The Weierstrass Approximation Theorems

Allan Pinkus

2 November 2004

Abstract. This is a survey of the Weierstrass Approximation Theorems and their various proofs.

1 Introduction

This survey is about the Weierstrass Approximation Theorems. We consider these results within a historical context and also discuss in detail many of the subsequent proofs. This is a shorter version of the paper Pinkus [2000] with some alterations.

The Weierstrass Approximation Theorems are two theorems that Weierstrass (1815–1897) published in 1885 in Weierstrass [1885] when he was 70 years old. They prove the density of algebraic polynomials in the space of continuous real-valued functions on a finite interval in the uniform norm, and the density of trigonometric polynomials in the space of 2π-periodic continuous real-valued functions on $\mathbb{R}$ in the uniform norm. These theorems did not arise from nowhere. They were born within a historical context and it is of some interest to try to understand their origins and their impact.

It has been said that two main themes stand out in Weierstrass' work. The first is called the *arithmetization of analysis*. This was a program to separate the calculus from geometry and to provide it with a proper solid analytic foundation. Providing a logical basis for the real numbers, for functions and for calculus was a necessary stage in the development of analysis. Weierstrass was one of the leaders of this movement in his lectures and in his papers. He not only brought a new standard of rigour to his own mathematics, but attempted to do the same to much of mathematical analysis.

The second theme which is everpresent in Weierstrass' work is that of power series (and function series). Weierstrass is said to have stated that his own work

Surveys in Approximation Theory
Volume 1, 2005. pp. 1–37.

ISBN x-x-x

in analysis was "nothing but power series", see Bell [1936, p. 462]. In fact we will see how Weierstrass perceived his approximation theorems as theorems on convergent series. These approximation theorems were also a counterbalance to Weierstrass' famous example of a continuous nowhere differentiable function. It is a generally accepted fact that the existence of continuous nowhere differentiable functions was known and lectured upon by Weierstrass in 1861. The approximation theorems are in a sense its converse. Every continuous function on $\mathbb{R}$ is a limit not only of infinitely differentiable or even analytic functions, but in fact of polynomials. Furthermore, this limit is uniform if we restrict the approximation to any finite interval. Thus the set of continuous functions contains very, very non-smooth functions, but they can each be approximated arbitrarily well by the ultimate in smooth functions. It is this dichotomy which very much lies at the heart of approximation theory.

2 The Fundamental Theorems of Approximation Theory

In this section we review the contents of Weierstrass' [1885] and its variants. We first fix some notation. $C(\mathbb{R})$ will denote the class of continuous real-valued functions on all of $\mathbb{R}$, $C[a,b]$, $-\infty < a < b < \infty$, 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, and sometimes should, be considered as the restriction to $[a,b]$ of functions in $C(\mathbb{R})$ which are $(b-a)$-periodic.)

The paper stating and proving what we, in approximation theory, call "the" Weierstrass theorems, i.e., those that prove the density of algebraic polynomials in the space $C[a,b]$ (for every $-\infty < a < b < \infty$) and trigonometric polynomials in $\widetilde{C}[0,2\pi]$, is Weierstrass [1885]. It was published when Weierstrass was 70 years old!! This is one paper, but it appeared in two parts. It seems that the significance of the paper was immediately appreciated, as the paper appeared in translation (in French) one year later in Weierstrass [1886]. Again it was published in two parts under the same title (but in different issues, which is somewhat confusing). The paper was "reprinted" in Weierstrass' collected works (Mathematische Werke). It is contained in Volume 3 that appeared in 1903, although parts of Volume 3 including, it seems, this paper, were edited by Weierstrass himself a few years previously. Here the two parts do appear as one paper. In addition, some changes were made. A half page was added at the beginning, ten pages of material were appended to the end of the paper, and some other minor changes were made. We will return to these additions later.

Weierstrass had an abiding interest in complex function theory and in representing functions by power series. The results he obtained in this paper should definitely be viewed from that perspective. In fact the title of this paper emphasizes this viewpoint. The paper is titled *On the possibility of giving an analytic representation to an arbitrary function of a real variable.* In this section we review what Weierstrass did in this paper.

Weierstrass starts his original paper with the statement that if f is continuous and bounded on all of $\mathbb{R}$ then, as is known,

$$\lim_{k\to 0^+} \frac{1}{k\sqrt{\pi}} \int_{-\infty}^{\infty} f(u) e^{-(\frac{u-x}{k})^2} \, \mathrm{d}u = f(x).$$

He then immediately notes that this may be generalized to any kernel ψ that is continuous, nonnegative, integrable and even on $\mathbb{R}$. For such ψ he sets

$$F(x,k) = \frac{1}{2k\omega} \int_{-\infty}^{\infty} f(u) \psi\left(\frac{u-x}{k}\right) \mathrm{d}u,$$

where

$$\omega = \int_0^{\infty} \psi(x) \, \mathrm{d}x,$$

and proves that

$$\lim_{k\to 0^+} F(x,k) = f(x)$$

for each x. He not only proves pointwise convergence, but also *uniform convergence on any finite interval.* The proof is standard. We will not repeat it here. Weierstrass also notes that there are entire ψ, as above, for which $F(\cdot,k)$ is entire for every $k > 0$. He explicitly states that $\psi(x) = e^{-x^2}$ is an example thereof. The consequence of the above is the following.

Theorem A. *Let f be continuous and bounded on $\mathbb{R}$. Then there exists a sequence of entire functions $F(x,k)$ (as functions of x for each positive k) such that for each x*

$$\lim_{k\to 0^+} F(x,k) = f(x).$$

Weierstrass seems very much taken with this result that every bounded continuous function on $\mathbb{R}$ is a pointwise limit of entire functions. In fact he prefaces Theorem A with the statement that this theorem "strikes me as remarkable and fruitful". For unknown reasons this sentence, and only this sentence, was deleted from the paper when it was reprinted in Weierstrass' Mathematische Werke.

As mentioned, on any finite interval, one may obtain uniform convergence. Furthermore, since $F(\cdot,k)$ is entire, the truncated power series of $F(\cdot,k)$ uniformly converges to $F(\cdot,k)$ on any finite interval. Each of the above statements is easily proved and gives:

Theorem B. *Let f be continuous and bounded on $\mathbb{R}$. Given a finite interval $[a,b]$ and an $\varepsilon > 0$, there exists an algebraic polynomial p for which*

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

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

Throughout the first part of Weierstrass [1885] and for much of the second part, Weierstrass is concerned with functions defined on all of $\mathbb{R}$. However later in the second part he does note that given any $f \in C[a,b]$, $-\infty < a < b < \infty$, we can define f to equal $f(a)$ on $(-\infty, a)$, and to equal $f(b)$ on (b, ∞). We can then apply the above Theorem B to obtain what is technically never explicitly stated, but nonetheless very implicitly stated, and what is today considered as the main result of this paper.

Fundamental Theorem of Approximation Theory. *Let $f \in C[a,b]$ where $-\infty < a < b < \infty$. Given $\varepsilon > 0$, there exists an algebraic polynomial p for which*

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

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

Returning to Weierstrass [1885], and bounded $f \in C(\mathbb{R})$, Weierstrass considers two sequences of positive values $\{c_n\}$ and $\{\varepsilon_n\}$, for which $\lim_{n\to\infty} c_n = \infty$, and $\sum_{n=1}^{\infty} \varepsilon_n < \infty$. From Theorem B it follows that for f as above there exists a polynomial p_n such that

$$|f(x) - p_n(x)| < \varepsilon_n$$

on $[-c_n, c_n]$.

Set $q_0 = p_1$ and $q_m = p_{m+1} - p_m$, $m = 1, 2, \ldots$. Then

$$\sum_{m=0}^{n} q_m(x) = p_{n+1}(x)$$

and, thus, in a pointwise sense

$$f(x) = \sum_{m=0}^{\infty} q_m(x). \tag{2.1}$$

Furthermore, let $[a,b]$ be a finite interval. Then for all m sufficiently large

$$|f(x) - p_m(x)| < \varepsilon_m$$

for all $x \in [a,b]$, implying also

$$|q_m(x)| < \varepsilon_m + \varepsilon_{m+1}$$

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

$$\sum_{m=M}^{\infty} |q_m(x)| < 2 \sum_{m=M}^{\infty} \varepsilon_m$$

for all $x \in [a,b]$ and the series

$$\sum_{m=0}^{\infty} q_m(x)$$

therefore converges absolutely and uniformly to f on $[a,b]$. This Weierstrass states as Theorem C. That is,

Theorem C. *Let f be continuous and bounded on $\mathbb{R}$. Then f may be represented, in many ways, by an infinite series of polynomials. This series converges absolutely for every value of x, and uniformly in every finite interval.*

Weierstrass and subsequent authors would often phrase or rephrase these approximation or density results (in this case Theorem B) in terms of infinite series. It was only many years later that this equivalent form went out of fashion. In fact such a phrasing was at the time significant. One should also recall that it was only a few years earlier that du Bois-Reymond had constructed a continuous function whose Fourier series diverged at a point, see du Bois-Reymond [1876]. Weierstrass' theorem was considered by many, including Weierstrass himself, to be a "representation theorem". The theorem was seen as a means of reconciling the "analytic" and "synthetic" viewpoints that had divided late 19th century mathematics, see Gray [1984] and also Siegmund-Schultze [1988]. Much of the remaining parts of Weierstrass [1885] is concerned with the construction (in some sense) of a good polynomial approximant or a good representation for f (as in (2.1)). Weierstrass was well aware that he could not possibly construct a good power series representation for f, but he did find, in some sense, a reasonable expansion of f in terms of Legendre polynomials.

In the latter part of Weierstrass [1885], Weierstrass proves the density of trigonometric polynomials in $\widetilde{C}[0, 2\pi]$. His proof is interesting and proceeds as follows using complex function theory.

Let ψ be an entire function that is nonnegative, integrable and even on $\mathbb{R}$ and has the following property. Given an $f \in \widetilde{C}[0, 2\pi]$, the functions

$$F(z,k) = \frac{1}{2k\omega} \int_{-\infty}^{\infty} f(u)\psi\left(\frac{u-z}{k}\right) \mathrm{d}u,$$

where

$$\omega = \int_0^{\infty} \psi(x)\,\mathrm{d}x,$$

are entire for each $k > 0$ (as a function of $z \in \mathbb{C}$) and satisfy

$$\lim_{k \to 0^+} F(x,k) = f(x)$$

uniformly on $[0, 2\pi]$. Weierstrass notes that such functions ψ exist, e.g., $\psi(u) = e^{-u^2}$.

