Density in Approximation Theory

Allan Pinkus

14 January 2005

Abstract. Approximation theory is concerned with the ability to approximate functions by simpler and more easily calculated functions. 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 in the set of functions we wish to approximate? In this work we survey some of the main density results and density methods.

MSC: 41-02, 41-03, 41A45

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, series, or as 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, the birth of approximation theory becomes an inevitable and unavoidable development.

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

Surveys in Approximation Theory
Volume 1, 2005. pp. 1–45.

ISSN 1555-578X

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 the partial sums of the Fourier series converge to the function in question.

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 in the set of functions we wish to approximate? That is, can we approximate any function in our set, as well as we might wish, using arbitrary functions from our given family? In this work we survey some of the main density results and density methods.

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 compact 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 further 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 some 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. A monograph many times the length of this work would not suffice to include all results. In addition, we do not prove all the results we quote. Writing a paper such as this involves compromises. We hope, nonetheless, that you the reader will find something here of interest.

2 The Weierstrass Approximation Theorems

We first fix some notation. We let $C[a,b]$ denote the class of continuous real-valued functions on the closed interval $[a,b]$, and $\widetilde{C}[a,b]$ the class of functions in $C[a,b]$ satisfying $f(a) = f(b)$. ($\widetilde{C}[a,b]$ may be regarded as the restriction to $[a,b]$ of $(b-a)$-periodic functions in $C(\mathbb{R})$.) We denote by Π_n the space of algebraic polynomials of degree at most n, i.e.,

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

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

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

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 and function series. He viewed the results obtained in this 1885 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*. 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.

It is interesting to read this paper of Weierstrass, as his perception of these approximation theorems was most 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.

Explicitly, let (ε_n) be any sequence of positive values for which $\sum_{n=1}^{\infty} \varepsilon_n < \infty$. Let p_n be an algebraic polynomial (which exists by Theorem 2.1) satisfying

$$\|f - p_n\| := \max_{a \le x \le b} |f(x) - p_n(x)| < \varepsilon_n, \qquad n = 1, 2, \ldots$$

Set $q_0 = p_1$ and $q_n = p_{n+1} - p_n$, $n = 1, 2, \ldots$ Then

$$f(x) = \sum_{n=0}^{\infty} q_n(x).$$

Thus every continuous function can be represented by a polynomial series that converges 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 proved was that every function in $\widetilde{C}[0, 2\pi]$ can 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 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.

Weierstrass' proof of Theorem 2.1 is rather straightforward. The same is not quite true of his proof of Theorem 2.2. He extends f from $[a,b]$ so that it is continuous and bounded on all of $\mathbb{R}$. He then smooths f by convolving it with the normalized heat (Gauss) kernel $(1/k\sqrt{\pi})e^{-(x/k)^2}$. This "smoothed" f_k is entire and is therefore uniformly approximable on the finite interval $[a,b]$ by its truncated power series. Moreover the f_k uniformly approximate f on $[a,b]$ as $k \to 0+$. Together this implies the desired result.

Over the next twenty-five or so years numerous alternative proofs were given to one or the other of these two Weierstrass results by a roster of some of the best analysts of the period. The proofs use diverse ideas and techniques. 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] and Pinkus [2005]. We explain, without going into all the details, 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 $|x|$ for all $|x| \le 1$. 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

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

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

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

and set

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

The σ_n are 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 was also the first proof which used a specifically given sequence of *linear* operators.

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*

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

hold 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 concerning 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

$$\int_0^1 x^n h(x)\,\mathrm{d}x = 0, \qquad n = 0, 1, \ldots,$$

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. If p_k is a sequence of polynomials that uniformly approximate h, then

$$\lim_{k\to\infty} \int_0^1 p_k(x)h(x)\,\mathrm{d}x = \int_0^1 [h(x)]^2\,\mathrm{d}x.$$

However

$$\int_0^1 p_k(x)h(x)\,\mathrm{d}x = 0$$

for every k, and thus

$$\int_0^1 [h(x)]^2\,\mathrm{d}x = 0$$

which, since h is continuous, implies $h = 0$.

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. $\square$

Shortly thereafter Helly [1912] applied these results to a question concerning the range of an integral operator. 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*

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

is the existence of a constant L for which

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

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*

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

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

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

we also have

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

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 come close). 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 Theorems 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 mentioned. 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 (form) 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 (form) 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 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 any 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 $f(x) \neq 0$.*

Proof: The necessity of the two conditions is obvious. We prove the sufficiency.

First some preliminaries. From the Weierstrass theorem we have the existence of a sequence of algebraic polynomials (p_n) (with constant term zero) that uniformly approximates the function $|t|$ on $[-c, c]$, any $c > 0$. As such, if f is in $\overline{A}$, the closure of A in the uniform norm, then so is $p_n(f)$ for each n which implies that $|f|$ is also in $\overline{A}$. Furthermore

$$\max\{f(x), g(x)\} = \frac{f(x) + g(x) + |f(x) - g(x)|}{2}$$

and

$$\min\{f(x), g(x)\} = \frac{f(x) + g(x) - |f(x) - g(x)|}{2}.$$

It thus follows that if $f, g \in \overline{A}$, then $\max\{f, g\}$ and $\min\{f, g\}$ are also in $\overline{A}$. This of course extends to the maximum and minimum of any finite number of functions.

Finally, let x, y be any distinct points in X, and $\alpha, \beta \in \mathbb{R}$. We claim that there exists an $h \in A$ satisfying the interpolation conditions $h(x) = \alpha$ and $h(y) = \beta$. By assumption there exists a $g \in A$ for which $g(x) \neq g(y)$, and functions f_1 and f_2 in A such that $f_1(x) \neq 0$ while $f_2(y) \neq 0$. If $g(x) = 0$ then we can construct the desired h as a linear combination of g and f_1. Similarly, if $g(y) = 0$ then we can construct the desired h as a linear combination of g and f_2. Assuming $g(x)$ and $g(y)$ are both not zero, the desired h can be constructed, for example, as a linear combination of g and g^2.

We now present a proof of the theorem. Given $f \in C(X)$, $\varepsilon > 0$ and $x \in X$, for every $y \in X$ let $h_y \in A$ satisfy $h_y(x) = f(x)$ and $h_y(y) = f(y)$. Since f and h_y are continuous there exists a neighborhood V_y of y for which $h_y(w) \geq f(w) - \varepsilon$ for all $w \in V_y$. The $\cup_{y \in X} V_y$ cover X. As X is compact, it has a finite subcover, i.e., there are points $y_1, \ldots, y_n$ in X such that

$$\bigcup_{i=1}^{n} V_{y_i} = X.$$

Let $g = \max\{h_{y_1}, \ldots, h_{y_n}\}$. Then $g \in \overline{A}$ and $g(w) \geq f(w) - \varepsilon$ for all $w \in X$.

The above g depends upon x, so we shall now denote it by g_x. It satisfies $g_x(x) = f(x)$ and $g_x(w) \geq f(w) - \varepsilon$ for all $w \in X$. As f and g_x are continuous there exists a neighborhood U_x of x for which $g_x(w) \leq f(w) + \varepsilon$ for all $w \in$

U_x. Since $\cup_{x \in X} U_x$ covers X, it has a finite subcover. Thus there exist points $x_1, \ldots, x_m$ in X for which

$$\bigcup_{i=1}^{m} U_{x_i} = X.$$

Let

$$F = \min\{g_{x_1}, \ldots, g_{x_m}\}.$$

Then $F \in \overline{A}$ and

$$f(w) - \varepsilon \le F(w) \le f(w) + \varepsilon$$

for all $w \in X$. Thus

$$\|f - F\| \le \varepsilon.$$

This implies that $f \in \overline{A}$. □

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 and proof as presented here is taken from Korovkin's book.

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

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* $f \in C[a,b]$.

Proof: Let $f \in C[a,b]$. As f is uniformly continuous, given $\varepsilon > 0$ there exists a $\delta > 0$ such that if $|x_1 - x_2| < \delta$ then $|f(x_1) - f(x_2)| < \varepsilon$.

