ for which

$$\text{span}\{f(\cdot - \alpha) : \alpha \in \mathbb{R}\}$$

is **not** dense in $C(\mathbb{R})$ (in the topology of uniform convergence on compacta). Such functions are called *mean periodic*, see Schwartz [1947].

Some functions in $C(\mathbb{R})$ have a further interesting property.

Proposition 5.3. *Assume $f = \widehat{g}$ (f is the Fourier transform of g) for some nontrivial $g \in L^1(\mathbb{R})$ with the support of g contained in an interval of length at most 2π. Then*

$$\text{span}\{f(\cdot - n) : n \in \mathbb{Z}\}$$

is dense in $C(\mathbb{R})$ (in the topology of uniform convergence on compacta).

Proof. Assume the above set is not dense in $C(\mathbb{R})$. Then there exists a Borel measure μ of bounded total variation and compact support E such that

$$\int_E f(x-n)d\,\mu(x) = 0$$

for all $n \in \mathbb{Z}$. Assume $f = \widehat{g}$, as above, and $\operatorname{supp}\{g\} \subseteq [a, a+2\pi]$. Thus for each $n \in \mathbb{Z}$

$$0 = \int_E f(x-n)d\,\mu(x) = \int_E \widehat{g}(x-n)d\,\mu(x) = \frac{1}{2\pi}\int_E \left(\int_a^{a+2\pi} g(t)e^{-i(x-n)t}dt\right)d\,\mu(x)$$
$$= \int_a^{a+2\pi}\left(\frac{1}{2\pi}\int_E e^{-ixt}d\,\mu(x)\right)e^{int}g(t)\,dt = \int_a^{a+2\pi} e^{int}g(t)\widehat{\mu}(t)\,dt$$

where $\widehat{\mu}$ is the Fourier transform of the measure μ. It is well known that $\widehat{\mu}$ is an entire function.

As all the Fourier coefficients of $g\,\widehat{\mu}$ on $[a, a+2\pi]$ vanish we have that $g\,\widehat{\mu}$ is identically zero thereon. This implies that g must vanish where $\widehat{\mu} \neq 0$. As $\widehat{\mu}$ is entire this implies that $g = 0$, a contradiction. ∎

The above is a simple example within a general theory. The interested reader should consult Atzmon, Olevskii [1996], Nikolski [1999], and references therein. Note that there is no function whose integer translates are dense in $L^2(\mathbb{R})$.

EXAMPLE 5.4. The following result is a special case of a general theorem of Schwartz [1944] (see also Pinkus [1996] and references therein). Here we consider $C(\mathbb{R})$, with the topology of uniform convergence on compacta. We are interested in determining the set of functions in $C(\mathbb{R})$ that are both translation and dilation invariant.

PROPOSITION 5.4. *If $\sigma \in C(\mathbb{R})$, $\sigma \neq 0$, then*

$$C(\mathbb{R}) = \overline{\operatorname{span}}\{\sigma(\alpha \cdot + \beta) : \alpha, \beta \in \mathbb{R}\}$$

if and only if σ is not a polynomial.

Proof. Let

$$\mathcal{M}_\sigma = \operatorname{span}\{\sigma(\alpha \cdot + \beta) : \alpha, \beta \in \mathbb{R}\}.$$

If $\overline{\mathcal{M}_\sigma} \neq C(\mathbb{R})$ then there exists a nontrivial Borel measure μ of bounded total variation and compact support such that

$$\int_{\mathbb{R}} \sigma(\alpha x + \beta)\,d\mu(x) = 0$$

for all $\alpha, \beta \in \mathbb{R}$. Since μ is nontrivial and polynomials are dense in $C(\mathbb{R})$ in the topology of uniform convergence on compact subsets, there must exist a $k \geq 0$ such that