Since f is 2π-periodic so is F, i.e.,

$$F(z + 2\pi, k) = F(z,k)$$

for all $z \in \mathbb{C}$ and $k > 0$. For each fixed $k > 0$, set

$$G(z,k) = F(\frac{\log z}{i}, k).$$

In general, since $\log z$ is a multiple-valued function, G would also be a multiple-valued function. However from the 2π-periodicity of F, it follows that G is single-valued and thus is an analytic function on $\mathbb{C}\backslash\{0\}$. Consequently, G has a Laurent series expansion of the form

$$G(z,k) = \sum_{n=-\infty}^{\infty} c_{n,k} z^n$$

which converges absolutely and uniformly to G on every domain bounded away from 0 and ∞. We will consider this expansion on the unit circle $|z| = 1$. Setting $z = e^{ix}$, it follows that

$$F(x,k) = \sum_{n=-\infty}^{\infty} c_{n,k} e^{inx}$$

where the series converges absolutely and uniformly to $F(x,k)$ for all real x. (In fact, it may be shown that if $\psi(u) = e^{-u^2}$, then $c_{n,k} = c_n e^{-n^2k^2/4}$, where the $\{c_n\}$ are the Fourier coefficients of f.) In other words, Weierstrass has given a proof of the fact that for $F(x,k)$ 2π-periodic and entire, its Fourier series converges absolutely and uniformly to $F(x,k)$ on $\mathbb{R}$. We now truncate this series to get an arbitrarily good approximant to $F(x,k)$ which itself, by a suitable choice of k, was an arbitrary good approximant to f. The truncated series is a trigonometric polynomial. This completes Weierstrass' proof, the result of which we now formally state.

Second Fundamental Theorem of Approximation Theory. *Let $f \in \widetilde{C}[0,2\pi]$. Given $\varepsilon > 0$, there exists a trigonometric polynomial t for which*

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

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

As we stated at the beginning of this section, when Weierstrass [1885] was reprinted in Weierstrass' Mathematische Werke there were two notable additions. These are of interest and worth mentioning. We recall that while this reprint appeared in 1903 there is reason to assume that Weierstrass himself edited this paper.

The first addition was a short (half page) "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 $x \in I\!R$. In addition the convergence of the above sum is uniform on every finite interval.

We can assume that this is the emphasis which Weierstrass wished to give his paper. It is a repeat of Theorem C (although the boundedness condition on f seems to have been overlooked) and curiously without mention of the fact that the f_i may be assumed to be polynomials.

The second addition is 10 pages appended to the end of the paper. In these 10 pages Weierstrass shows how to extend the results of this paper (or, to be more precise, the results concerning algebraic polynomials) to approximating continuous functions of several variables. He does this by setting $F(x_1, \ldots, x_n, k)$ equal to

$$\frac{1}{2^n k^n \omega^n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(u_1, \ldots, u_n) \psi(\frac{u_1 - x_1}{k}) \cdots \psi(\frac{u_n - x_n}{k}) \, \mathrm{d}u_1 \cdots \mathrm{d}u_n$$

and then essentially mimicking the proofs of Theorems A and B. However Picard [1891a] published already in 1891 an alternative proof of Weierstrass' theorems and showed how to extend the results to functions of several variables. As such, Weierstrass' priority to this result is somewhat in question.

3 Additional Proofs of the Fundamental Theorems

In this section we present various alternative proofs of Weierstrass' theorems on the density of algebraic and trigonometric polynomials on finite intervals in $I\!R$. We believe that the echo of these proofs have an abiding value. Some of the papers we cite contain additional results or emphasize other points of view. We ignore such digressions. The proofs we present divide roughly into three groups. The first group contains proofs that, in one form or another, are based on singular integrals. The proofs of Weierstrass, Picard, Fejér, Landau, and de la Vallée Poussin belong here. The second group of proofs is based on the idea of approximating a particular function. In this group we find the proofs of Runge/Phragmén, Lebesgue, Mittag-Leffler, and Lerch. Finally, there is the third group that contain the proofs which do not quite belong to either of the above groups. Here we find proofs due to Lerch, Volterra and Bernstein. These are what we term the "early proofs". They all appeared prior to 1913. Note the pantheon of names that were drawn to this theorem. The main focus of these proofs are the Weierstrass theorems themselves rather than any far-reaching generalizations thereof. There are later proofs coming from different and broader formulations. However we discuss only one of these later proofs. It is that due to Kuhn which we consider to be wonderfully elegant and simple. For historical consistency we have chosen to present these proofs in more or less chronological order. This lengthens the paper, but we hope the advantages of this approach offset the deficiencies.

We start by formally stating certain facts which will be obvious to most readers, but perhaps not to everyone. The first two statements follow from a change of variables, and are stated without proof.

Proposition 1. *Algebraic polynomials are dense in $C[a,b]$ iff they are dense in $C[0,1]$.*

Analogously we have the less used:

Proposition 2. *The trigonometric polynomials*

$$\mathrm{span}\{1, \sin x, \cos x, \sin 2x, \cos 2x, \ldots\}$$

are dense in $\widetilde{C}[0,2\pi]$ iff

$$\mathrm{span}\{1, \sin \frac{2\pi x}{b-a}, \cos \frac{2\pi x}{b-a}, \sin 2\frac{2\pi x}{b-a}, \cos 2\frac{2\pi x}{b-a}, \ldots\}$$

are dense in $\widetilde{C}[a,b]$.

We now show that the density of algebraic polynomials in $C[a,b]$, and trigonometric polynomials in $\widetilde{C}[0,2\pi]$, are in fact equivalent statements. That is, we prove that each of the fundamental theorems follows from the other, see also Natanson [1964, p. 16–19].

Proposition 3. *If trigonometric polynomials are dense in $\widetilde{C}[0,2\pi]$, then algebraic polynomials are dense in $C[a,b]$.*

Proof: We present two proofs of this result. The first proof may be found in Picard [1891a].

Assume, without loss of generality, that $0 \le a < b < 2\pi$. Extend $f \in C[a,b]$ to some $\widetilde{f} \in \widetilde{C}[0,2\pi]$. Since trigonometric polynomials are dense in $\widetilde{C}[0,2\pi]$, there exists a trigonometric polynomial t that is arbitrarily close to $\widetilde{f}$ on $[0,2\pi]$, and thus to f on $[a,b]$. Every trigonometric polynomial is a finite linear combination of $\sin nx$ and $\cos nx$. As such each is an entire function. Thus t is an entire function having an absolutely and uniformly convergent power series expansion. By suitably truncating this power series we obtain an algebraic polynomial that is arbitrarily close to t, and thus ultimately to f.

A slight variant on the above bypasses the need to extend f to $\widetilde{f}$. Assume $f \in C[0,2\pi]$, and define

$$g(x) = f(x) + \frac{f(0) - f(2\pi)}{2\pi} x.$$

Then $g \in \widetilde{C}[0,2\pi]$. We now apply the reasoning of the previous paragraph to obtain an algebraic polynomial p arbitrarily close to g on $[0,2\pi]$, whence it follows that

$$p(x) - \frac{f(0) - f(2\pi)}{2\pi} x$$

is arbitrarily close to f on $[0,2\pi]$.

A different and more commonly quoted proof is the following which does not depend upon the truncation of a power series. According to de la Vallée Poussin [1918], [1919], the idea in this proof is due to Bernstein.

Given $f \in C[-1,1]$, set

$$g(\theta) = f(\cos\theta), \qquad -\pi \le \theta \le \pi.$$

Then $g \in \widetilde{C}[-\pi,\pi]$ and g is even. As such given $\varepsilon > 0$ there exists a trigonometric polynomial t for which

$$|g(\theta) - t(\theta)| < \varepsilon$$

for all $\theta \in [-\pi,\pi]$. We divide t into its even and odd parts, i.e.,

$$t_e(\theta) = \frac{t(\theta) + t(-\theta)}{2}$$

$$t_o(\theta) = \frac{t(\theta) - t(-\theta)}{2}$$

and note that t_e and t_o are also trigonometric polynomials. (Equivalently, t_e is composed of the cosine terms of t, while t_o is composed of the sine terms of t.)

Since g is even we have

$$\max\{|(g-t)(\theta)|, |(g-t)(-\theta)|\}$$

$$= \max\{|(g-t_e)(\theta) - t_o(\theta)|, |(g-t_e)(\theta) + t_o(\theta)|\} \ge |(g-t_e)(\theta)|,$$

and, thus,

$$|g(\theta) - t_e(\theta)| < \varepsilon$$

for all $\theta \in [-\pi,\pi]$. In other words, since g is even we may assume that t is even.

Let

$$t(\theta) = \sum_{m=0}^{n} a_m \cos m\theta.$$

Each $\cos m\theta$ is a polynomial of exact degree m in $\cos\theta$. In fact

$$\cos m\theta = T_m(\cos\theta)$$

where the T_m are the Chebyshev polynomials (see e.g., Rivlin [1974]). Setting

$$p(x) = \sum_{m=0}^{n} a_m T_m(x),$$

we have

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

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

Proposition 4. *If algebraic polynomials are dense in* $C[a,b]$*, then trigonometric polynomials are dense in* $\widetilde{C}[0,2\pi]$.

Proof: The first proof of this fact was the one given by Weierstrass in Section 2. To our surprise (and chagrin) we have essentially found only one other proof of this result, and it is not simple. The proof we give here is de la Vallée Poussin's [1918], [1919] variation on a proof in Lebesgue [1898].

Let $f \in \widetilde{C}[0,2\pi]$ and consider f as being defined on all of $\mathbb{R}$. Set

$$g(\theta) = \frac{f(\theta)+f(-\theta)}{2}$$

and

$$h(\theta) = \frac{f(\theta)-f(-\theta)}{2}\sin\theta.$$

Both g and h are continuous even functions of period 2π.