For each $y \in [a,b]$, set

$$p_u(x) = f(y) + \varepsilon + \frac{2\|f\|(x-y)^2}{\delta^2}$$

and

$$p_\ell(x) = f(y) - \varepsilon - \frac{2\|f\|(x-y)^2}{\delta^2}.$$

Since

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

for $|x-y| < \delta$, and

$$|f(x) - f(y)| < \frac{2\|f\|(x-y)^2}{\delta^2}$$

for $|x-y| > \delta$, it is readily verified that

$$p_\ell(x) \le f(x) \le p_u(x)$$

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

Since the L_n are positive linear operators, this implies that

$$(L_n p_\ell)(x) \le (L_n f)(x) \le (L_n p_u)(x) \tag{4.1}$$

for all $x \in [a,b]$, and in particular for $x = y$.

For the given fixed f, ε and δ the p_u and p_ℓ are quadratic polynomials that depend upon y. Explicitly

$$p_u(x) = \left(f(y) + \varepsilon + \frac{2\|f\|y^2}{\delta^2} \right) - \left(\frac{4\|f\|y}{\delta^2} \right) x + \left(\frac{2\|f\|}{\delta^2} \right) x^2.$$

Since the coefficients are bounded independently of $y \in [a,b]$, and

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

uniformly on $[a,b]$, it follows that there exists an N such that for all $n \ge N$, and every choice of $y \in [a,b]$ we have

$$|(L_n p_u)(x) - p_u(x)| < \varepsilon$$

and

$$|(L_n p_\ell)(x) - p_\ell(x)| < \varepsilon$$

for all $x \in [a, b]$. That is, $L_n p_u$ and $L_n p_\ell$ converge uniformly in both x and y to p_u and p_ℓ, respectively. Setting $x = y$ we obtain

$$(L_n p_u)(y) < p_u(y) + \varepsilon = f(y) + 2\varepsilon$$

and

$$(L_n p_\ell)(y) > p_\ell(y) - \varepsilon = f(y) - 2\varepsilon.$$

Thus given $\varepsilon > 0$ there exists an N such that for all $n \geq N$ and every $y \in [a, b]$ we have from (4.1)

$$f(y) - 2\varepsilon < (L_n f)(y) < f(y) + 2\varepsilon.$$

This proves the theorem. □

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 $U_n = \text{span}\{u_1, \ldots, u_n\}$, $n = 1, 2, \ldots$, 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 $C[0, 1]$. Recall from section 2 that for each such f

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

We can consider the (B_n) as a sequence of positive linear operators mapping $C[a, b]$ 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 $[0, 1]$.

Example 4.3. Recall from section 2 that the Fejér operators σ_n maps $\widetilde{C}[0, 2\pi]$ into T_n. It is easily checked that σ_n is a positive linear operator. Furthermore, $\sigma_n(1; x) = 1$, $\sigma_n(\sin x; x) = (n/(n+1)) \sin x$, and $\sigma_n(\cos x; x) = (n/(n+1)) \cos x$. Thus from the periodic version of the Bohman-Korovkin theorem $\sigma_n(g)$ converges uniformly to g on $[0, 2\pi]$, for each $g \in \widetilde{C}[0, 2\pi]$.

5 Some Univariate 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*

$$\sum_{k=1}^{\infty} \frac{1}{\lambda_k} = \infty. \tag{5.1}$$

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.) We present here the classical proof due to Müntz, with some additions from Szász [1916] that put Müntz's argument into a more elegant form.

Proof: Let

$$M_n = \operatorname{span}\{x^{\lambda_0}, \ldots, x^{\lambda_n}\},$$

and

$$E(f, M_n)_\infty = \min_{p \in M_n} \|f - p\|_\infty.$$

Based on the Weierstrass theorem it is both necessary and sufficient to prove that

$$\lim_{n \to \infty} E(x^m, M_n)_\infty = 0$$

for each $m = 0, 1, 2, ...$

To estimate $E(x^m, M_n)_\infty$ we first calculate

$$E(f, M_n)_2 = \min_{p \in M_n} \|f - p\|_2,$$

where $\|\cdot\|_2$ is the $L^2[0,1]$ norm. It is well known that

$$E^2(f, M_n)_2 = \frac{G(x^{\lambda_0}, \dots, x^{\lambda_n}, f)}{G(x^{\lambda_0}, \dots, x^{\lambda_n})}$$

where $G(f_1, \dots, f_k)$ is the Gramian of $f_1, \dots, f_k$, i.e.,

$$G(f_1, \dots, f_k) = \det\left(\langle f_i\,, f_j\rangle\right)_{i,j=1}^k .$$

As

$$\langle x^p\,, x^q\rangle = \int_0^1 x^p x^q \,\mathrm{d}x = \frac{1}{p+q+1}$$

and

$$\det\left(\frac{1}{a_i + b_j}\right)_{i,j=1}^r = \frac{\prod_{1\le j<i\le r}(a_i - a_j)(b_i - b_j)}{\prod_{i,j=1}^r (a_i + b_j)},$$

a simple calculation leads to

$$E^2(x^m, M_n)_2 = \frac{\prod_{k=0}^n (m - \lambda_k)^2}{(2m+1)\prod_{k=0}^n (m + \lambda_k + 1)^2}.$$

Thus, as is easily proven,

$$\lim_{n\to\infty} E(x^m, M_n)_2 = 0$$

if and only if

$$\lim_{n\to\infty} \prod_{k=0}^n \frac{m - \lambda_k}{m + \lambda_k + 1} = 0,$$

i.e.,

$$\prod_{k=0}^\infty \left(1 - \frac{2m+1}{m + \lambda_k + 1}\right) = 0.$$

Assuming $m \ne \lambda_k$ for every k (otherwise there was no reason to do this calculation) we have $1 \ne (2m+1)/(m + \lambda_k + 1) > 0$ and

$$\lim_{k\to\infty} \frac{2m+1}{m + \lambda_k + 1} = 0.$$

Thus

$$\prod_{k=0}^\infty \left(1 - \frac{2m+1}{m + \lambda_k + 1}\right) = 0$$

if and only if

$$\sum_{k=0}^{\infty} \frac{2m+1}{m+\lambda_k+1} = \infty,$$

that in turn is equivalent to

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

independent of m. So a necessary and sufficient condition for density in the $L^2[0,1]$ norm is that (5.1) holds.

We now consider $C[0,1]$. Assume

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

Then $E(x^m, M_n)_2$ does not tend to zero as $n \to \infty$ for every m that is not one of the λ_k. As

$$E(f, M_n)_2 \leq E(f, M_n)_\infty$$

for every $f \in C[0,1]$, we have that the system

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

is not fundamental in $C[0,1]$. Furthermore, if $\lambda_0 > 0$ then all the functions $x^{\lambda_0}, x^{\lambda_1}, x^{\lambda_2}, \ldots$ vanish at $x = 0$, and density cannot possibly hold.

Let us now assume that (5.1) holds, and $\lambda_0 = 0$. We will show how to uniformly approximate each x^m, $m \geq 1$. For $x \in [0,1]$

$$\begin{aligned}
|x^m - \sum_{k=1}^{n} a_k x^{\lambda_k}| &= \left| \int_0^x \left(m t^{m-1} - \sum_{k=1}^{n} a_k \lambda_k t^{\lambda_k - 1} \right) \mathrm{d}t \right| \\
&\leq \int_0^1 \left| m t^{m-1} - \sum_{k=1}^{n} a_k \lambda_k t^{\lambda_k - 1} \right| \mathrm{d}t \\
&\leq \left(\int_0^1 \left| m t^{m-1} - \sum_{k=1}^{n} a_k \lambda_k t^{\lambda_k - 1} \right|^2 \mathrm{d}t \right)^{1/2}.
\end{aligned}$$

Thus we can approximate x^m arbitrarily well in the uniform norm from the system $x^{\lambda_1}, x^{\lambda_2}, \ldots$ if we can approximate x^{m-1} arbitrarily well in the $L^2[0,1]$ norm from the system $x^{\lambda_1 - 1}, x^{\lambda_2 - 1}, \ldots$. We know that the latter holds if

$$\sum_{k \geq k_0} \frac{1}{\lambda_k - 1} = \infty$$

where k_0 is such that $\lambda_{k_0} - 1 > 0$. From (5.1) and since the λ_k are an increasing sequence tending to ∞, this condition necessarily holds. This proves the sufficiency. $\square$

The above method of showing how the L^2 result implies the $C[0,1]$ result is due to Szász, and simplifies a more complicated argument due to Müntz that uses Fejér's proof of the Weierstrass theorem. 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].