$$\int_{\mathbb{R}} x^k\,d\mu(x) \neq 0.$$

It is relatively simple to show that for each $\phi \in C_0^\infty(\mathbb{R})$, (infinitely differentiable and having compact support) the convolution $(\sigma * \phi)$ is contained in $\overline{\mathcal{M}_\sigma}$. Since both σ and ϕ are in $C(\mathbb{R})$, and ϕ has compact support, this can be proven by taking limits of Riemann sums of the convolution integral. We also consider taking derivatives as a limiting operation in taking divided differences. Since $(\sigma * \phi) \in C^\infty(\mathbb{R})$, and thus it and all its derivatives are uniformly continuous on every compact set, it follows that for each $\alpha, \beta \in \mathbb{R}$

$$\frac{\partial^n}{\partial \alpha^n}(\sigma * \phi)(\alpha x + \beta) = x^n (\sigma * \phi)^{(n)}(\alpha x + \beta) \in \overline{\mathcal{M}_\sigma}.$$

Thus

$$\int_{\mathbb{R}} x^n (\sigma * \phi)^{(n)}(\alpha x + \beta)\, d\mu(x) = 0,$$

for all $\alpha, \beta \in \mathbb{R}$ and $n \in \mathbb{Z}_+$. Setting $\alpha = 0$, we see that

$$(\sigma * \phi)^{(n)}(\beta) \int_{\mathbb{R}} x^n\, d\mu(x) = 0$$

for each choice of $\beta \in \mathbb{R}$, $n \in \mathbb{Z}_+$ and $\phi \in C_0^\infty(\mathbb{R})$. This implies, since $\int_{\mathbb{R}} x^k\, d\mu(x) \neq 0$, that

$$(\sigma * \phi)^{(k)} = 0$$

for all $\phi \in C_0^\infty(\mathbb{R})$. That is, $\sigma^{(k)} = 0$ in the weak sense. However, as is well-known, this implies that $\sigma^{(k)} = 0$ in the strong (usual) sense. That is, σ is a polynomial of degree at most $k - 1$.

The converse direction is simple. If σ is a polynomial of degree m, then $\mathcal{M}_\sigma$ is exactly the space of polynomials of degree m, and is therefore not dense in $C(\mathbb{R})$. ∎

EXAMPLE 5.5. Let $\langle \cdot, \cdot \rangle$ denote the usual inner (scalar) product on $\mathbb{R}^n$. We prove the following result.

PROPOSITION 5.5. *For each $\sigma \in C(\mathbb{R})$*

$$\operatorname{span}\{\sigma(\langle \mathbf{a}, \cdot \rangle + b) \,:\, \mathbf{a} \in \mathbb{R}^n, b \in \mathbb{R}\}$$

is dense in $C(\mathbb{R}^n)$ (uniform convergence on compacta) if and only if σ is not a polynomial.

Proof. If σ is a polynomial of degree m, then each $\sigma(\langle \mathbf{a}, \cdot \rangle + b)$ is contained in the space of polynomials of total degree at most m on $\mathbb{R}^n$, and is certainly not dense in $C(\mathbb{R}^n)$.

Assume σ is not a polynomial. Choose an f in $C(\mathbb{R}^n)$, X any compact subset of $\mathbb{R}^n$, and $\varepsilon > 0$. From an elementary application of the Stone-Weierstrass theorem we have the existence of $g_k \in C(\mathbb{R})$ and $\mathbf{a}^k \in \mathbb{R}^n$, $k = 1, \ldots, m$, such that

$$\left| f(\mathbf{x}) - \sum_{k=1}^{m} g_k(\langle \mathbf{a}^k, \mathbf{x} \rangle) \right| < \varepsilon$$

for all $\mathbf{x} \in X$. Let $[c, d]$ be a finite interval of $\mathbb{R}$ containing all values $\langle \mathbf{a}^k, \mathbf{x} \rangle$ for $\mathbf{x} \in X$ and $k = 1, \ldots, m$, i.e.,

$$\bigcup_{k=1}^{m} \{\langle \mathbf{a}^k, \mathbf{x} \rangle \,:\, \mathbf{x} \in X\} \subseteq [c, d].$$