Define

$$\phi(x) = g(\arccos x), \qquad \psi(x) = h(\arccos x).$$

These are well-defined functions in $C[-1,1]$. Thus, given $\varepsilon > 0$ there exist algebraic polynomials p and q for which

$$|\phi(x) - p(x)| < \frac{\varepsilon}{4}, \qquad |\psi(x) - q(x)| < \frac{\varepsilon}{4}$$

for all $x \in [-1,1]$. As g and h are even, it follows that

$$|g(\theta) - p(\cos\theta)| < \frac{\varepsilon}{4}, \qquad |h(\theta) - q(\cos\theta)| < \frac{\varepsilon}{4}$$

for all θ. From the definition of g and h, we obtain

$$\big|f(\theta)\sin^2\theta - \big[p(\cos\theta)\sin^2\theta + q(\cos\theta)\sin\theta\big]\big| < \frac{\varepsilon}{2}$$

for all θ.

We apply this same analysis to the function $f(\theta+\pi/2)$ to obtain algebraic polynomials r and s for which

$$\Big|f(\theta+\frac{\pi}{2})\sin^2\theta - \big[r(\cos\theta)\sin^2\theta + s(\cos\theta)\sin\theta\big]\Big| < \frac{\varepsilon}{2}$$

for all θ. Substituting for $\theta + \pi/2$ gives

$$\big|f(\theta)\cos^2\theta - \big[r(\sin\theta)\cos^2\theta - s(\sin\theta)\cos\theta\big]\big| < \frac{\varepsilon}{2}.$$

Thus the trigonometric polynomial

$$p(\cos\theta)\sin^2\theta + q(\cos\theta)\sin\theta + r(\sin\theta)\cos^2\theta - s(\sin\theta)\cos\theta$$

is an ε-approximant to f. □

After these preliminaries we can now look at the inherent methods and ideas used in various alternative proofs of either of the two Weierstrass fundamental theorems of approximation theory. We present these proofs in more or less the order in which they appeared in print.

Picard. Émile Picard (1856–1941) (Hermite's son-in-law) had an abiding interest in the Weierstrass' theorems and in Picard [1891a] gave the first in a series of different proofs of the Weierstrass theorems. This proof also appears in Picard's famous textbook [1891b]. Later editions of this textbook expanded upon this, often including other methods of proof, but not always with complete references. Picard's proof, like that of Weierstrass, is based on a smoothing procedure using singular integrals. Picard, however, chose to use the Poisson integral. His proof proceeds as follows.

Assume $f \in \widetilde{C}[0, 2\pi]$. As f is continuous and 2π-periodic on $\mathbb{R}$, it is uniformly continuous thereon. As such, given $\varepsilon > 0$ there exists a $\delta > 0$ such that for $|x - \theta| < \delta$ we have $|f(x) - f(\theta)| < \varepsilon$. Let

$$P(r, \theta) = \frac{1}{2\pi}\int_0^{2\pi} \frac{1 - r^2}{1 - 2r\cos(x - \theta) + r^2} f(x)\,\mathrm{d}x$$

denote the Poisson integral of f.

We claim that, with the above notation,

$$|P(r, \theta) - f(\theta)| < \varepsilon + \frac{\|f\|(1 - r^2)}{r(1 - \cos\delta)}$$

for all θ. This may be explicitly proven as follows.

$$\begin{aligned} P(r, \theta) - f(\theta) &= \frac{1}{2\pi}\int_0^{2\pi} \frac{1 - r^2}{1 - 2r\cos(x - \theta) + r^2}[f(x) - f(\theta)]\,\mathrm{d}x \\ &= \frac{1}{2\pi}\int_{|x-\theta|<\delta} \frac{1 - r^2}{1 - 2r\cos(x - \theta) + r^2}[f(x) - f(\theta)]\,\mathrm{d}x \\ &\quad + \frac{1}{2\pi}\int_{\delta\le|x-\theta|\le\pi} \frac{1 - r^2}{1 - 2r\cos(x - \theta) + r^2}[f(x) - f(\theta)]\,\mathrm{d}x. \end{aligned}$$

Now

$$\frac{1}{2\pi}\int_{|x-\theta|<\delta} \frac{1 - r^2}{1 - 2r\cos(x - \theta) + r^2}|f(x) - f(\theta)|\,\mathrm{d}x$$

$$< \frac{\varepsilon}{2\pi} \int_0^{2\pi} \frac{1-r^2}{1-2r\cos(x-\theta)+r^2}\,\mathrm{d}x = \varepsilon.$$

In addition

$$\frac{1}{2\pi} \int_{\delta \le |x-\theta| \le \pi} \frac{1-r^2}{1-2r\cos(x-\theta)+r^2} |f(x)-f(\theta)|\,\mathrm{d}x$$

$$\le 2\|f\| \frac{1}{2\pi} \int_{\delta \le |x-\theta| \le \pi} \frac{1-r^2}{1-2r\cos(x-\theta)+r^2}\,\mathrm{d}x \le \frac{\|f\|(1-r^2)}{r(1-\cos\delta)}.$$

This last inequality is a consequence of

$$1-2r\cos(x-\theta)+r^2 \ge 2r-2r\cos\delta = 2r(1-\cos\delta)$$

which holds for all x, θ satisfying $\delta \le |x-\theta| \le \pi$.

As a function of r,

$$\frac{\|f\|(1-r^2)}{r(1-\cos\delta)}$$

decreases to zero as r increases to 1. Choose some $r_1 < 1$ for which

$$\frac{\|f\|(1-r_1^2)}{r_1(1-\cos\delta)} < \varepsilon.$$

Thus

$$|f(\theta) - P(r_1,\theta)| < 2\varepsilon$$

for all θ.

Let

$$a_0/2 + \sum_{n=1}^{\infty} [a_n \cos nx + b_n \sin nx]$$

denote the Fourier series of f. Recall that the Fourier series of $P(r,\theta)$ is given by

$$a_0/2 + \sum_{n=1}^{\infty} r^n \left[a_n \cos nx + b_n \sin nx\right].$$

Since the a_n and b_n are uniformly bounded, the above Fourier series converges absolutely, and uniformly converges to $P(r,\theta)$ for each $r < 1$. Thus there exists an m for which

$$\left| P(r_1,\theta) - \left[a_0/2 + \sum_{n=1}^{m} r_1^n (a_n \cos nx + b_n \sin nx) \right] \right| < \varepsilon$$

for all θ. Set

$$g(\theta) = a_0/2 + \sum_{n=1}^{m} r_1^n (a_n \cos nx + b_n \sin nx).$$

We have "constructed" a trigonometric polynomial satisfying

$$|f(\theta) - g(\theta)| < 3\varepsilon$$

for all θ. In other words we have proven that in the uniform norm, trigonometric polynomials are dense in the space of continuous 2π-periodic functions.

As noted in the proof of Proposition 3, Picard then proves the Weierstrass theorem for algebraic polynomials based on the above result. Picard ends his paper by noting that the same procedure can be used to obtain parallel results for continuous functions of many variables. He was the first to publish an extension of the Weierstrass theorems to several variables.

As Picard [1891a] states, this proof is based on an inequality obtained by H. A. Schwarz in his well-known paper Schwarz [1871]. In fact, as Cakon [1987] points out, almost the entire Picard proof can be found in Schwarz [1871]. What is perhaps surprising is that Weierstrass did not notice this connection.

Lerch I. M. Lerch (1860–1922) was a Czech mathematician of some renown (see Skrasek [1960] and MacTutor [2004]) who attended some of Weierstrass' lectures. Lerch wrote two papers, Lerch [1892] and Lerch [1903], that included proofs of the Weierstrass theorem for algebraic polynomials. Unfortunately the paper Lerch [1892] is in Czech, difficult to procure, and I have found no reference to it anywhere in the literature except in Lerch [1903] and in a footnote in Borel [1905] (but Borel did not see the paper). Subsequent authors mentioned in this work were seemingly totally ignorant of this paper. Many of these authors quote Volterra [1897], although Lerch [1892] contains a similar proof with the same ideas. It is for the reader to decide whether, in these circumstances, Lerch deserves prominence or only precedence.

We here explain the proof as is essentially contained in Lerch [1892]. We defer the discussion of Lerch [1903] to a more appropriate place. Let $f \in C[a, b]$. Since f is uniformly continuous on $[a, b]$, it can be uniformly approximated thereon by a polygonal (piecewise linear) line. Lerch notes that every polygonal line g may be uniformly approximated by a Fourier cosine series of the form

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos \frac{x-a}{b-a} n\pi,$$

where

$$a_n = \frac{2}{b-a} \int_a^b g(x) \cos \frac{x-a}{b-a} n\pi \, \mathrm{d}x.$$

It was, at the time, well-known to any mathematician worth his salt that the Fourier cosine series of a continuous function with a finite number of maxima and minima uniformly converges to the function. This result goes back to Dirichlet [1829], see e.g. Sz.-Nagy [1965, p. 399]. Alternatively it is today a standard result contained in every Fourier series text that if the derivative of a continuous function is piecewise continuous with one-sided derivatives at each point, then its Fourier cosine series converges uniformly. Both these results follow from the analogous results for periodic functions and the usual Fourier series. Both these results hold for our polygonal line. As this Fourier cosine series converges uniformly to our polygonal line we may truncate it to obtain a trigonometric polynomial (but not a trigonometric polynomial as in Proposition 2) which approximates our polygonal line arbitrarily well. Finally, as the trigonometric polynomial is an entire function we can suitably truncate its power series expansion to obtain our desired algebraic polynomial approximant.

Volterra. The next published proof of Weierstrass' theorems is due to Volterra [1897]. V. Volterra (1860–1940) proved only the density of trigonometric polynomials in $\widetilde{C}[0, 2\pi]$. As he was aware of Picard [1891a], this should not detract from his proof.

Volterra was unaware of Lerch [1892], but his proof is much the same. Let $f \in \widetilde{C}[0, 2\pi]$. Since f is continuous on a closed interval, it is also uniformly continuous thereon. As such, it is possible to find a polygonal line that approximates f arbitrarily well. One can also assume that the polygonal line is 2π-periodic. It thus suffices to prove that one can arbitrarily well approximate any continuous, 2π-periodic, polygonal line by trigonometric polynomials. As stated in the proof of Lerch, the Fourier series of the polygonal line uniformly converges to the function. We now suitably truncate the Fourier series to obtain the desired approximation.

C. Runge (1856–1927), E. Phragmén (1863–1937), H. Lebesgue (1875–1941) and G. Mittag-Leffler (1846–1927) all contributed proofs of the Weierstrass approximation theorems, and their proofs are related both in character and idea. What did each do?