The above proof of Müntz and Szász as well as most of the functional analytic proofs, that use analytic methods, first prove the L^2 result. Rudin's approach is more direct, and we reproduce it here.

Rudin's Proof: Assume $0 = \lambda_0 < \lambda_1 < \cdots$. If (x^{λ_n}) is not fundamental in $C[0,1]$ then from the Hahn-Banach theorem and Riesz representation theorem there exists a Borel measure μ of bounded total variation such that

$$\int_0^1 x^{\lambda_n}\,\mathrm{d}\mu(x) = 0,$$

$n = 0, 1, 2, \ldots$. As $\lambda_0 = 0$ and $\lambda_n > 0$ for all $n > 1$ we may assume the above holds for $n = 1, 2, \ldots$ and μ has no mass concentrated at 0. Set

$$f(z) = \int_0^1 x^z\,\mathrm{d}\mu(x).$$

For $x \in (0,1]$ and $\operatorname{Re} z > 0$ we have that $x^z = e^{z \ln x}$ and $|x^z| = x^{\operatorname{Re} z} \le 1$. It therefore follows that f is analytic and bounded in the right half plane, and of course satisfies

$$f(\lambda_n) = 0, \qquad n = 1, 2, \ldots$$

Now set

$$g(z) = f\left(\frac{1+z}{1-z}\right).$$

The transformation $(1+z)/(1-z)$ maps the unit disc to the right half plane. Thus $g \in H^\infty$, the space of bounded analytic functions in the unit disc, and $g(\alpha_n) = 0$ where

$$\alpha_n = \frac{\lambda_n - 1}{\lambda_n + 1}.$$

Now it is a known result associated with Blaschke products that the (α_n) are the zeros, in the unit disc, of a nontrivial $g \in H^\infty$ if and only if

$$\sum_{n=1}^\infty (1 - |\alpha_n|) < \infty.$$

It is readily checked that $\sum_{n=1}^{\infty}(1-|\alpha_n|)<\infty$ if and only if $\sum_{n=1}^{\infty} 1/\lambda_n<\infty$. Thus if $\sum_{n=1}^{\infty} 1/\lambda_n=\infty$, then $g=0$ which implies that $f=0$. But then

$$0=f(k)=\int_0^1 x^k \,\mathrm{d}\mu(x)$$

for all $k=1,2,\dots$ which implies by the Weierstrass theorem that $\mu=0$. Thus if $\sum_{n=1}^{\infty} 1/\lambda_n=\infty$ then the $(x^{\lambda_n})_{n=0}^{\infty}$ are fundamental in $C[0,1]$.

Assume that $\sum_{n=1}^{\infty} 1/\lambda_n<\infty$. How can we construct the desired measure μ? One way is as follows. Set

$$f(z)=\frac{1}{(z+2)^2}\prod_{n=0}^{\infty}\frac{\lambda_n-z}{2+\lambda_n+z}.$$

The function f is a meromorphic function with poles at -2 and $-\lambda_n-2$, and zeros at the λ_n. f is also bounded in $\operatorname{Re} z>-1$ since each factor is less than 1 in absolute value thereon. For each z satisfying $\operatorname{Re} z>-1$ we have by Cauchy's formula

$$f(z)=\frac{1}{2\pi i}\int_{\Gamma_R}\frac{f(w)}{w-z}dw$$

where Γ_R is the right semi-circle of radius $R\,(>1+|z|)$, centered at -1, together with the line from $-1-iR$ to $-1+iR$. Letting $R\to\infty$, it may be readily shown that the integral over the semi-circle tends to zero, and we obtain

$$f(z)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{f(-1+is)}{1+z-is}\,\mathrm{d}s.$$

As

$$\frac{1}{1+z-is}=\int_0^1 x^{z-is}\,\mathrm{d}x$$

for $\operatorname{Re} z>-1$ we have

$$f(z)=\int_0^1 x^z\left(\frac{1}{2\pi}\int_{-\infty}^{\infty} f(-1+is)e^{-is\ln x}ds\right)\mathrm{d}x.$$

Set

$$\mathrm{d}\mu(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty} f(-1+is)e^{-is\ln x}\,\mathrm{d}s.$$

This is the Fourier transform of $f(-1+is)$ at $\ln x$ and is bounded and continuous on $(0,1]$, since the factor $1/(2+z)^2$ in the definition of f ensures that $f(-1+is)$ is a function in L^1. Thus we have obtained our desired measure μ. $\square$

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

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

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

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

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

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

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

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

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

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

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

As

$$\int_a^b g(zx)\,\mathrm{d}\mu(x) = 0$$

and g is entire it may be shown, differentiating by z, that

$$\int_a^b x^n g^{(n)}(zx)\,\mathrm{d}\mu(x) = 0$$

for every nonnegative integer n. Setting $z = 0$ gives us

$$g^{(n)}(0) \int_a^b x^n\,\mathrm{d}\mu(x) = 0, \qquad n = 0, 1, \ldots$$

Thus

$$\int_a^b x^n\,\mathrm{d}\mu(x) = 0,$$

for all $n \in N_g$. But $\mathrm{span}\{x^n : n \in N_g\}$ is dense in $C[a,b]$, so μ is the trivial measure.

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

$$\int_a^b x^n\,\mathrm{d}\mu(x) = 0$$

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

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

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

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 $0 < \alpha < \beta < \infty$.

A question related to Müntz type problems is that of the fundamentality of the functions $(e^{\lambda_n x})$, where (λ_n) is a sequence of complex numbers. This has been considered in the space of 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 a great deal of 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. *Akhiezer's Theorem.* Let Γ be a subset of $\mathbb{R}\backslash[-1, 1]$, and consider the set

$$\mathcal{N}_\Gamma = \text{span}\left\{\frac{1}{t - \gamma} : \gamma \in \Gamma\right\}.$$

When is $\mathcal{N}_\Gamma$ dense in $C[-1, 1]$? One result is similar to Theorem 5.2. It may be found in Feinerman-Newman [1974, p. 116–117], but the proof therein is somewhat different.

Proposition 5.3. *If Γ has either a finite accumulation point in $\mathbb{R}\backslash[-1, 1]$ or ∞ is an accumulation point, then $\mathcal{N}_\Gamma$ is dense in $C[-1, 1]$.*

Proof: Assume $\overline{\mathcal{N}_\Gamma} \neq C[-1, 1]$. Then there exists a Borel measure μ of bounded total variation such that

$$\int_{-1}^{1} \frac{1}{t - \gamma} \, \mathrm{d}\mu(t) = 0$$

for all $\gamma \in \Gamma$. Set

$$f(z) = \int_{-1}^{1} \frac{1}{t - z} \, \mathrm{d}\mu(t).$$

Note that $f(\gamma) =$ for all $\gamma \in \Gamma$. It is readily verified that f is analytic on $\mathbb{C}\backslash[-1, 1]$, and analytic also at infinity.

Thus if Γ has either a finite accumulation point in $\mathbb{R}\backslash[-1, 1]$ or ∞ is an accumulation point, then $f = 0$. For $|z| > 1 \ge |t|$

$$f(z) = -\sum_{n=0}^{\infty} \left(\frac{1}{z}\right)^{n+1} \int_{-1}^{1} t^n \, \mathrm{d}\mu(t).$$

As $f = 0$, this then implies that

$$\int_{-1}^{1} t^n \,\mathrm{d}\mu(t) = 0$$

for all n, which from the Hahn-Banach and Weierstrass theorems implies that μ is the trivial measure. This proves the proposition. $\square$

What can be said if the only accumulation points of Γ are 1 or -1 or both? The result is known, contains the previous Proposition 5.3 as a special case, and was proven by Akhiezer, see Achieser [1956, p. 254–256].