From Proposition 5.4 we have the existence of $c_{ik}, \alpha_{ik}, \beta_{ik} \in \mathbb{R}$, $i = 1, \ldots, n_k$, $k = 1, \ldots, m$ for which

$$\left| g_k(t) - \sum_{i=1}^{n_k} c_{ik} \sigma(\alpha_{ik} t + \beta_{ik}) \right| < \frac{\varepsilon}{m}$$

for all $t \in [c,d]$ and $k = 1,\ldots,m$. Thus for all $\mathbf{x} \in X$

$$\left| f(\mathbf{x}) - \sum_{k=1}^{m} \sum_{i=1}^{n_k} c_{ik}\sigma(\alpha_{ik}\langle \mathbf{a}^k, \mathbf{x}\rangle + \beta_{ik}) \right| < 2\varepsilon,$$

which proves the density. ∎

Proposition 5.5 is a basic result in one of the models of neural network theory, see Leshno, Lin, Pinkus, Schocken [1993] and Pinkus [1999].

Example 5.6. Here are two examples where we consider the density of positive cones. That is, we present some applications of Theorem 3.8.

Let Π denote the space of all algebraic polynomials and Π_+ the convex cone of all algebraic polynomials with nonnegative coefficients. We first prove the following result due to Bonsall [1958].

Theorem 5.6. *The uniform closure of* Π_+ *on* $[-1,0]$ *is exactly the set of* f *in* $C[-1,0]$ *for which* $f(0) \geq 0$.

Proof. Let $g_n(x) = (1+x)^n$ and $\phi_n(x) = x^n$ for all $n \in \mathbb{Z}_+$. Note that g_n and ϕ_n are in Π_+, and

$$g_n = \sum_{k=0}^{n} \binom{n}{k} \phi_k.$$

Assume F is a continuous linear functional F on $C[-1,0]$ satisfying $F(\phi_n) \geq 0$ for every $n \in \mathbb{Z}_+$. Then

$$F(g_n) \geq \binom{n}{k} F(\phi_k).$$

Now $\|g_n\| = 1$ for all n. Thus

$$\|F\| \geq \binom{n}{k} F(\phi_k),$$

for each $n \geq k$. Fix $k \geq 1$ and let $n \to \infty$. This implies that $F(\phi_k) = 0$ for all $k = 1, 2, \ldots$. Thus each $\pm\phi_k$, $k \geq 1$, is in the uniform closure of Π_+ on $[-1,0]$. As every $f \in C[-1,0]$ satisfying $f(0) = 0$ is in the uniform closure of the space generated by the $\pm\phi_k$, $k \geq 1$, the result now easily follows.

Bonsall actually proves that each linear functional F, as above, is necessarily of the exact form $F(f) = cf(0)$ where $c = F(1) \geq 0$. This he proves as follows. For each $f \in C[-1,0]$ and $\varepsilon > 0$, let $p \in \Pi$ satisfy

$$\|f - p\| < \varepsilon$$

Thus we have $|f(0) - p(0)| < \varepsilon$. Since $F(\phi_k) = 0$, $k \geq 1$, we obtain

$$|F(f) - F(1)p(0)| = |F(f - p(0))| = |F(f - p)| < \varepsilon\|F\|.$$

Furthermore

$$|F(f) - F(1)f(0)| = |F(f) - F(1)p(0) + F(1)p(0) - F(1)f(0)| < 2\varepsilon\|F\|.$$