Mittag-Leffler, in 1900, was the last of the above four to publish on this subject. However he seems to have been the first to point out, in print, Runge and Phragmén's contributions. As such we start this story with Mittag-Leffler. The paper Mittag-Leffler [1900] is an "extract from a letter to E. Picard". This was, at the time, a not uncommon format for an article. Journals were still in their infancy, but were replacing correspondence as the primary mode of dissemination of mathematical research. Thus this combination of these two forms. The article came in response to what Picard had written in his "Lectures on Mathematics" given at the Decennial Celebration at Clark University, Picard [1899]. In this grand review Picard mentions the importance, in the development of the understanding of functions, of Weierstrass' example of a continuous nowhere differentiable function, and of Weierstrass' theorem on the representation of every continuous function on a finite interval as an absolutely and uniformly convergent series of polynomials. Picard then goes on to mention his own proof and that of Volterra [1897]. Mittag-Leffler [1900] points out that Weierstrass' theorem also follows from work of Runge [1885, 1885/86] although, as he notes, it is not explicitly contained anywhere in either of these two papers. He then explains his own proof, to which we shall return later. How did Mittag-Leffler know about Weierstrass' theorem following from the work of Runge? Firstly, Mittag-Leffler was the editor of Acta Mathematica and, as he writes, he was the one who published Runge's paper. (Mittag-Leffler founded Acta Mathematica in 1882 and was its editor for 45 years.) Moreover in the paper of Mittag-Leffler [1900] there is a very interesting long footnote which seems to have been somewhat overlooked. It starts as follows: *I found on this subject among my papers an article of Phragmén, from the year 1886, which goes thus.* What follows is two pages where Phragmén (who was 23 years old at the time) explains how Weierstrass' theorem can follow from Runge's work, Phragmén's simplification thereof, and also how to get from this the Weierstrass theorem on the density of trigonometric polynomials in $\widetilde{C}[0, 2\pi]$ (with some not insignificant additional work). Before we explain this in detail, let us start with the general idea behind these various proofs.

Let $f \in C[0, 1]$. Since f is continuous on a closed interval, it is also uniformly continuous thereon. As Lerch and Volterra pointed out, it is thus possible to find a polygonal line g (which today we might also call a spline of degree 1 with simple knots) that approximates f uniformly to within any given $\varepsilon > 0$, i.e., for which

$$|f(x) - g(x)| < \varepsilon,$$

for all $x \in [0, 1]$. This polygonal line is the first idea in these proofs. The second idea is to show that there is an arbitrarily good polynomial approximant to the relatively "simpler" g. This will then suffice to prove that we can find a polynomial that approximates our original f arbitrarily well. The third and more fundamental idea is to reduce the problem of finding a good polynomial approximant to g (which depends upon f) to that of finding a good polynomial approximant to one and only one function, independent of f. Each of Runge, Mittag-Leffler and Lebesgue do this in a different way.

Runge/Phragmén. We first fix some notation. Let $0 = x_0 < x_1 < \cdots < x_m = 1$ be the abscissae (knots) of the polygonal line g. There are various ways of writing g. One elementary way is:

$$g(x) = g_1(x) + \sum_{i=1}^{m-1} [g_{i+1}(x) - g_i(x)]\, h(x - x_i) \tag{3.1}$$

where g_i is the linear polynomial agreeing with g on $[x_{i-1}, x_i]$ and

$$h(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$

g_i may be explicitly given as

$$g_i(x) = y_{i-1} + \left(\frac{x - x_{i-1}}{x_i - x_{i-1}} \right) (y_i - y_{i-1})$$

where $y_j = g(x_j)$, $j = 0, 1, \ldots, m$.

What Runge did in his 1885/86 paper is the following. He considered the function

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

which has the property that

$$\lim_{n\to\infty} \phi_n(x) = \begin{cases} 1, & |x| < 1 \\ 1/2, & |x| = 1 \\ 0, & |x| > 1 \end{cases}.$$

Set $\psi_n(x) = 1 - \phi_n(1 + x)$. Then restricted to $[-1, 1]$ we have

$$\lim_{n\to\infty} \psi_n(x) = \begin{cases} 1, & 0 < x < 1 \\ 1/2, & x = 0 \\ 0, & -1 < x < 0 \end{cases}.$$

Since each ψ_n is increasing on $[-1, 1]$, and $\psi_{n+1}(x) > \psi_n(x)$ for $x \in (0, 1]$, while $\psi_{n+1}(x) < \psi_n(x)$ for $x \in (-1, 0)$, it follows that given any $\delta > 0$, small, the functions ψ_n are bounded on $[-1, 1]$ and uniformly converge to the function h on $[-1, -\delta] \bigcup [\delta, 1]$ for any given δ.

Since the linear polynomial $g_{i+1} - g_i$ vanishes at x_i, a short calculation verifies that for each $x_i \in (0, 1)$

$$[g_{i+1}(x) - g_i(x)]\, \psi_n(x - x_i)$$

uniformly converges to

$$[g_{i+1}(x) - g_i(x)]\, h(x - x_i)$$

on $[0, 1]$. Replacing the h in (3.1) by ψ_n we obtain a series of functions which uniformly approximate g.

These functions

$$\Psi_n(x) = g_1(x) + \sum_{i=1}^{m-1} [g_{i+1}(x) - g_i(x)]\, \psi_n(x - x_i)$$

are not polynomials or entire functions. But they are rational functions. Thus any continuous function on a finite real interval can be uniformly approximated by rational functions. This is the main result of Runge [1885/86]. It was published the same year as Weierstrass' paper.

Runge also discussed what could be said in the case of continuous functions on all of $\mathbb{R}$. In that context he noted that from one of his results in Runge [1885] one could always replace Ψ_n by another rational function, real on $\mathbb{R}$, with exactly two conjugate poles.

Phragmén in the above-mentioned footnote in Mittag-Leffler [1900] (but according to Mittag-Leffler written in 1886), remarks that apparently Runge overlooked in Runge [1885/86] (or did not think important) the fact that he could replace rational functions by polynomials. Runge quite explicitly had the tools to do this from Runge [1885].

What is the relevant result from Runge [1885]? It is the following, which we state in an elementary form. Assume D is a compact set and $\mathbb{C}\backslash D$ is connected. Let R be a rational function with poles outside D. Then given any point $w \in \mathbb{C}\backslash D$ there are rational functions, with only the one pole w, that approximate R arbitrarily well on D. This is not a difficult result to prove. Here, essentially, is Runge's proof. The rational function R can be decomposed as $R = \sum_{j=1}^{n} R_j$ where each R_j is a rational function with only one pole w_j. We now show how to move each w_j to w in a series of finite steps. For each j we choose $a_0, \ldots, a_m$, where $a_0 = w_j$ and $a_m = w$, and the a_i are chosen so that

$$|a_{i-1} - a_i| < |z - a_i|, \qquad i = 1, \ldots, m$$

for all $z \in D$. This can be done. At each stage we will construct a rational function G_i ($G_0 = R_j$) with only the simple pole a_i, and such that G_i is arbitrarily close to G_{i-1}. This follows from the fact that for given $k \in \mathbb{N}$ the function

$$\frac{1}{(z - a_{i-1})^k}$$

can be arbitrarily well approximated on D by

$$\left[\frac{1}{(z - a_{i-1})}\left[1 - \left(\frac{a_{i-1} - a_i}{z - a_i}\right)^n\right]\right]^k$$

by taking n sufficiently large. Note that the latter is a rational function with a pole only at a_i. Runge further noted that by a linear fractional transformation (and a bit of care) the pole could be shifted to ∞, whence the rational function becomes a polynomial. As Phragmén points out, if the function f to be approximated on $[0, 1]$ is real, we can replace the polynomial approximant G obtained above by $Re\, G$ on $[0, 1]$ which is also a polynomial and which better approximates f thereon. Thus Weierstrass' theorem is proved.

Phragmén also notes that it is really not necessary to use the results of Runge [1885]. If we go back to Runge [1885/86] and consider his construction therein, we see that each of the rational approximants are real on $[0, 1]$, and have denominator $1 + (1 + x)^{2n}$ for some n. Any such R may be decomposed as

$$R = g + r_1 + r_2$$

where g is a polynomial, r_1 is a rational function, all of whose poles lie in the upper half-plane, and $r_2(z) = \overline{r_1(\overline{z})}$ is a rational function, all of whose poles are conjugate to the poles of r_1 and lie in the lower half-plane. It is possible to choose a point z_1 in the lower half plane such that there exists a circle centered at z_1 containing $[0,1]$, but not containing any poles of r_1. As such the Taylor series of r_1 about z_1 converges uniformly to r_1 in $[0,1]$. Truncate it to obtain a polynomial p_1 that approximates r_1 arbitrarily well on $[0,1]$. It follows that $p_2(z) = \overline{p_1(\overline{z})}$ has the corresponding property with respect to r_2. As such

$$P = g + p_1 + p_2$$

is a real polynomial that can be chosen to approximate f arbitrarily well.

Another simple option, not mentioned by Phragmén, is simply to use the result of Runge [1885], to move the poles of any rational approximant away from $[0,1]$ so that a circle can be put about $[0,1]$ which does not contain any poles, and then use the truncated power series as above. Phragmén's proof of the density of trigonometric polynomials in $\widetilde{C}[0,2\pi]$ is more complicated and we will not present it here.

In any case, as we have seen, the algebraic Weierstrass theorem is a fairly simple consequence of Runge's [1885] and [1885/86] results. It is unfortunate and somewhat astonishing that Runge did not think of it.

Lebesgue. Let us now give Lebesgue's proof of Weierstrass' theorem as found in Lebesgue [1898]. This is one of the more elegant and cited proofs of Weierstrass' theorem. It is interesting to note that this was Lebesgue's first published paper. He was, at the time of publication, a 23 year old student at the École Normale Supérieure. He obtained his doctorate in 1902.

A more "modern" form of writing the g of (3.1) is as a spline. That is,

$$g(x) = ax + b + \sum_{i=1}^{m-1} c_i(x - x_i)_+^1$$

where

$$x_+^1 = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

and $ax + b = g_1(x)$. (This easily follows from the form (3.1). As $g_{i+1}(x) - g_i(x)$ is a linear polynomial that vanishes at x_i, it is necessarily of the form $c_i(x - x_i)$ for some constant c_i.) Since

$$2x_+^1 = |x| + x$$

the above form of g may also be rewritten as

$$g(x) = Ax + B + \sum_{i=1}^{m-1} C_i|x - x_i| \tag{3.2}$$

for some real constants A, B, and C_i.