Akhiezer's Theorem 5.4. *Let $(\gamma_n)_{n=1}^{\infty}$ be a sequence in $I\!R \backslash [-1, 1]$, and consider the set*

$$\mathcal{N} = \operatorname{span}\left\{\frac{1}{t - \gamma_n} \,:\, n = 1, 2, \ldots\right\}.$$

Then $\mathcal{N}$ is dense in $C[-1, 1]$ if and only if

$$\sum_{n=1}^{\infty} 1 - |\gamma_n - \sqrt{\gamma_n^2 - 1}| = \infty.$$

See also Borwein, Erdélyi [1995, p. 208] where a different method of proof is used. They also give the above condition as

$$\sum_{n=1}^{\infty} \sqrt{\gamma_n^2 - 1} = \infty,$$

and these two conditions are in fact equivalent. Akhiezer's proof of this theorem is delicate and detailed, dependent on the construction of specific best approximants. We will not reproduce it here. Michael Sodin has a proof which uses complex variable theory.

Example 5.4. 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, for example, both Example 5.2 and 5.3.

There is a characterization of those $f \in C(I\!R)$ for which

$$\operatorname{span}\{f(\cdot - \alpha) \,:\, \alpha \in I\!R\}$$

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

Some functions in $C(I\!R)$ have a further interesting property.

Proposition 5.5. *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*

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

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

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

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

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

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

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. $\square$

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 5.5. *The Bernstein Approximation Problem.* Assume ω is a *weight* on $\mathbb{R}$ by which we will mean a non-negative, measurable, bounded function. For each f in $C(\mathbb{R})$ satisfying

$$\lim_{|x|\to\infty} \omega(x)f(x) = 0$$

set

$$\|f\|_\omega = \sup_{x\in\mathbb{R}} \omega(x)|f(x)|,$$

and let $C_\omega(\mathbb{R})$ denote the real normed linear space of those f as above with $\|f\|_\omega < \infty$. The Bernstein approximation problem was first formulated in Bernstein [1924]. It asks for necessary and sufficient conditions on a weight ω

such that (algebraic) polynomials are dense in $C_\omega(\mathbb{R})$. That is, for each f in $C_\omega(\mathbb{R})$ and $\varepsilon > 0$ there exists a polynomial p for which $\|f - p\|_\omega < \varepsilon$. This immediately implies that ω must satisfy

$$\lim_{|x|\to\infty} \omega(x)p(x) = 0$$

for every polynomial p.

In Bernstein [1924] can be found the following result. Assume $\omega(x) = 1/q(x)$, where

$$q(x) = \sum_{n=0}^{\infty} a_n x^{2n}$$

with $a_0 > 0$, $a_n \geq 0$ for all n, and q not the constant function. Then a necessary and sufficient for polynomials to be dense in $C_\omega(\mathbb{R})$ is that

$$\int_1^\infty \frac{\ln q(x)}{1+x^2}\,\mathrm{d}x = \infty.$$

In general a condition of this form is necessary, but not sufficient. It is often sufficient for "reasonable" weights.

The literature on this problem is rather extensive including the review articles Ahiezer [1956] and Mergelyan [1956], see also Lorentz, v. Golitschek and Makovoz [1996, p. 28-33], and Timan [1963, p. 16–19]. The article of Mergelyan, as well as Prolla [1977], includes a proof of this next result. Let $\mathcal{M}_\omega$ denote the set of polynomials p satisfying $\omega(x)|p(x)| \leq 1 + |x|$ for all $x \in \mathbb{R}$, and set

$$M_\omega(z) = \sup\{|p(z)| : p \in \mathcal{M}_\omega\}.$$

Mergelyan's Theorem 5.6. *Let ω be as above. Then a necessary and sufficient for polynomials to be dense in $C_\omega(\mathbb{R})$ is that*

$$M_\omega(z) = \infty$$

for every $z \in \mathbb{C}\backslash\mathbb{R}$.

Unfortunately this condition is not easy to check.

Here is a condition that is easier to check, but which only holds for certain weights. Assume $\omega = \exp\{-Q\}$, ω is even, and Q is a convex function of $\ln x$ on $(0, \infty)$. Then

$$\int_0^\infty \frac{\ln \omega(x)}{1+x^2}\,\mathrm{d}x = -\infty$$

is both necessary and sufficient for polynomials to be dense in $C_\omega(\mathbb{R})$, see Mhaskar [1996, p. 331].

Example 5.6. *Markov Systems.* Assume we are given a sequence of functions $(u_m)_{m=0}^{\infty}$ in $C[a,b]$. What we have been asking is when this sequence is fundamental, i.e., linear combinations are dense. That is when, for each f in $C[a,b]$ and $\varepsilon > 0$, there exists a finite linear combination u of the $(u_m)_{m=0}^{\infty}$ such that

$$\|f - u\|_\infty < \varepsilon.$$

The only general result characterizing the density of such sequences is the somewhat tautological Theorem 3.6. However, when the sequence $(u_m)_{m=0}^{\infty}$ has a particular type of Chebyshev property, then P. Borwein proved a surprisingly interesting condition equivalent to density.

To explain his result we first need some definitions. Given the sequence $(u_m)_{m=0}^{\infty}$ in $C[a,b]$ we set

$$U_n = \operatorname{span}\{u_0, u_1, \ldots, u_n\}$$

for each $n \in \mathbb{Z}_+$. We say that U_n is a *Chebyshev space* if no $u \in U_n$, $u \neq 0$, has more than n distinct zeros in $[a,b]$. We say that the sequence $(u_m)_{m=0}^{\infty}$ is a *Markov sequence* if U_n is a Chebyshev space for $n = 0, 1 \ldots$. There are numerous examples of Markov sequences. For example, $(x^{\lambda_m})_{m=0}^{\infty}$ is a Markov sequence on $[a,b]$ where $a > 0$ and the $(\lambda_m)_{m=0}^{\infty}$ are arbitrary distinct real values, while $(1/(x - c_m))_{m=0}^{\infty}$ is a Markov sequence on any $[a,b]$ where the $(c_m)_{m=0}^{\infty}$ are distinct values in $\mathbb{R}\backslash[a,b]$.

In what follows we assume that the $(u_m)_{m=0}^{\infty}$ is a Markov sequence. Let $t_n \in U_n$ be of the form $t_n = u_n - v_n$ with $v_n \in U_{n-1}$, satisfying

$$\|t_n\|_\infty = \min_{v \in U_{n-1}} \|u_n - v\|_\infty.$$

It is well known, from the Chebyshev and Markov properties, that t_n is uniquely defined and has n zeros in (a,b). Let $x_1 < \cdots < x_n$ denote these n zeros and set $x_0 = a$, $x_{n+1} = b$. The *mesh* of t_n is defined by

$$m_n = \max_{i=1,\ldots,n+1} (x_i - x_{i-1}).$$

It is readily proven that for any $k < n$ the function t_k has at most one zero between any two consecutive zeros of t_n. From this it follows that

$$\lim_{n\to\infty} m_n = 0$$

if and only if

$$\liminf_{n\to\infty} m_n = 0.$$

The following result may be found in Borwein [1990], and also in Borwein, Erdélyi [1995, p. 155–158].

Borwein's Theorem 5.7. *Assume $(u_m)_{m=0}^{\infty}$ is a Markov sequence in $C^1[a,b]$ and $u_0 = 1$. Then the sequence $(u_m)_{m=0}^{\infty}$ is dense in $C[a,b]$ if and only if*

$$\lim_{n\to\infty} m_n = 0.$$

A similar result relating density to Bernstein-type inequalities is in Borwein, Erdélyi [1995b], and Borwein, Erdélyi [1995, p. 206–211].