As this is valid for each $\varepsilon > 0$ we obtain

$$F(f) = F(1)f(0). \ ∎$$

There is an alternative method of proving this result via the Stone-Weierstrass theorem (actually a slight generalization thereof). Consider the set of f in $C[-1,0]$ satisfying $f(0)=0$. Now $e^{\alpha x}-1$ is in the uniform closure of Π_+ on $[-1,0]$ for $\alpha>0$. (Truncate the power series expansion about 0.) Furthermore if f and g are in this closure then so is fg. As $e^{\alpha x}-1$ approaches -1 uniformly on $[-1,\delta]$ for any $\delta<0$ as $\alpha\to\infty$, and is bounded on $[\delta,0]$, it follows that for f in the uniform closure of Π_+ on $[-1,0]$ and satisfying $f(0)=0$ we also have $-f$ in this same closure. In addition $p(x)=x$ is nonzero for all $x\neq 0$ and separates points. Thus from an elementary generalization of the Stone-Weierstrass theorem the uniform closure of Π_+ on $[-1,0]$ contains the set of all f in $C[-1,0]$ satisfying $f(0)=0$. The result now follows.

Thus for any $a<b<0$ the uniform closure of Π_+ on $[a,b]$ is exactly all of $C[a,b]$. What happens if $[a,b]\subseteq[0,\infty)$? It is well known that in this case the uniform closure of Π_+ is simply the set of analytic functions in $[a,b]$ given by a power series about 0 with nonnegative coefficients which converges on $[a,b]$.

There are also somewhat surprising results due to Nussbaum, Walsh [1998], generalizing work of Toland [1996]. These results are used to investigate when the spectral radius of a positive, bounded linear operator belong to its spectrum. A special case of what they prove is the following:

THEOREM 5.7. *For any $a<-1$ the uniform closure of Π_+ on $[a,1]$ contains the set of all f in $C[a,1]$ that vanish identically on $[-1,1]$.*

Proof. We present two proofs of this result. The first proof uses the Hahn-Banach theorem and is contained in Nussbaum, Walsh [1998]. The second proof is constructive.

Assume we are given a continuous linear functional F on $C[a,1]$ satisfying $F(x^n)\geq 0$ for all $n\in\mathbb{Z}_+$. From the Riesz representation theorem, this implies the existence of a Borel measure μ of bounded total variation satisfying

$$\int_a^1 x^n d\mu(x)\geq 0$$

for all $n\in\mathbb{Z}_+$. We will prove that $\operatorname{supp}\{\mu\}\subseteq[-1,1]$. As this is true then

$$\int_a^1 f(x)d\mu(x)=0$$

for every f in $C[a,1]$ that vanishes identically on $[-1,1]$, proving our theorem.

To this end, consider

$$G(z)=\int_a^1 \frac{1}{z-x}d\mu(x).$$

G is analytic in $\mathbb{C}\backslash[a,1]$, and vanishes at ∞. For $|z|>\lambda=\sup\{|x|: x\in\operatorname{supp}\{\mu\}\}$ we have

$$G(z)=\sum_{n=1}^{\infty}\frac{c_{n-1}}{z^n}$$

where

$$c_n=\int_a^1 x^n d\mu(x)\geq 0.$$

Note that $H(z) = G(1/z)$ is analytic in $\mathbb{C}\backslash\{(-\infty, 1/a] \cup [1,\infty)\}$ and has about the origin a power series expansion with nonnegative coefficients. From a theorem of Pringsheim, if the radius of convergence of the power series is $\rho > 0$ then the point $z = \rho$ is a singular point of the analytic function represented by the power series. As the power series converges on $[0, 1)$ the radius of convergence is at least 1, and therefore H is analytic in $\mathbb{C}\backslash\{(-\infty, -1] \cup [1,\infty)\}$ and G analytic in $\mathbb{C}\backslash[-1, 1]$. That is, G is in fact analytic in $[a, -1)$. This implies, see Nussbaum, Walsh [1998, p. 2371], that the measure μ has no support in $[a, -1)$.