Lebesgue [1898] considers the form (3.2) of g, and argues as follows. To approximate g arbitrarily well by a polynomial it suffices to be able to approximate $|x|$ arbitrarily well by a polynomial in $[-1, 1]$ (or in fact in any neighbourhood of the origin). If for given $\eta > 0$ there exists a polynomial p satisfying

$$||x| - p(x)| < \eta$$

for all $x \in [-1, 1]$, then

$$||x - x_i| - p(x - x_i)| < \eta$$

for all $x \in [0, 1] \subset [x_i - 1, x_i + 1]$ (since $0 \le x_i \le 1$). By a judicious choice of η, depending on the predetermined constants C_i in (3.2), it then follows that

$$\left| g(x) - \left[Ax + B + \sum_{i=1}^{m-1} C_i p(x - x_i) \right] \right| < \varepsilon$$

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

Thus our problem has been reduced to that of approximating just the one function $|x|$. How can this be done? As Lebesgue explains, one can write

$$|x| = \sqrt{x^2} = \sqrt{1 - (1 - x^2)} = \sqrt{1 - z}$$

where $z = 1 - x^2$, and then expand the above radical by the binomial formula to obtain a power series in $z = 1 - x^2$ which converges uniformly to $|x|$ in $[-1, 1]$. One finally just truncates the power series.

To be more explicit, we have

$$(1 - z)^{1/2} = \sum_{n=0}^{\infty} \binom{1/2}{n} (-z)^n$$

where

$$\binom{1/2}{n} = \frac{\frac{1}{2}(\frac{1}{2} - 1) \cdots (\frac{1}{2} - n + 1)}{n!} = \frac{(-1)^{n-1} \frac{1}{2} \frac{1}{2} \frac{3}{2} \cdots \frac{2n-3}{2}}{n!}.$$

Thus

$$(1 - z)^{1/2} = 1 - \sum_{n=1}^{\infty} a_n z^n$$

with $a_1 = 1/2$, 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 $(1-z)^{1/2}$ in $|z| \le 1$. It is easily checked that the radius of convergence of this power series is 1. An application of Stirling's formula shows that

$$a_n = \frac{e}{2\sqrt{\pi}} \frac{1}{n^{3/2}}(1+o(1))$$

so that the series also has the correct convergence properties for $|z| = 1$. A different proof of this same fact may be found in Todd [1961, p. 11]. This finishes Lebesgue's proof.

An alternative argument (see Ostrowski [1951, p. 168] or Feinerman, Newman [1974, p. 5]) gets around the more delicate analysis at $|z| = 1$ by noting that $(1-z)^{1/2}$ may be uniformly approximated on $[0,1]$ by $(1-\rho z)^{1/2}$ as $\rho \uparrow 1$. (In fact it is easily checked that for $0 < \rho < 1$

$$|(1-z)^{1/2} - (1-\rho z)^{1/2}| \le (1-\rho)^{1/2}$$

for all $z \in [0,1]$.) Now the power series for $(1-\rho z)^{1/2}$, namely

$$(1-\rho z)^{1/2} = 1 - \sum_{n=1}^{\infty} a_n \rho^n z^n,$$

is absolutely and uniformly convergent in $|z| < \rho^{-1}$ and thus in $|z| \le 1$.

Bourbaki [1949, p. 55] (see also Dieudonné [1969, p. 137]) presents an ingenious argument to obtain a sequence of polynomials which uniformly approximate $|x|$. For $t \in [0,1]$ define a sequence of polynomials recursively as follows. Let $p_0(t) \equiv 0$ and

$$p_{n+1}(t) = p_n(t) + \frac{1}{2}(t - p_n^2(t)),$$

$n = 0, 1, 2, \ldots$. It is readily verified that for each fixed $t \in [0,1]$, $p_n(t)$ is an increasing sequence bounded above by $\sqrt{t}$. The former is a consequence of the latter which is proven as follows. Assume $0 \le p_n(t) \le \sqrt{t}$. Then

$$\begin{aligned}\sqrt{t} - p_{n+1}(t) =& \sqrt{t} - p_n(t) - \frac{1}{2}(t - p_n^2(t)) \\ =& (\sqrt{t} - p_n(t))(1 - \frac{1}{2}(\sqrt{t} + p_n(t))) \\ \ge& 0\end{aligned}$$

since $\sqrt{t} + p_n(t) \le 2\sqrt{t} \le 2$ for $t \in [0,1]$. Thus for each $t \in [0,1]$

$$\lim_{n\to\infty} p_n(t) = p(t)$$

exists. Since $p(t)$ is nonnegative and satisfies

$$p(t) = p(t) - \frac{1}{2}(t - p^2(t))$$

we have $p(t) = \sqrt{t}$. The $\{p_n\}$ are real-valued continuous functions (polynomials) which increase, and converge pointwise to a continuous function p. This implies that the convergence is uniform (Dini's theorem). Let $q_n(x) = p_n(x^2)$ for $x \in [-1, 1]$. Then the polynomials $\{q_n\}$ converge uniformly to $\sqrt{x^2} = |x|$ on $[-1, 1]$. A similar and equivalent proof may be found in Sz.-Nagy [1965, p. 77]. (Sz.-Nagy attributes his procedure to C. Visser.)

Mittag-Leffler. The proof by Mittag-Leffler as given in Mittag-Leffler [1900] is the following. He also considers the g as given in (3.1), and sets

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

It is easily checked that

$$\lim_{n\to\infty} \chi_n(x) = \begin{cases} 1, & 0 < x \le 1 \\ 0, & x = 0 \\ -1, & -1 \le x < 0 \end{cases}.$$

Furthermore, since each χ_n is increasing on $[-1, 1]$, and $\chi_{n+1}(x) > \chi_n(x)$ for $x \in (0, 1]$, while $\chi_{n+1}(x) < \chi_n(x)$ for $x \in (-1, 0)$, it follows that given $\delta > 0$, small, the function χ_n uniformly converges to 1 on $[\delta, 1]$ and to -1 on $[-1, -\delta]$. Thus the functions

$$h_n = \frac{\chi_n + 1}{2}$$

are bounded on $[-1, 1]$ and uniformly approximate the function h of (3.1) on $[-1, -\delta] \bigcup [\delta, 1]$ for any given δ. Furthermore the χ_n and thus the h_n are entire (analytic) functions.

As previously, since $g_{i+1} - g_i$ is a linear polynomial vanishing at x_i, a short calculation verifies that for each $x_i \in (0, 1)$

$$[g_{i+1}(x) - g_i(x)]\, h_n(x - x_i)$$

uniformly converges to

$$[g_{i+1}(x) - g_i(x)]\, h(x - x_i)$$

on $[0, 1]$. Replacing the h in (3.1) by h_n we obtain a series of functions $\{H_n\}$ that uniformly approximate g. Finally, since h_n is an entire function, each of the functions H_n is an entire function. As such they may be approximated arbitrarily well by a truncation of their power series. This again proves Weierstrass' theorem.

Fejér. L. Fejér (1880–1959) was a student of H. A. Schwarz. What we will report on here is taken from Fejér [1900] (he had just turned 20 when the paper appeared). This fundamental paper formed the basis for Fejér's doctoral thesis obtained in 1902 from the University of Budapest. The paper contains what is today described as the "classic" theorem on Cesàro $(C,1)$ summability of Fourier series. As we are interested in Weierstrass' theorem, we will restrict ourselves, a priori, to $f \in \widetilde{C}[0,2\pi]$, and prove that the Cesàro sum of the Fourier series of any such f converges uniformly to f. Note that this is the first proof of Weierstrass' theorem (in the trigonometric polynomial case) that actually provides, by a linear process, a sequence of easily calculated approximants.

Let $\sigma_0(x) = 1/2$, and

$$\sigma_m(x) = \frac{1}{2} + \cos x + \cos 2x + \cdots + \cos mx$$

for $m = 1, 2, \ldots$. Set

$$G_n(x) = \frac{\sigma_0(x) + \cdots + \sigma_{n-1}(x)}{n}.$$

A calculation shows that

$$G_n(x) = \frac{1}{2n} \frac{1 - \cos nx}{1 - \cos x} = \frac{1}{2n} \left[\frac{\sin\left(\frac{nx}{2}\right)}{\sin\left(\frac{x}{2}\right)} \right]^2 .$$

Furthermore it is easily seen that

$$\frac{1}{\pi} \int_0^{2\pi} G_n(x)\,\mathrm{d}x = 1.$$

G_n is a nonnegative kernel that integrates to 1 (and, as we shall show approaches the Dirac-Delta function at 0 as n tends to infinity, i.e., convolution against G_n approaches the identity operator).

Assume $f \in \widetilde{C}[0,2\pi]$. Let

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos kx + b_k \sin kx$$

denote the Fourier series of f. Let $s_0(x) = a_0/2$, and

$$s_m(x) = \frac{a_0}{2} + \sum_{k=1}^{m} a_k \cos kx + b_k \sin kx$$

denote the partial sums of the Fourier series of f. The functions s_m do not necessarily converge uniformly, or pointwise, to f as $m \to \infty$. This is a well-known result of du Bois-Reymond [1876]. However let us now set

$$S_n(x) = \frac{s_0(x) + \cdots + s_{n-1}(x)}{n} = \frac{1}{\pi} \int_0^{2\pi} f(y) G_n(y - x)\,\mathrm{d}y.$$

Explicitly the S_n are given by

$$S_n(x) = \frac{a_0}{2} + \sum_{k=1}^{n-1} \left(1 - \frac{k}{n}\right) \left[a_k \cos kx + b_k \sin kx\right].$$

Surprisingly, the S_n always converge uniformly to f.

Theorem 5. *For each $f \in \widetilde{C}[0, 2\pi]$, the trigonometric polynomials S_n converge uniformly to f as $n \to \infty$.*

Proof: From the above

$$S_n(x) = \frac{1}{\pi}\int_0^{2\pi} f(y)G_n(y-x)\,\mathrm{d}y = \frac{1}{2n\pi}\int_0^{2\pi} f(y)\frac{1-\cos n(y-x)}{1-\cos(y-x)}\,\mathrm{d}y.$$

Since $f \in \widetilde{C}[0, 2\pi]$, f may be considered to be uniformly continuous on all of $I\!R$. Thus given $\varepsilon > 0$ there exists a $\delta > 0$ such that if $|x-y| < \delta$, then

$$|f(x) - f(y)| < \frac{\varepsilon}{2}.$$

In what follows we assume $\delta < \pi/2$.