Example 5.7. The following result is a special case of a general theorem of Schwartz [1944] (see also Pinkus [1996] and references therein). Here we again 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 5.8. *If $\sigma \in C(\mathbb{R})$, $\sigma \neq 0$, then*

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

if and only if σ is not a polynomial.

Proof: Let

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

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

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

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

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

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 $\alpha, \beta \in \mathbb{R}$

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

Thus

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

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

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

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

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

for all $\phi \in C_0^\infty(I\!R)$. 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 $C(I\!R)$. $\square$

Example 5.8. *Splines* are piecewise polynomials with a high order of continuity. For $a = \xi_0 < \xi_1 < \cdots < \xi_k < \xi_{k+1} = b$ we set

$$\mathcal{S}_n(\xi_1, \ldots, \xi_k) = \{s \in C^{(n-1)}[a,b] : \, s \in \Pi_n|_{(\xi_{i-1}, \xi_i)}, i = 1, \ldots, k+1\}.$$

Hence a function s belongs to $\mathcal{S}_n(\xi_1, \ldots, \xi_k)$ if it is a $C^{(n-1)}$ function, i.e., has a certain global level of smoothness, and is a polynomial of degree at most n on each of the intervals (ξ_{i-1}, ξ_i). We say that $\mathcal{S}_n(\xi_1, \ldots, \xi_k)$ is the space of *splines of degree n with the simple knots* $(\xi_1, \ldots, \xi_k)$. When using splines one fixes the degree and permits the number (and placement) of the knots to vary. From the perspective of numerical computations, approximation by splines enjoys many advantages over approximation by algebraic and trigonometric polynomials. As the number of knots increases the corresponding space of splines may or may not "become dense" in $C[a,b]$. Whether it does or not simply depends upon if the knots become dense in $[a,b]$.

To be more exact, for each $k = 1, 2, \ldots$ let

$$S_k = \mathcal{S}_n(\xi_1^k, \ldots, \xi_k^k)$$

for some set of k knots as above, where $\xi_0^k = a$ and $\xi_{k+1}^k = b$. For each such k, let

$$m_k = \max_{i=0,\ldots,k} (\xi_{i+1}^k - \xi_i^k),$$

denote the maximum mesh length. Then we have, see for example, de Boor [1968],

Proposition 5.9. *For each $f \in C[a,b]$ there exist $s_k \in S_k$ such that*

$$\lim_{k \to \infty} \|f - s_k\|_\infty = 0$$

if and only if $\lim_{k \to \infty} m_k = 0$.

Proof: We first assume that $\lim_{k \to \infty} m_k = 0$. The set

$$(1, x, \ldots, x^n, (x - \xi_1^k)_+^n, \ldots, (x - \xi_k^k)_+^n),$$

where x_+^n equals x^n for $x \geq 0$ and 0 for $x < 0$, is a basis for S_k. However there are also 'better' bases. They are given by B-splines

$$(B_1^k, \ldots, B_{n+k+1}^k).$$

Each B_i^k is nonnegative, $\sum_{i=1}^{n+k+1} B_i^k = 1$ on $[a, b]$, and $\overline{\operatorname{supp}} B_i^k = [\xi_{i-n-1}^k, \xi_i^k]$ where $\xi_{-n}^k \leq \cdots \leq \xi_{-1}^k \leq a$, and $b \leq \xi_{k+2}^k \leq \cdots \leq \xi_{n+k+1}^k$.

For each $\delta > 0$, let

$$\omega(f; \delta) = \max_{|x-y| \leq \delta} |f(x) - f(y)|$$

denote the usual *modulus of continuity* of f. As f is continuous on $[a, b]$ it is also uniformly continuous thereon and thus

$$\lim_{\delta \to 0^+} \omega(f; \delta) = 0.$$

Now choose $t_i^k \in \operatorname{supp} B_i^k \cap [a, b]$, $i = 1, \ldots, n + k + 1$, and set

$$s_k(x) = \sum_{i=1}^{n+k+1} f(t_i^k) B_i^k(x).$$

(s_k is called a *quasi-interpolant*.) Then for $x \in [\xi_j^k, \xi_{j+1}^k]$, $j \in \{0, 1, \ldots, k\}$ we have

$$\begin{aligned} |f(x) - s_k(x)| &= |f(x) - \sum_{i=1}^{n+k+1} f(t_i^k) B_i^k(x)| \\ &\leq \sum_{i=1}^{n+k+1} |f(x) - f(t_i^k)| B_i^k(x) \\ &= \sum_{i=j}^{n+j+2} |f(x) - f(t_i^k)| B_i^k(x) \end{aligned}$$

since $B_i^k(x) = 0$ for $i \notin \{j, \ldots, n + j + 2\}$. Thus

$$|f(x) - s_k(x)| \leq \max_{i=j,\ldots,n+j+2} |f(x) - f(t_i)| \leq \omega(f; (n+2)m_k),$$

and

$$\|f - s_k\| \leq \omega(f; (n+2)m_k),$$

from which we obtain

$$\lim_{k \to \infty} \|f - s_k\| = 0.$$

If the m_k do not tend to zero, then there is a subinterval $[c, d]$ of $[a, b]$ of positive length, and a subsequence (k_m) of (k) such that S_{k_m} has no knots in $[c, d]$. That is, each function in S_{k_m} is a polynomial of degree at most n on $[c, d]$. If $f \in C[a, b]$ is not a polynomial of degree at most n on $[c, d]$, then there exists a $C > 0$ such that

$$\|f - s\| \geq C$$

for all $s \in S_{k_m}$. $\square$

Each $\mathcal{S}_n(\xi_1,\ldots,\xi_k)$ is a linear space of splines of degree n with k fixed knots. One can also consider the nonlinear set $\mathcal{S}_{n,k}$ of splines of degree n with k *free knots*, i.e.,

$$\mathcal{S}_{n,k} = \{s(x) = \sum_{i=0}^{n} c_i x^i + \sum_{i=1}^{k} d_i (x-\xi)_+^n \,:\, c_i, d_i \in I\!R,\ a < \xi_1 < \cdots < \xi_k < b\}.$$

(Note that the set $\mathcal{S}_{n,k}$ is not closed.)

The following is a consequence of our previous density result.

Proposition 5.10. *Let $p \in (1,\infty)$. Assume $f \in L^p[a,b]\backslash\overline{\mathcal{S}_{n,k}}$ and $s^* \in \overline{\mathcal{S}_{n,k}}$ satisfies*

$$\|f - s^*\|_p \le \|f - s\|_p$$

for all $s \in \mathcal{S}_{n,k}$. Then $s^* \notin \overline{\mathcal{S}_{n,k-1}}$.

Proof: Let F be a continuous linear functional on $L^p[a,b]$ that satisfies

$$\|F\| = 1$$

and

$$F(f - s^*) = \|f - s^*\|\,.$$

As $f - s^* \ne 0$ and $p \in (1,\infty)$, such an F exists and is unique.

Our proof is by contradiction. If $s^* \in \overline{\mathcal{S}_{n,k-1}}$, then for each $d \in I\!R$ and $\xi \in [a,b]$ we have

$$\|f - s^*\|_p \le \|f - s^* - d(x-\xi)_+^n\|_p.$$

Thus from Theorem 3.5 we have

$$F((x-\xi)_+^n) = 0$$

for each $\xi \in [a,b]$. In addition, as

$$\|f - s^*\|_p \le \|f - s^* - p\|_p$$

for every $p \in \Pi_n$, we have

$$F(x^k) = 0, \qquad k = 0,\ldots,n.$$

From Proposition 5.9

$$\mathrm{span}\{1, x, \ldots, x^n, (x-\xi)_+^n \,:\, \xi \in [a,b]\}$$

is dense in $C[a,b]$ and thus in $L^p[a,b]$. From Theorem 3.6 this implies that $F = 0$. A contradiction. $\square$

The exact same argument proves, for example, that if

$$\mathcal{P}_k = \{\sum_{j=1}^{k} a_j x^{m_j} \, : \, a_j \in \mathbb{R}, m_j \in \mathbb{Z}_+, j = 1, \ldots, k\},$$

then in $L^p[a,b]$, $1 < p < \infty$, any best approximation to f from $\mathcal{P}_k$ is never contained in $\mathcal{P}_{k-1}$.