The second proof of this result is based on a variation of a proof to be found in Orlicz [1992, p. 99]. For $n \in \mathbb{N}$, odd, consider the function

$$g_n(x) = \int_0^x e^{t^n/n} - 1\,dt.$$

Note that the integrand is uniformly bounded on $[a, 1]$ and

$$\lim_{n\to\infty} e^{t^n/n} - 1 = \begin{cases} 0, & -1 \le t \le 1 \\ -1, & a \le t < -1. \end{cases}$$

As

$$e^{t^n/n} - 1 = \sum_{k=1}^{\infty} \left(\frac{t^n}{n}\right)^k \frac{1}{k!}$$

this function is in the uniform closure of Π_+. Thus, so is g_n . Set

$$G(x) = \begin{cases} -(x + 1), & a \le x \le -1 \\ 0, & -1 \le x \le 1. \end{cases}$$

Then

$$\lim_{n\to\infty} (G(x) - g_n(x)) = 0$$

uniformly on $[a, 1]$. That is, for all $x \in [-1, 1]$

$$\left| \int_0^x (e^{t^n/n} - 1)\,dt \right| \le e^{1/n} - 1,$$

while for $x \in [a, -1]$

$$\begin{aligned}\left| -(x+1) - \int_0^x (e^{t^n/n} - 1)\,dt \right| &= \left| -(x+1) - \int_{-1}^x (e^{t^n/n} - 1)\,dt + \int_{-1}^0 (e^{t^n/n} - 1)\,dt \right| \\ &\le \int_x^{-1} e^{t^n/n} dt + \int_{-1}^0 (1 - e^{t^n/n}) dt \le \int_a^{-1} e^{t^n/n} dt + (1 - e^{-1/n}).\end{aligned}$$

Thus G is in the uniform closure of Π_+ on $[a, 1]$.

Moreover, as seen above, the function $e^{t^n/n} - 1$ is uniformly bounded and approaches

$$H(x) = \begin{cases} 0, & -1 \le x \le 1, \\ -1, & a \le x < -1. \end{cases}$$

The convergence to H is uniform in $[a, 1]$, away from any neighbourhood of -1. Thus $GH = -G$ is also in the uniform closure of Π_+ on $[a, 1]$, and therefore the uniform closure of Π_+ on $[a, 1]$ contains the algebra generated by G. An elementary generalization of the Stone-Weierstrass theorem implies that the uniform closure of Π_+ on $[a, 1]$ contains the set of all f in $C[a, 1]$ which vanish identically on $[-1, 1]$. ∎

The above result is an extension of Theorem II′ in Orlicz [1992, p. 96]. Orlicz proved that for every $f \in C[a,-1]$ for which $f(-1)=0$ and each $\varepsilon > 0$, there exists a $p \in \Pi$ of the form

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

simultaneously satisfying

$$\|f - p\|_{[a,-1]} < \varepsilon$$

and

$$\sum_{k=0}^{n} |a_k| < \varepsilon.$$

References

N. I. Achieser [1956], *Theory of Approximation*, Frederick Ungar, New York, originally published in Russian in 1947.

F. Altomare and M. Campiti [1994], *Korovkin-type Approximation Theory and its Applications*, Walter de Gruyter, Berlin - New York.

A. Atzmon and A. Olevskiĭ [1996], *Completeness of integer translates in function spaces on* $\mathbb{R}$, J. Approx. Theory 87, 291–327.

B. Baillaud and H. Bourget [1905], *Correspondance d'Hermite et de Stieltjes*, Gauthier-Villars, Paris, Tome II.

S. Banach [1929], *Sur les fonctionnelles linéaires*, Studia Math. 1, 211–216 and 223–239.

S. Banach [1932], *Théorie des Opérations Linéaires*, Hafner, New York.

S. N. Bernstein [1912], *Sur l'ordre de la meilleure approximation des fonctions continues par les polynômes de degré donné*, Mem. Cl. Sci. Acad. Roy. Belg. 4, 1–103.

S. N. Bernstein [1912/13], *Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités*, Comm. Soc. Math. Kharkov 13, 1–2. Also appears in Russian translation in Bernstein's *Collected Works*.