Since G_n integrates to 1 we have

$$S_n(x) - f(x) = \frac{1}{\pi}\int_0^{2\pi}[f(y)-f(x)]G_n(y-x)\,\mathrm{d}y$$

$$= \frac{1}{\pi}\int_{|y-x|<\delta}[f(y)-f(x)]G_n(y-x)\,\mathrm{d}y + \frac{1}{\pi}\int_{\delta\le|y-x|\le\pi}[f(y)-f(x)]G_n(y-x)\,\mathrm{d}y.$$

We estimate each of the above two integrals.

On $|y-x| < \delta$ we have $|f(x)-f(y)| < \frac{\varepsilon}{2}$. Thus

$$\left|\frac{1}{\pi}\int_{|y-x|<\delta}[f(y)-f(x)]G_n(y-x)\,\mathrm{d}y\right| < \frac{\varepsilon}{2}\frac{1}{\pi}\int_{|y-x|<\delta}G_n(y-x)\,\mathrm{d}y$$

$$< \frac{\varepsilon}{2}\frac{1}{\pi}\int_0^{2\pi}G_n(y-x)\,\mathrm{d}y = \frac{\varepsilon}{2}.$$

We have here used the crucial fact that G_n is nonnegative and integrates to 1 over any interval of length 2π.

From the explicit form of G_n and the inequality $|f(y)-f(x)| \le 2\|f\|$ we have

$$\left|\frac{1}{\pi}\int_{\delta\le|y-x|\le\pi}[f(y)-f(x)]G_n(y-x)\,\mathrm{d}y\right| \le \frac{2\|f\|}{2n\pi}\int_{\delta\le|y-x|\le\pi}\frac{1-\cos n(y-x)}{1-\cos(y-x)}\,\mathrm{d}y.$$

Now $|1-\cos n(y-x)| \le 2$, while on $\delta \le |y-x| \le \pi$ we have $1-\cos(y-x) \ge 1-\cos\delta$. Thus

$$\left|\frac{1}{\pi}\int_{\delta\le|y-x|\le\pi}[f(y)-f(x)]G_n(y-x)\,\mathrm{d}y\right| \le \frac{2\|f\|}{2n\pi}\frac{2}{1-\cos\delta}2\pi = \frac{4\|f\|}{n(1-\cos\delta)}.$$

For n sufficiently large

$$\frac{4\|f\|}{n(1-\cos\delta)} < \frac{\varepsilon}{2}.$$

Thus for such n

$$|S_n(x) - f(x)| < \varepsilon. \quad \square$$

Applying the method of the (second) proof of Proposition 3 to the above we see that to each $f \in C[-1,1]$ we may obtain a sequence of algebraic polynomials

$$p_n(x) = \frac{a_0}{2} + \sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right) a_k T_k(x)$$

where

$$a_k = \frac{2}{\pi}\int_{-1}^{1} \frac{f(x)T_k(x)}{\sqrt{1-x^2}}\,\mathrm{d}x,$$

$k = 0, 1, \ldots$. These explicitly defined p_n (each of degree at most $n-1$) uniformly approximate f.

Lerch II. The paper Lerch [1903] contains yet another proof of the density of algebraic polynomials in $C[0,1]$. In his previous proof, in Lerch [1892], Lerch had used general properties of Fourier series to prove the Weierstrass theorem for algebraic polynomials. His proof here is different in that while the same general scheme is used, he only needs to consider the Fourier series of two specific functions, and their properties. In this sense it is more elementary than his previous proof.

We recall from Lerch [1892] that it suffices to be able to arbitrarily approximate the polygonal line g as given in (3.1). Lerch rewrites (3.1) in the form

$$g(x) = \sum_{i=1}^{m} \ell_i(x)$$

where

$$\ell_i(x) = \begin{cases} 0, & x < x_{i-1} \\ y_{i-1} + \left(\frac{x - x_{i-1}}{x_i - x_{i-1}}\right)(y_i - y_{i-1}), & x_{i-1} \le x < x_i \\ 0, & x_i \le x \end{cases}$$

(when defining ℓ_m we should, for precision, define it to equal y_m at $x_m = 1$).

As we mentioned, Lerch bases his proof on quite explicit Fourier series. It is well known and easily checked that

$$\frac{1}{2} - x = \sum_{n=1}^{\infty} \frac{\sin 2n\pi x}{n\pi}, \qquad 0 < x < 1, \tag{3.3}$$

while

$$x^2 - x + \frac{1}{6} = \sum_{n=1}^{\infty} \frac{\cos 2n\pi x}{n^2\pi^2}, \qquad 0 \le x \le 1. \tag{3.4}$$

There is a problem with the convergence of the Fourier series in (3.3). This series converges uniformly to $1/2 - x$ on any $[a, b]$, $0 < a < b < 1$, but does not converge uniformly in any neighbourhood of $x = 0$ or $x = 1$. (In fact its value at $x = 0$ and $x = 1$ is 0.) However the series in (3.4) does converge absolutely and uniformly to the given function on all of $[0, 1]$. It is also readily checked, using the 1-periodicity of the Fourier series, that the function

$$\frac{1}{2}(x_i - x_{i-1})(y_i + y_{i-1}) + \sum_{n=1}^{\infty} \frac{y_{i-1}\sin 2n\pi(x - x_{i-1}) - y_i \sin 2n\pi(x - x_i)}{n\pi}$$
$$-\frac{1}{2}\frac{(y_i - y_{i-1})}{(x_i - x_{i-1})}\sum_{n=1}^{\infty} \frac{\cos 2n\pi(x - x_{i-1}) - \cos 2n\pi(x - x_i)}{n^2\pi^2}$$

is the Fourier series of ℓ_i and that there is uniform convergence of this series to ℓ_i on any compact subset of $[0, 1]$ not containing x_{i-1} and x_i.

Thus

$$\frac{1}{2}\sum_{i=1}^{m}(x_i - x_{i-1})(y_i + y_{i-1}) + \sum_{n=1}^{\infty} \frac{y_0 \sin 2n\pi x - y_m \sin 2n\pi(x - 1)}{n\pi}$$
$$-\frac{1}{2}\sum_{i=1}^{m}\frac{(y_i - y_{i-1})}{(x_i - x_{i-1})}\sum_{n=1}^{\infty} \frac{\cos 2n\pi(x - x_{i-1}) - \cos 2n\pi(x - x_i)}{n^2\pi^2}$$

is the Fourier series of g. Note that this series converges uniformly to g also at $x_1, \ldots, x_{m-1}$. There remains the problem of convergence at $x_0 = 0$ and $x_m = 1$. (However if $g \in \widetilde{C}[0, 1]$, i.e., g is 1-periodic, then $y_0 = y_m$ and the problematic term has disappeared. In this case, we have constructed the Fourier series of g which converges absolutely and uniformly to g on $[0, 1]$. Truncate this Fourier series to obtain a trigonometric polynomial which approximates g arbitrarily well. This proves the density of trigonometric polynomials.) If $y_0 \neq y_m$ then we may, as does Lerch, again apply (3.3) to obtain

$$\frac{1}{2}\sum_{i=1}^{m}(x_i - x_{i-1})(y_i + y_{i-1}) + (y_0 - y_m)(\frac{1}{2} - x)$$
$$-\frac{1}{2}\sum_{i=1}^{m}\frac{(y_i - y_{i-1})}{(x_i - x_{i-1})}\sum_{n=1}^{\infty} \frac{\cos 2n\pi(x - x_{i-1}) - \cos 2n\pi(x - x_i)}{n^2\pi^2}.$$

(Alternatively, just shift g by a polynomial so that the new g satisfies $g(0) = g(1)$.) This series converges absolutely and uniformly to g on all of $[0, 1]$. Truncating this infinite series we obtain an entire function (trigonometric polynomial) that approximates g arbitrarily well. We now appropriately truncate the power series of this entire function to obtain the desired algebraic polynomial.

Unfortunately there is no indication, in Lerch [1903], that he was aware of any of the other published proofs of the Weierstrass theorem. A careful reading of this proof shows that it is essentially a quasi-constructive version of Lebesgue's proof.

Landau. The proof of E. Landau (1877–1938) in Landau [1908] follows the tradition of the proofs of Weierstrass, Picard and Fejér in that the essential underlying mechanism in his proof is a singular integral. However it is more direct than the former two in its judicious choice of the kernel. Let $f \in C[a,b]$ where, without loss of generality, it will be assumed that $0 < a < b < 1$. Extend f to be a continuous function on all of $[0,1]$.

Define

$$k_n = \int_{-1}^{1} (1-u^2)^n \, \mathrm{d}u$$

and set

$$p_n(x) = \frac{1}{k_n} \int_0^1 f(y) \left[1-(x-y)^2\right]^n \, \mathrm{d}y.$$

Note that p_n is a polynomial of degree at most $2n$ in x. What Landau proves is that the sequence of polynomials $\{p_n\}$ converge uniformly to f on $[a,b]$. Landau's sequence of polynomial approximants differ from those of the previous proofs (except for Fejér's proof) in that they are explicitly given, and in that they are obtained via a linear method.