The key ingredients of the above argument are the density of the set under consideration (in the above examples, splines and algebraic polynomials) in the normed linear space E, and the fact that E is *smooth*. That is, to each nonzero element of the space E there is a unique continuous linear functional of norm one that attains its norm on the given element.

Example 5.9. 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 positive cone of all algebraic polynomials with nonnegative coefficients. We first prove the following result due to Bonsall [1958].

Theorem 5.11. *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 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 $\|g_n\| = \|\phi_n\| = 1$ 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 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 $F(\phi_k)=0$, $k\ge 1$, 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

$$F(f) = F(1)f(0). \quad \square$$

There is an alternative method of proving this result via a slight generalization of the Stone-Weierstrass theorem. 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 in $[a,b]$.

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

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

Assume we are given any continuous linear functional F on $C[a,1]$ satisfying $F(x^n)\ge 0$ 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

$$\int_a^1 x^n \,\mathrm{d}\mu(x) \ge 0$$

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

$$\int_a^1 f(x)\,\mathrm{d}\mu(x) = 0$$

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

To this end, consider

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

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

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

Note that $H(z) = G(1/z)$ is analytic in $\mathbb{C}\backslash\{(-\infty,1/a]\bigcup[1,\infty)\}$ 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 $\mathbb{C}\backslash\{(-\infty,-1]\bigcup[1,\infty)\}$ 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 $\mathrm{d}\mu$ has no support in $[a,-1)$.

The following constructive 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

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

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

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

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

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

Then

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

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

$$\left|\int_0^x e^{t^n/n}-1\,\mathrm{d}t\right|\le e^{1/n}-1,$$

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

$$\left|-(x+1)-\int_0^x\left(e^{t^n/n}-1\right)\,\mathrm{d}t\right|$$

$$=\left|-(x+1)-\int_{-1}^x\left(e^{t^n/n}-1\right)\,\mathrm{d}t+\int_{-1}^0\left(e^{t^n/n}-1\right)\,\mathrm{d}t\right|$$

$$\le\int_x^{-1}e^{t^n/n}\,\mathrm{d}t+\int_{-1}^0\left(1-e^{t^n/n}\right)\,\mathrm{d}t\le\int_a^{-1}e^{t^n/n}\,\mathrm{d}t+(1-e^{-1/n}).$$

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 $[-1,1]$. $\square$

The above result is an extension of Theorem II′ in Orlicz [1999, p. 96]. Orlicz proved that for every $f\in C[a,-1]$ satisfying $f(-1)=0$ and for each $\varepsilon>0$, there exists a $p\in\Pi$ of the form

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

simultaneously satisfying

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

and

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

6 Some Multivariate Density Results

In the section we consider applications of the results of the previous sections to multivariate functions.

Example 6.1. We start with an application of the Stone-Weierstrass theorem. Let $h_1, \ldots, h_m$ be any m fixed real-valued continuous functions defined on X, a compact set. Let

$$\mathcal{M} = \mathrm{span}\{g\big(\sum_{i=1}^{m} a_i h_i\big) : \mathbf{a} = (a_1, \ldots, a_m) \in I\!R^m,\ g \in C(I\!R)\}.$$

When is $\mathcal{M}$ dense in $C(X)$?

Proposition 6.1. *$\overline{\mathcal{M}} = C(X)$ if and only if for each $x, y \in X$, $x \neq y$, there exists $i \in \{1, \ldots, m\}$ such that $h_i(x) \neq h_i(y)$.*

Proof: If there exists an $x \neq y$ for which $h_i(x) = h_i(y)$, $i = 1, ..., m$, then for every $f \in \mathcal{M}$ we have $f(x) = f(y)$ and obviously $\overline{\mathcal{M}} \neq C(X)$. On the other hand, assume that for each $x, y \in X$, $x \neq y$, there exists $i \in \{1, \ldots, m\}$ such that $h_i(x) \neq h_i(y)$. Consider the linear span of the set

$$\big(\sum_{i=1}^{m} a_i h_i\big)^k$$

as we vary over all $\mathbf{a} \in I\!R^m$ and $k = 0, 1, 2, \ldots$. This is an algebra generated by

$$h_1^{\ell_1} \cdots h_m^{\ell_m}$$

where the ℓ_i are non-negative integers. Furthermore this algebra contains the constant function and separates points. Thus the density follows from the Stone–Weierstrass theorem. $\square$

Example 6.2. Is it true that for arbitrary compact sets X and Y we always have that $\mathrm{span}\{C(X)\times C(Y)\}$ is dense in $C(X\times Y)$? If X and Y are compact subsets of $I\!R$ this follows from the fact that algebraic polynomials are dense in $C(X\times Y)$, and each algebraic polynomial is a linear combination of products of monomials in X and monomials in Y. Similarly to Example 6.1 we have:

Dieudonné Theorem 6.2. *If X and Y are compact, then the linear space $\mathrm{span}\{C(X)\times C(Y)\}$ is dense in $C(X\times Y)$.*

Proof: For $f \in C(X)$ and $g \in C(Y)$ the function $(f\times g)(x, y) = f(x)g(y)$ is in $C(X\times Y)$. Furthermore all finite sums of the form $f_1\times g_1 + \cdots + f_m\times g_m$ clearly form a subalgebra of $C(X\times Y)$ that contains the constant function and separates points. Thus by the Stone-Weierstrass theorem $\mathrm{span}\{C(X)\times C(Y)\}$ is dense in $C(X\times Y)$. $\square$

This theorem, originally proven in Dieudonné [1937] by other methods, easily extends to a product of any finite number of compact spaces. It may also be found in Nachbin [1967] and Prolla [1977].

Example 6.3. Here is a simple application of the functional analytic approach to density. We consider $C(\mathbb{R}^n)$ with the topology of uniform convergence on compacta. As we recall, the set of functions for which the span of all their translates are not dense in $C(\mathbb{R}^n)$ are called *mean-periodic* functions. There is no known characterization of mean-periodic functions in $C(\mathbb{R}^n)$ for $n \geq 2$. However not many functions can be mean-periodic. For example

Proposition 6.3. *If $g \in C(\mathbb{R}^n) \cap L^1(\mathbb{R}^n)$, $g \neq 0$, then*

$$C(\mathbb{R}^n) = \overline{\operatorname{span}}\{g(\cdot - \mathbf{a}) : \mathbf{a} \in \mathbb{R}^n\}.$$

Proof: The continuous linear functionals on $C(\mathbb{R}^n)$ are represented by Borel measures of bounded total variation and compact support. If the above space is not dense in $C(\mathbb{R}^n)$, then there exists such a nontrivial measure μ satisfying

$$\int_{\mathbb{R}^n} g(\mathbf{x} - \mathbf{a})\,\mathrm{d}\mu(\mathbf{x}) = 0,$$

for all $\mathbf{a} \in \mathbb{R}^n$. Both g and μ have "nice" Fourier transforms. Since the above is a convolution we must have

$$\widehat{g}(\mathbf{w})\widehat{\mu}(\mathbf{w}) = 0.$$

Now $\widehat{\mu}$ is an entire function, while $\widehat{g}$ is continuous. Since $\widehat{g}$ must vanish where $\widehat{\mu} \neq 0$, it follows that $\widehat{g} = 0$ and thus $g = 0$, a contradiction. $\square$

Example 6.4. Let $\langle \cdot\,, \cdot \rangle$ denote the usual inner (scalar) product on $\mathbb{R}^n$. Applying Propositions 6.1 and 5.8 we prove the following result.

Proposition 6.4. *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 thus the above span 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 application of Proposition 6.1 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 $I\!R$ 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 5.8 we have the existence of $c_{ik}, \alpha_{ik}, \beta_{ik} \in I\!R$, $i = 1, \ldots, n_k$, $k = 1, \ldots, m$, 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. $\square$

Proposition 6.4 is a basic result in one of the models of neural network theory, see Leshno, Lin, Pinkus, Schocken [1993] and Pinkus [1999].