S. Bernstein [1913], *Sur les recherches récentes relatives à la meilleure approximation des fonctions continues par des polynômes*, in: Proceedings of the Fifth International Congress of Mathematicians (Cambridge, 1912), E. W. Hobson and A. E. H. Love (eds.), Cambridge, Vol. I, 256–266. Also appears in Russian translation in Bernstein's *Collected Works*.

H. Bohman [1952], *On approximation of continuous and of analytic functions*, Ark. Mat. 2, 43–56.

F. F. Bonsall [1958], *Linear operators in complete positive cones*, Proc. London Math. Soc. 8, 53–75.

P. Borwein and T. Erdélyi [1995], *Polynomials and Polynomial Inequalities*, Springer-Verlag, New York.

R. C. Buck [1962], *Studies in Modern Analysis, Vol. 1*, Mathematical Association of America, Washington, D. C. Contains a reprint of Stone [1948].

R. B. Burckel and S. Saeki [1983], *An elementary proof of the Müntz-Szász theorem*, Expo. Math. 4, 335–341.

E. W. Cheney [1966], *Introduction to Approximation Theory*, McGraw-Hill, New York.

J. B. Conrey [2003], *The Riemann Hypothesis*, Notices of the A. M. S. 50, 341–353.

L. Fejér [1900], *Sur les fonctions bornées et intégrables*, C. R. Acad. Sci. Paris 131, 984–987.

R. P. Feinerman and D. J. Newman [1974], *Polynomial Approximation*, Williams and Wilkins, Baltimore.

M. v. Golitschek [1983], *A short proof of Muñtz's theorem*, J. Approx. Theory 39, 394–395.

H. Hahn [1927], *Über lineare Gleichungssysteme in linearen Räumen*, J. Reine Angew. Math. 157, 214–229.

E. Helly [1912], *Über linearer Functionaloperationen*, Sitzung. der Oester. Akad. Wissen. (Wien) 121, 265–297.

P. P. Korovkin [1953], *On convergence of linear positive operators in the space of continuous functions*, Dokl. Akad. Nauk SSSR 90, 961–964.

P. P. Korovkin [1960], *Linear Operators and Approximation Theory*, Hindustan Publ. Corp., Delhi. (The Russian original appeared in 1959.)

H. Lebesgue [1898], *Sur l'approximation des fonctions*, Bull. Sciences Math. 22, 278–287.

M. Lerch [1892], *O hlavni vete theorie funkci vytvorujicich* [On the main theorem on generating functions], (in Czech), Rozpravy Ceske Akademie v. Praze 1, 681–685.

M. Lerch [1903], *Sur un point de la théorie des fonctions génératrices d'Abel*, Acta Math. 27, 339–351.

M. Leshno, V. Ya. Lin, A. Pinkus and S. Schocken [1993], *Multilayer feedforward networks with a non-polynomial activation function can approximate any function*, Neural Networks 6, 861–867.

B. Ja. Levin [1964], *Distribution of Zeros of Entire Functions*, Transl. Math. Monographs 5, Amer. Math. Soc., Providence.

B. Ja. Levin [1996], *Lectures on Entire Functions*, Transl. Math. Monographs 150, Amer. Math. Soc., Providence.

N. Levinson [1940], *Gap and Density Theorems*, Amer. Math. Soc., Colloquium Publ. 26, New York.

W. A. J. Luxemburg [1976], *Müntz-Szász type approximation results and the Paley-Weiner theorem*, in: Approximation Theory II, G. G. Lorentz, C. K. Chui, L. L. Schumaker (eds.), Academic Press, New York, 437–448.

W. A. J. Luxemburg and J. Korevaar [1971], *Entire functions and Müntz-Szász type approximation*, Trans. Amer. Math. Soc. 157, 23–37.