We first present Landau's original proof. In this proof we will use the following estimates. For every $0 < \delta < 1$,

$$\int_{\delta \le |u| \le 1} (1-u^2)^n \, \mathrm{d}u \le \int_{\delta \le |u| \le 1} (1-\delta^2)^n \, \mathrm{d}u < 2(1-\delta^2)^n.$$

Similarly

$$k_n = \int_{-1}^{1} (1-u^2)^n \, \mathrm{d}u \ge \int_{|u| \le 1/\sqrt{n}} (1-u^2)^n \, \mathrm{d}u \ge \int_{|u| \le 1/\sqrt{n}} \left(1-\frac{1}{n}\right)^n \mathrm{d}u$$

$$= \frac{2}{\sqrt{n}} \left(1-\frac{1}{n}\right)^n.$$

Thus

$$\frac{1}{k_n} \int_{\delta \le |u| \le 1} (1-u^2)^n \, \mathrm{d}u \le \sqrt{n}(1-\delta^2)^n \left(1-\frac{1}{n}\right)^{-n}.$$

Note that for every fixed $\delta \in (0,1)$ we have

$$\lim_{n \to \infty} \sqrt{n}(1-\delta^2)^n \left(1-\frac{1}{n}\right)^{-n} = 0.$$

Now choose $\varepsilon > 0$. Since f is uniformly continuous on $[0,1]$ there exists a $\delta > 0$ such that if $x, y \in [0,1]$ satisfies $|x-y| < \delta$, then

$$|f(x) - f(y)| < \varepsilon/3.$$

Assume $0 < \delta < \min\{a, 1-b\}$. Choose N so that for all $n \geq N$

$$2\|f\|\sqrt{n}(1-\delta^2)^n \left(1 - \frac{1}{n}\right)^{-n} < \varepsilon/3.$$

For every $x \in [a,b]$,

$$|p_n(x) - f(x)| = \left| \frac{1}{k_n} \int_0^1 f(y) \left[1-(x-y)^2\right]^n \, \mathrm{d}y - f(x) \right|$$

$$\leq \frac{1}{k_n} \int_0^1 |f(y)-f(x)| \left[1-(x-y)^2\right]^n \, \mathrm{d}y + |f(x)| \left| 1 - \frac{1}{k_n} \int_0^1 \left[1-(x-y)^2\right]^n \, \mathrm{d}y \right|.$$

We bound the integral

$$\frac{1}{k_n} \int_0^1 |f(y) - f(x)| \left[1-(x-y)^2\right]^n \, \mathrm{d}y$$

by considering separately integration over $\{y : |x-y| < \delta\}$ and over $\{y : \delta \leq |x-y|\}$ for $y \in [0,1]$.

Now

$$\frac{1}{k_n} \int_{|x-y|<\delta} |f(y) - f(x)| \left[1-(x-y)^2\right]^n \, \mathrm{d}y$$
$$< \frac{\varepsilon}{3} \frac{1}{k_n} \int_{|x-y|<\delta} \left[1-(x-y)^2\right]^n \, \mathrm{d}y < \frac{\varepsilon}{3}.$$

Furthermore

$$\frac{1}{k_n} \int_{\substack{\delta \leq |x-y| \\ 0 \leq y \leq 1}} |f(y) - f(x)| \left[1-(x-y)^2\right]^n \, \mathrm{d}y \leq \frac{2\|f\|}{k_n} \int_{\delta \leq |u| \leq 1} [1-u^2]^n \, \mathrm{d}u$$

$$\leq 2\|f\|\sqrt{n}(1-\delta^2)^n \left(1 - \frac{1}{n}\right)^{-n} < \varepsilon/3.$$

Finally

$$|f(x)| \left| 1 - \frac{1}{k_n} \int_0^1 \left[1-(x-y)^2\right]^n \, \mathrm{d}y \right|$$
$$\leq \frac{\|f\|}{k_n} \left| \int_1^1 [1-u^2]^n \, \mathrm{d}u - \int_{-x}^{1-x} [1-u^2]^n \, \mathrm{d}u \right|.$$

Since $x \in [a, b]$ and $\delta < \min\{a, 1-b\}$, we have

$$\frac{\|f\|}{k_n}\left|\int_{-1}^{1}[1-u^2]^n\,\mathrm{d}u - \int_{-x}^{1-x}[1-u^2]^n\,\mathrm{d}u\right| \le \frac{\|f\|}{k_n}\int_{\delta\le|u|\le 1}[1-u^2]^n\,\mathrm{d}u$$

$$\le \|f\|\sqrt{n}(1-\delta^2)^n\left(1-\frac{1}{n}\right)^{-n} < \varepsilon/3.$$

This proves the result. □

For completeness and as a matter of interest, it easily follows from integration by parts that

$$k_n = \int_{-1}^{1}[1-u^2]^n\,\mathrm{d}u = \frac{2^{2n+1}(n!)^2}{(2n+1)!}.$$

Applying Stirling's formula it may be shown that

$$\lim_{n\to\infty}\sqrt{n}k_n = \sqrt{\pi}.$$

The following is a variation on and simplification of Landau's proof. It is due to Jackson [1934]. As above, assume $f \in C[a, b]$ with $0 < a < b < 1$. Extend f to be a continuous function on all of $\mathbb{R}$ which also vanishes identically off $[0, 1]$. This latter fact, together with a change of variable argument, gives

$$\begin{aligned} p_n(x) =& \frac{1}{k_n}\int_0^1 f(y)\left[1-(x-y)^2\right]^n\,\mathrm{d}y \\ =& \frac{1}{k_n}\int_{-1}^{1} f(x+u)(1-u^2)^n\,\mathrm{d}u \end{aligned}$$

and thus we get the simpler

$$p_n(x) - f(x) = \frac{1}{k_n}\int_{-1}^{1}\left[f(x+u)-f(x)\right](1-u^2)^n\,\mathrm{d}u.$$

Let ε and δ be as above. For $|u| \ge \delta$, we have

$$|f(x+u)-f(x)| \le 2\|f\| \le \frac{2\|f\|u^2}{\delta^2},$$

while for $|u| < \delta$ we have

$$|f(x+u)-f(x)| < \frac{\varepsilon}{3}.$$

Thus

$$|f(x+u) - f(x)| < \frac{\varepsilon}{3} + \frac{2\|f\|u^2}{\delta^2}$$

for all $x, u \in [0,1]$. Substituting it follows that

$$|p_n(x) - f(x)| < \frac{1}{k_n}\int_{-1}^{1} \frac{\varepsilon}{3}(1-u^2)^n \, \mathrm{d}u + \frac{1}{k_n}\int_{-1}^{1} \frac{2\|f\|u^2}{\delta^2}(1-u^2)^n \, \mathrm{d}u$$

$$= \frac{\varepsilon}{3} + \frac{2\|f\|}{\delta^2 k_n}\int_{-1}^{1} u^2(1-u^2)^n \, \mathrm{d}u.$$

Set

$$j_n = \int_{-1}^{1} u^2(1-u^2)^n \, \mathrm{d}u.$$

Integration by parts yields

$$j_n = \frac{-u(1-u^2)^{n+1}}{2(n+1)}\Big|_{-1}^{1} + \int_{-1}^{1} \frac{(1-u^2)^{n+1}}{2(n+1)} \, \mathrm{d}u = \frac{k_{n+1}}{2(n+1)}.$$

Since $(1-u^2) \leq 1$ on $[-1,1]$ we also have $k_{n+1} \leq k_n$. Thus

$$j_n \leq \frac{k_n}{2(n+1)}.$$

Substituting we obtain

$$|p_n(x) - f(x)| < \frac{\varepsilon}{3} + \frac{\|f\|}{\delta^2(n+1)}.$$

We now choose n sufficiently large so that

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

for all $x \in [0,1]$ and thus on $[a,b]$.

For much more concerning the "Landau" polynomials, see Butzer, Stark [1986], and the many references therein.

A few months after the appearance of Landau [1908], Lebesgue "responded" with Lebesgue [1908] which appeared in the same journal and is an "extract from a letter addressed to E. Landau". Despite Lebesgue's flowery opening *Je me félicite de m'etre rencontré avec vous sur un point particulier ...*, Lebesgue then goes on to inform Landau that he actually had the same proof for more than two years, but his manuscript was not yet ready (he is probably referring to his treatise Lebesgue [1909]). But since Landau did publish, then Lebesgue feels called upon to tell Landau (and the world) about some of his reflections on this matter. Aside from the entertainment value of this exchange between two stars, Lebesgue does make two valid points. The first has less to do with Landau's particular proof than with the proofs of Weierstrass, Picard, Fejér, and Landau. Lebesgue notes that these proofs can and should be considered within the general context of integral convolutions with sequences of non-negative kernels, where the convolution approaches the identity. This was subsequently elaborated upon in Lebesgue [1909]. Furthermore in the latter half of this short paper Lebesgue goes on to ask questions about the order of approximation. This is a clear indication that the subject is evolving.

De la Vallée Poussin. The treatise de la Vallée Poussin [1908] also contains a proof of Weierstrass' theorem using this exact same integral. In fact Ch. J. de la Vallée Poussin (1866–1962) devotes over 30 pages of his paper to a study of its various approximation properties (and not only the question of density). A footnote on p. 197 therein states that de la Vallée Poussin was made aware of Landau's paper only while editing his own paper. (Landau's paper appeared in January of 1908.) So it seems that three outstanding mathematicians almost simultaneously discovered this method of proving Weierstrass' theorem. As Landau states, this integral had in fact already been introduced by Stieltjes in a letter to Hermite dated September 12, 1893 (see Baillaud, Bourget [1905]).

In addition, de la Vallée Poussin introduced, in the second half of de la Vallée Poussin [1908], what he regarded as the periodic analogues of the Landau polynomials. These are

$$I_n(x) = \frac{1}{h_n} \int_{-\pi}^{\pi} f(y) \left[\cos\left(\frac{y-x}{2}\right)\right]^{2n} \mathrm{d}y$$

where

$$h_n = \int_{-\pi}^{\pi} \left[\cos\left(\frac{y}{2}\right)\right]^{2n} \mathrm{d}y = \frac{\pi(2n)!}{2^{2n-1}(n!)^2}.$$

I_n is a trigonometric polynomial of degree at most n. The proof of the fact that the I_n uniformly converge to f for $f \in \widetilde{C}[-\pi, \pi]$ is very similar to the proof of the analogous result for the Landau polynomials. We will not repeat the proof here. For more concerning this proof, this paper, and de la Vallée Poussin's other contributions to approximation theory, we recommend Butzer, Nessel [1993].

Bernstein. What we will arbitrarily call the last of the early proofs of the Weierstrass theorems is due to S. N. Bernstein (1880–1968) and appeared in Bernstein [1912/13]. (The thesis advisor of Bernstein's first doctorate was Picard.) This paper is reproduced in Stark [1981]. A translation into Russian appears in his somewhat more accessible collected works. This proof is very different from the previous proofs, and has had a profound impact in various areas. It is here that Bernstein introduces what we today call Bernstein polynomials.

The Bernstein polynomial of $f \in C[0, 1]$ is defined by

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

Bernstein demonstrates, using probabilistic ideas, that the B_n converge uniformly to f on $[0,1]$. The proof of this fact, as generally given today, is slightly different from Bernstein's original proof and has the added advantage of providing "error estimates". We will here present Bernstein's original proof, although it is somewhat overinvolved.