Example 6.5. *Ridge Functions.* Ridge functions were considered in the previous example. They are functions of the form $g(\langle \mathbf{a}, \cdot\rangle)$ for some fixed 'direction' $\mathbf{a}$ and some function $g \in C(I\!R)$. They are functions constant on the hyperplanes $\{\langle \mathbf{a}, \mathbf{x}\rangle = t\}$ for every $t \in I\!R$.

Let Ω be a subset of $I\!R^n$. In what follows we assume that Ω is a subset of S^{n-1}, i.e., all elements of Ω are of norm 1. (This is simply a convenient normalization.) The question we ask is: What are necessary and sufficient conditions on Ω such that the set of all ridge functions with directions from Ω are dense in $C(I\!R^n)$. The result we prove is due to Vostrecov, Kreines [1961], see also Lin, Pinkus [1993]. We will apply both the Weierstrass theorem and the Riesz representation theorem in obtaining these conditions.

Let

$$\mathcal{M}(\Omega) = \operatorname{span}\{g(\langle \mathbf{a}, \cdot\rangle) : \mathbf{a} \in \Omega,\, g \in C(I\!R)\}.$$

Note that we vary over all $\mathbf{a} \in \Omega$ and *all* $g \in C(I\!R)$.

Theorem 6.5. *The linear space $\mathcal{M}(\Omega)$ is dense in $C(I\!R^n)$ in the topology of uniform convergence on compacta if and only if the only homogeneous polynomial (of n variables) that vanishes identically on Ω is the zero polynomial.*

Proof: $(\Rightarrow)$. Assume there exists a nontrivial homogeneous polynomial p of degree k that vanishes on Ω. Let

$$p(\mathbf{y}) = \sum_{|\mathbf{m}|=k} b_{\mathbf{m}}\mathbf{y}^{\mathbf{m}},$$

where $\mathbf{m} = (m_1, \ldots, m_n) \in \mathbb{Z}_+^n$, $|\mathbf{m}| = m_1 + \cdots + m_n$ and $\mathbf{y}^{\mathbf{m}} = y_1^{m_1} \cdots y_n^{m_n}$.

Choose any $\phi \in C_0^\infty(\mathbb{R}^n)$, $\phi \neq 0$. For each $\mathbf{m} \in \mathbb{Z}_+^n$, $|\mathbf{m}| = k$, set

$$D^{\mathbf{m}} = \frac{\partial^k}{\partial x_1^{m_1} \cdots \partial x_n^{m_n}},$$

and define

$$\psi(\mathbf{x}) = \sum_{|\mathbf{m}|=k} b_{\mathbf{m}} D^{\mathbf{m}} \phi(\mathbf{x}).$$

Note that $\psi \in C_0^\infty(\mathbb{R}^n)$, $\psi \neq 0$, ($\mathrm{supp}\psi \subseteq \mathrm{supp}\phi$), and

$$\widehat{\psi} = i^k \widehat{\phi} p$$

where $\widehat{\cdot}$ denotes the Fourier transform. As p is homogeneous, $p(\lambda\mathbf{a}) = \widehat{\psi}(\lambda\mathbf{a}) = 0$ for all $\mathbf{a} \in \Omega$ and $\lambda \in \mathbb{R}$.

We claim that

$$\int_{\mathbb{R}^n} g(\langle \mathbf{a}, \mathbf{x} \rangle) \psi(\mathbf{x}) \, \mathrm{d}\mathbf{x} = 0$$

for all $\mathbf{a} \in \Omega$ and $g \in C(\mathbb{R})$, i.e., the nontrivial linear functional defined by integrating against ψ annihilates $\mathcal{M}(\Omega)$. From the Riesz representation theorem this implies that $\mathcal{M}(\Omega)$ is not dense in $C(\mathbb{R}^n)$.

We prove this as follows. For $\mathbf{a} \in \Omega$ we write

$$0 = \widehat{\psi}(\lambda\mathbf{a}) = \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} \psi(\mathbf{x}) e^{-i\lambda\langle \mathbf{a}, \mathbf{x} \rangle} \, \mathrm{d}\mathbf{x}$$

$$= \frac{1}{(2\pi)^{n/2}} \int_{-\infty}^{\infty} \Big[\int_{\langle \mathbf{a}, \mathbf{x} \rangle = t} \psi(\mathbf{x}) d\mathbf{x} \Big] e^{-i\lambda t} \, \mathrm{d}t.$$

Since this holds for all $\lambda \in \mathbb{R}$, we have that

$$\int_{\langle \mathbf{a}, \mathbf{x} \rangle = t} \psi(\mathbf{x}) \, \mathrm{d}\mathbf{x} = 0$$

for all t. Thus for any $g \in C(\mathbb{R})$,

$$\int_{\mathbb{R}^n} g(\langle \mathbf{a}, \mathbf{x} \rangle) \psi(\mathbf{x}) \, \mathrm{d}\mathbf{x} = \int_{-\infty}^{\infty} \Big[\int_{\langle \mathbf{a}, \mathbf{x} \rangle = t} \psi(\mathbf{x}) d\mathbf{x} \Big] g(t) \, \mathrm{d}t = 0.$$

($\Leftarrow$). Assume that for a given $k \in \mathbb{N}$ no nontrivial homogeneous polynomial p of degree k vanishes identically on Ω. We will prove that $\mathcal{M}(\Omega)$ includes all homogeneous polynomials of degree k (and thus all polynomials of degree at most k). If the above holds for all $k \in \mathbb{Z}_+$ it then follows that $\mathcal{M}(\Omega)$ contains all polynomials and therefore $\overline{\mathcal{M}(\Omega)} = C(\mathbb{R}^n)$.

Let $\mathbf{a} \in \Omega$ and set $g(\langle \mathbf{a}, \mathbf{x} \rangle) = (\langle \mathbf{a}, \mathbf{x} \rangle)^k$, whence $(\langle \mathbf{a}, \mathbf{x} \rangle)^k \in \mathcal{M}(\Omega)$. Since $D^{\mathbf{m}_1} \mathbf{x}^{\mathbf{m}_2} = \delta_{\mathbf{m}_1, \mathbf{m}_2} k!$, for $\mathbf{m}_1, \mathbf{m}_2 \in \mathbb{Z}_+^n$, $|\mathbf{m}_1| = |\mathbf{m}_2| = k$, it easily follows that every linear functional ℓ on the finite dimensional linear space of homogeneous polynomials H_k^n of degree k may be represented by some $q \in H_k^n$ via

$$\ell(p) = q(D)p$$

for each $p \in H_k^n$.

For any given $q \in H_k^n$,

$$q(D)\, (\langle \mathbf{a}, \mathbf{x} \rangle)^k = k!\, q(\mathbf{a}).$$

If the linear functional ℓ annihilates $(\langle \mathbf{a}, \mathbf{x} \rangle)^k$ for all $\mathbf{a} \in \Omega$, then its representor $q \in H_k^n$ vanishes on Ω. By assumption this implies that $q = 0$. The fact that no nontrivial linear functional on H_k^n annihilates $(\langle \mathbf{a}, \mathbf{x} \rangle)^k$ for all $\mathbf{a} \in \Omega$ implies

$$H_k^n = \operatorname{span}\{(\langle \mathbf{a}, \mathbf{x} \rangle)^k : \mathbf{a} \in \Omega\}.$$

Thus $H_k^n \subseteq \mathcal{M}(\Omega)$. $\square$

Example 6.6. There are other results of the same general flavor as that found in Example 6.4. For example, assume $\|\cdot\|$ is the usual Euclidean norm on $\mathbb{R}^n$. Then we have from Pinkus [1996] the following two results.