C. H. Müntz [1914], *Über den Approximationssatz von Weierstrass*, in: H. A. Schwarz's Festschrift, Berlin, 303–312.

L. Nachbin [1967], *Elements of Approximation Theory*, Van Nostrand, Princeton.

N. Nikolski [1999], *Remarks concerning completeness of translates in function spaces*, J. Approx. Theory 98, 303–315.

R. D. Nussbaum and B. Walsh [1998], *Approximation by polynomials with nonnegative coefficients and the spectral theory of positive operators*, Trans. Amer. Math. Soc. 350, 2367–2391.

W. Orlicz [1992], *Linear Functional Analysis*, World Scientific, Singapore.

R. Paley and N. Wiener [1934], *Fourier Transforms in the Complex Domain*, Amer. Math. Soc., Colloquium Publ. 19, New York.

E. Picard [1891], *Sur la représentation approchée des fonctions*, C. R. Acad. Sci. Paris 112, 183–186.

A. Pinkus [1996], *TDI-subspaces of $C(\mathbb{R}^d)$ and some density problems from neural networks*, J. Approx. Theory 85, 269–287.

A. Pinkus [1999], *Approximation theory of the MLP model in neural networks*, Acta Numerica 8, 143–195.

A. Pinkus [2000], *Weierstrass and approximation theory*, J. Approx. Theory 107, 1–66.

J. B. Prolla [1993], *Weierstrass–Stone, the Theorem*, Peter Lang, Frankfurt.

F. Riesz [1909], *Sur les opérations fonctionnelles linéaires*, C. R. Acad. Sci. Paris 149, 974–977. Also appears in *Oeuvres*, Volume I, 400–402.

F. Riesz [1910], *Sur certains systèmes d'équations fonctionnelles et l'approximation des fonctions continues*, C. R. Acad. Sci. Paris 150, 674–677. Also appears in *Oeuvres* of F. Riesz. This paper is on pages 403–406 of Volume I. Unfortunately in some copies of the *Oeuvres* there was a mix-up with a previous paper, and the pages are 403 and 404 followed by 398 and 399.

F. Riesz [1911], *Sur certains systèmes singuliers d'équations intégrales*, Ann. Sci. Ecole Norm. Sup. 28, 33–62. Also appears in *Oeuvres*, Volume II, 798–827.

L. C. G. Rogers [1981], *A simple proof of Müntz's theorem*, Math. Proc. Camb. Phil. Soc. 90, 1–3.

W. Rudin [1966], *Real and Complex Analysis*, McGraw-Hill, New York.

E. Schmidt [1905], *Entwicklung willkürlicher Funktionen nach Systemen vorgeschriebener*, dissertation, Göttingen.

L. Schwartz [1943], *Étude des sommes d'exponentielles réelles*, Hermann, Paris.

L. Schwartz [1944], *Sur certaines familles non fondamentales de fonctions continues*, Bull. Soc. Math. France 72, 141–145.

L. Schwartz [1947], *Théorie générale des fonctions moyenne-périodiques*, Ann. Math. 48, 857–928.

M. H. Stone [1937], *Applications of the theory of Boolean rings to general topology*, Trans. Amer. Math. Soc. 41, 375–481.

M. H. Stone [1948], *A generalized Weierstrass approximation theorem*, Math. Magazine 21, 167–184, 237–254.

J. F. Toland [1996], *Self-adjoint operators and cones*, J. London Math. Soc. 53, 167–183.

K. Weierstrass [1885], *Über die analytische Darstellbarkeit sogenannter willkürlicher Functionen einer reellen Veränderlichen*, Sitzungsberichte der Akademie zu Berlin, 633–639 and 789–805. (This appeared in two parts. An expanded version of this paper with ten additional pages also appeared in *Mathematische Werke*, Vol. 3, 1–37, Mayer & Müller, Berlin, 1903.)

K. Weierstrass [1886], *Sur la possibilité d'une représentation analytique des fonctions dites arbitraires d'une variable réelle*, J. Math. Pures Appl. 2, 105–113 and 115-138. (This is a translation of Weierstrass [1885] and, as the original, it appeared in two parts and in subsequent issues, but under the same title. This journal was, at the time, called *Journal de Liouville*.)