Proposition 6.6. *For each* $\sigma \in C(\mathbb{R}_+)$

$$\operatorname{span}\{\sigma(\rho\|\cdot - \mathbf{a}\|) : \rho > 0, \mathbf{a} \in \mathbb{R}^n\}$$

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

Proposition 6.7. *For each* $\sigma \in C(\mathbb{R})$

$$\operatorname{span}\{\sigma\left(a \prod_{i=1}^{n} (\cdot - b_i)\right) : a, b_1, \ldots, b_n \in \mathbb{R}\}$$

is dense in $C(\mathbb{R}^n)$ *(uniform convergence on compacta) if and only if* σ *is not of the form*

$$\sigma(t) = \sum_{j=0}^{r} c_{0j} t^j + \sum_{j=1}^{r} \sum_{i=1}^{n-1} c_{ij} t^j (\ln |t|)^i$$

for some finite r *and coefficients* (c_{ij}).

Example 6.7. A very interesting result, with applications in Radon transform theory, is that conjectured by Lin and Pinkus and proved by Agranovsky, Quinto [1996]. It characterizes the set of centers of radial functions needed for density. The complete answer is only known in $\mathbb{R}^2$.

Theorem 6.8. *Let $\mathcal{A} \subseteq \mathbb{R}^2$. Then*

$$\operatorname{span}\{g(\|\cdot - \mathbf{a}\|) : g \in C(\mathbb{R}),\, \mathbf{a} \in \mathcal{A}\}$$

is not dense in $C(\mathbb{R}^2)$ (uniform convergence on compacta) if and only if $\mathcal{A}$ is composed of a finite number of points together with a subset of a set of straight lines having a common intersection point and where the angles between each of the lines is a rational multiple of π (a Coxeter system of lines).

Example 6.8. The functions $\|\cdot - \mathbf{a}\|^2$ are shifts of the polynomial $q(\mathbf{x}) = \sum_{i=1}^n x_i^2$. Assume we are given an arbitrary polynomial p. Under what exact conditions do we have that

$$\operatorname{span}\{g(p(\cdot - \mathbf{a})) : g \in C(\mathbb{R}),\, \mathbf{a} \in \mathbb{R}^n\}$$

is dense in $C(\mathbb{R}^n)$? This next result, as well as variations thereof, can be found in Pinkus, Wajnryb [1995].

Theorem 6.9. *Let p be an arbitrary polynomial in $\mathbb{R}^n$. Then for $n = 1, 2, 3$,*

$$\operatorname{span}\{g(p(\cdot - \mathbf{a})) : g \in C(\mathbb{R}),\, \mathbf{a} \in \mathbb{R}^n\}$$

is dense in $C(\mathbb{R}^n)$ (uniform convergence on compacta) if and only if

$$\operatorname{span}\{p(\cdot - \mathbf{a}) : \mathbf{a} \in \mathbb{R}^n\}$$

separates points.

By "separates points" we mean that for any given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\mathbf{x} \neq \mathbf{y}$, there exists a $\mathbf{a} \in \mathbb{R}^n$ for which

$$p(\mathbf{x} - \mathbf{a}) \neq p(\mathbf{y} - \mathbf{a}).$$

This condition is obviously necessary. The sufficiency is far from trivial. For $n = 4$ it is also sufficient if p is a homogeneous polynomial. However for $n \geq 4$ this condition is not always sufficient.

Example 6.9. *Müntz's Theorem.* The Müntz problem in the multivariate setting is significantly more difficult than in the univariate setting. Some sufficient conditions have been given, but the problem still remains very much open. The interested reader is urged to look at Bloom [1992] and Kroó [1994] and references therein.

References

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

Agranovsky, M. L. and E. T. Quinto [1996] Injectivity sets for the Radon transform over circles and complete systems of radial functions, *J. Funct. Anal.* **139**, 383–414.

Ahiezer, N. I. [1956] On the weighted approximation of continuous functions by polynomials on the entire number axis, *Uspehi Mat. Nauk* **11**, 3–43. Also appears in translation in English in *American Math. Soc. Transl., Series 2* **22**, 95–137, 1962.

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

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

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

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

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

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

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

Bernstein, S. [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, Vol. I*, (Cambridge, 22-28 August 1912), E. W. Hobson, A. E. H. Love, eds, Cambridge, England, 256–266. Also appears in Russian translation in Bernstein's Collected Works.

Bernstein, S. [1924] Le problème de l'approximation des fonctions continues sur tout l'axe réel et l'une de ses applications, *Bull. Soc. Math. France* **52**, 399–410.

Bloom, T. [1992] A multivariable version of the Müntz-Szász theorem. The Madison Symposium on Complex Analysis, (Madison, WI, 1991) , *Contemp. Math.* **137**, 85–92.

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

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

Boor, C. de [1968] On uniform approximation by splines, *J. Approx. Theory* **1**, 219–235.

Borwein, P. [1990] Zeros of Chebyshev polynomials in Markov systems, *J. Approx. Theory* **63**, 56–64.

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

Borwein, P. and T. Erdélyi [1995] Dense Markov spaces and unbounded Bernstein inequalities, *J. Approx. Theory* **81**, 66–77.

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

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

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

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

Dieudonné, J. [1937] Sur les fonctions continues numérique définies dans une produit de deux espaces compacts, *Comptes Rendus Acad. Sci. Paris* **205**, 593–595.

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

Fejér, L. [1900] Sur les fonctions bornées et intègrables, *Comptes Rendus Acad. Sci. Paris* **131**, 984–987.

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

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

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

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

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

Kroó, A. [1994] A geometric approach to the multivariate Müntz problem, *Proc. Amer. Math.Soc.* **121**, 199–208.

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

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

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

Leshno, M., 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.

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

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

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

Lin, V. Ya. and A. Pinkus [1993] Fundamentality of ridge functions, *J. Approx. Theory* **75**, 295–311.

Lorentz, G. G., M. v. Golitschek, and Y. Makovoz [1996] "Constructive Approximation. Advanced problems", Springer Verlag, Berlin.

Luxemburg, W. A. J. [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.

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

Mergelyan, S. N. [1956] Weighted approximation by polynomials, *Uspehi Mat. Nauk* **11**, 107–152. Also appears in translation in English in *American Math. Soc. Transl., Series 2* **10**, 59–106, 1958.

Mhaskar, H. N. [1996] "Introduction to the Theory of Weighted Polynomial Approximation", World Scientific, Singapore.

Müntz, C. H. [1914] Über den Approximationssatz von Weierstrass, in *Mathematische Abhandlungen Hermann Amandus Schwarz zu seinem fünfzigjährigen Doktojubiläum am 6. August 1914 gewidmet von Freunden und Schülern*, C. Carathéodory, G. Hessenberg, et al., eds, Springer, Berlin, 303–312.

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

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

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

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

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

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

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

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

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

Pinkus, A. [2005] The Weierstrass approximation theorems, *Surveys in Approx. Theory*, (to appear).

Pinkus, A. and B. Wajnryb [1995] A Problem of Approximation Using Multivariate Polynomials, *Uspekhi Mat. Nauk* **50**, 89–110. = *Russian Math. Surveys* **50** (1995), 319–340.

Prolla, J. B. [1977] "Approximation of Vector Valued Functions", North-Holland, Amsterdam.

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

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

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

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

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

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

Runge, C. [1885/86] Über die Darstellung willkürlicher Functionen, *Acta Math.* **7**, 387–392.

Schmidt, E. [1905] "Entwicklung willkürlicher Funktionen nach Systemen vorgeschriebener", dissertation, Göttingen (Germany).

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

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

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

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

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

Szász, O. [1916] Über die Approximation stetiger Funktionen durch lineare Aggregate von Potenzen , *Math. Ann.* **77**, 482–496.

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

Vostrecov, B. A. and M. A. Kreines [1961] Approximation of continuous functions by superpositions of plane waves, *Dokl. Akad. Nauk SSSR* **140**, 1237–1240. = *Soviet Math. Dokl.* **2** (1961), 1326–1329.

Weierstrass, K. [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.)

Weierstrass, K. [1886] Sur la possibilité d'une représentation analytique des fonctions dites arbitraires d'une variable réelle, *J. Math. Pure et 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*)

Allan Pinkus
Department of Mathematics
Technion
Haifa, 32000
Israel
`pinkus@tx.technion.ac.il`
`http://www.technion.ac.il/~pinkus/`