Since $f \in C[0,1]$, given $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$|x - y| < \delta$$

implies

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

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

$$\overline{f}(x) = \max\{f(y) : y \in [x-\delta, x+\delta] \cap [0,1]\}$$

and

$$\underline{f}(x) = \min\{f(y) : y \in [x-\delta, x+\delta] \cap [0,1]\}.$$

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

$$0 \le \overline{f}(x) - f(x) < \frac{\varepsilon}{2},$$

and

$$0 \le f(x) - \underline{f}(x) < \frac{\varepsilon}{2}.$$

For fixed $\delta > 0$ as above, set

$$\eta_n(x) = \sum_{\{m : |x-(m/n)| > \delta\}} \binom{n}{m} x^m (1-x)^{n-m}.$$

From the decomposition

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

$$= \sum_{\{m : |x-(m/n)| \le \delta\}} f\left(\frac{m}{n}\right) \binom{n}{m} x^m (1-x)^{n-m}$$

$$+ \sum_{\{m : |x-(m/n)| > \delta\}} f\left(\frac{m}{n}\right) \binom{n}{m} x^m (1-x)^{n-m},$$

it easily follows that

$$\underline{f}(x)[1-\eta_n(x)] - \|f\|\eta_n(x) \le B_n(x) \le \overline{f}(x)[1-\eta_n(x)] + \|f\|\eta_n(x).$$

Bernstein then states that according to Bernoulli's theorem there exists an N such that for all $n > N$ and all $x \in [0,1]$ we have

$$\eta_n(x) < \frac{\varepsilon}{4\|f\|}.$$

Thus as a consequence of

$$f(x) + [\underline{f}(x) - f(x)] - \eta_n(x)[\|f\| + \underline{f}(x)] \le B_n(x)$$

and

$$B_n(x) \le f(x) + [\overline{f}(x) - f(x)] + \eta_n(x)[\|f\| - \overline{f}(x)],$$

we obtain

$$f(x) - \frac{\varepsilon}{2} - \frac{\varepsilon}{4\|f\|}2\|f\| < B_n(x) < f(x) + \frac{\varepsilon}{2} + \frac{\varepsilon}{4\|f\|}2\|f\|,$$

which gives

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

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

For completeness we now verify Bernstein's statement regarding $\eta_n(x)$. (For a probabilistic explanation of this quantity and estimate, see e. g. Levasseur [1984].) To this end confirm that

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

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

and

$$\sum_{m=0}^{n} \frac{m^2}{n^2} \binom{n}{m} x^m (1-x)^{n-m} = x^2 + \frac{x(1-x)}{n}.$$

Then

$$\begin{aligned}
\eta_n(x) &= \sum_{\{m:|x-(m/n)|>\delta\}} \binom{n}{m} x^m (1-x)^{n-m} \\
&\le \sum_{\{m:|x-(m/n)|>\delta\}} \left(\frac{x-\frac{m}{n}}{\delta}\right)^2 \binom{n}{m} x^m (1-x)^{n-m} \\
&\le \frac{1}{\delta^2} \sum_{m=0}^{n} \left(x - \frac{m}{n}\right)^2 \binom{n}{m} x^m (1-x)^{n-m} \\
&= \frac{1}{\delta^2}\left[x^2 - 2x\cdot x + x^2 + \frac{x(1-x)}{n}\right] \\
&= \frac{x(1-x)}{n\delta^2} \\
&\le \frac{1}{4n\delta^2}.
\end{aligned}$$

for all $x \in [0,1]$. Thus for each fixed $\delta > 0$ we can in fact choose N such that for all $n \geq N$ and all $x \in [0,1]$

$$\eta_n(x) < \frac{\varepsilon}{4\|f\|}.$$

This ends Bernstein's proof.

Bernstein's proof is beautiful and elegant! It constructs in a simple, linear (but unexpected) manner a sequence of approximating polynomials depending explicitly on the values of f at rational points. No further information regarding f is used. This was not the first attempt to find a proof of the Weierstrass theorem using a suitable partition of unity. In Borel [1905, p. 79–82], which seems to have been the first textbook devoted mainly to approximation theory, we find the following formula for constructing a sequence of polynomials approximating every $f \in C[0,1]$.

E. Borel (1871–1956) proved that the sequence of polynomials

$$p_n(x) = \sum_{m=0}^{n} f\left(\frac{m}{n}\right) q_{n,m}(x)$$

uniformly approximates f where the $q_{n,m}$ are fixed polynomials independent of f. His $q_{n,m}$ are constructed as follows. Set

$$g_{n,m}(x) = \begin{cases} 0, & \left|x - \frac{m}{n}\right| > \frac{1}{n} \\ nx - (m-1), & \frac{m-1}{n} \leq x \leq \frac{m}{n} \\ -nx + (m+1), & \frac{m}{n} \leq x \leq \frac{m+1}{n}. \end{cases}$$

Note that the $g_{n,m}$ are non-negative, sum to 1, and $g_{n,m}(m/n) = 1$. Let (by the Weierstrass theorem) $q_{n,m}$ be any polynomial satisfying

$$|g_{n,m}(x) - q_{n,m}(x)| < \frac{1}{n^2}$$

for all $x \in [0,1]$. It is now not difficult to verify that the p_n do approximate f. However the Bernstein polynomials are so much more satisfying in so many ways.

Kuhn's Proof. There are many elegant and simple proofs of Weierstrass' theorem. But perhaps the most elementary proof (of which we are aware) is the following due to Kuhn [1964]. Kuhn's proof uses one basic inequality, namely Bernoulli's inequality

$$(1+h)^n \geq 1 + nh$$

which is valid for $h \geq -1$ and $n \in \mathbb{N}$.

We present Kuhn's proof except that we save a step by recalling (see (3.1)) that we need only approximate continuous polygonal lines which we can write as

$$g(x) = g_1(x) + \sum_{i=1}^{m-1} [g_{i+1}(x) - g_i(x)]h(x - x_i)$$

where the $0 = x_0 < x_1 < \cdots < x_m = 1$ are the abscissae of the polygonal line g, each g_i is linear, $g_{i+1} - g_i$ vanishes at x_i, and

$$h(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$

This form was used in the proofs of Runge/Phragmén, of Mittag-Leffler and of Lebesgue. In fact, in the first two of these proofs it was noted that it suffices to find a sequence of polynomials bounded on $[-1, 1]$ and approximating h uniformly on $[-1, -\delta] \cup [\delta, 1]$, for any given $\delta > 0$.

Kuhn simply writes down such a sequence of polynomials, namely

$$p_n(x) = \left[1 - \left(\frac{1-x}{2}\right)^n\right]^{2^n}.$$

(Note that the polynomials $\{x[2p_n(x) - 1]\}$ uniformly converge to $|x|$ on $[-1, 1]$. See Lebesgue's proof.)

It is more convenient to consider the simpler

$$q_n(x) = (1 - x^n)^{2^n},$$

which is just a shift and rescale of p_n. On $[0, 1]$ the q_n are decreasing and satisfy $q_n(0) = 1$, $q_n(1) = 0$. The requisite facts concerning the p_n therefore reduce to showing

$$\lim_{n\to\infty} q_n(x) = \begin{cases} 1, & 0 \leq x < 1/2 \\ 0, & 1/2 < x \leq 1 \end{cases}.$$

Let $x \in [0, 1/2)$. Then from Bernoulli's inequality

$$1 \geq q_n(x) = (1 - x^n)^{2^n} \geq 1 - (2x)^n.$$

Since $0 \leq 2x < 1$, we have

$$\lim_{n\to\infty} q_n(x) = 1.$$

Let $x \in (1/2, 1)$. Then using Bernoulli's inequality we obtain

$$\frac{1}{q_n(x)} = \frac{1}{(1-x^n)^{2^n}} = \left(1 + \frac{x^n}{1-x^n}\right)^{2^n} \geq 1 + \frac{(2x)^n}{1-x^n} > (2x)^n$$

and thus

$$0 < q_n(x) < \frac{1}{(2x)^n}.$$

As $2x > 1$, it follows that

$$\lim_{n\to\infty} q_n(x) = 0.$$

The monotonicity of the q_n implies that this approximation is appropriately uniform. This ends Kuhn's proof.

References

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

Bell, E. T. [1936] "Men of Mathematics", Scientific Book Club, London.

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.

du Bois-Reymond, P. [1876] Untersuchungen über die Convergenz und Divergenz der Fourierschen Darstellungsformeln, *Abhandlungen der Mathematisch-Physicalischen Classe der K. Bayerische Akademie der Wissenshaften* **12**, 1–13.

Borel, É. [1905] "Lecons sur les Fonctions de Variables Réelles et les Développements en Séries de Polynomes", Gauthier-Villars, Paris. (2nd edition, 1928).

Bourbaki, N. [1949] "Topologie Générale (Livre III). Espaces Fonctionnels Dictionnaire (Chapitre X)", Hermann & Cie, Paris.

Butzer, P. L. and R. J. Nessel [1993] Aspects of de la Vallée Poussin's work in approximation and its influence, *Archive Hist. Exact Sciences* **46**, 67–95.

Butzer, P. L. and E. L. Stark [1986] The singular integral of Landau alias the Landau polynomials - Placement and impact of Landau's article "Über die Approximation einer stetigen Funktion durch eine ganze rationale Funktion", in *Edmund Landau, Collected Works, Volume 3*, P. T. Bateman, L. Mirsky, H. L. Montgomery, W. Schall, I. J. Schoenberg, W. Schwarz, H. Wefelscheid, eds, Thales-Verlag, Essen, 83–111.

Cakon, R. [1987] "Alternative Proofs of Weierstrass Theorem of Approximation: An Expository Paper", dissertation, Master's Thesis, Department of Mathematics, The Pennsylvania State University.

Dieudonné, J. [1969] "Foundations of Modern Analysis", Academic Press, New York.

Dirichlet, L. [1829] Sur la convergence des séries trigonométriques qui servent à représenter une fonction arbitraire entre des limites données, *J. für Reine und Angewandte Math.* **4**, 157–169.

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

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

Gray, J. D. [1984] The shaping of the Riesz representation theorem: A chapter in the history of analysis, *Arch. Hist. Exact Sciences* **31**, 127–187.

Jackson, D. [1934] A proof of Weierstrass's theorem, *Amer. Math. Monthly* **41**, 309–312.

Kuhn, H. [1964] Ein elementarer Beweis des Weierstrassschen Approximationssatzes, *Arch. Math.* **15**, 316–317.

Landau, E. [1908] Über die Approximation einer stetigen Funktion durch eine ganze rationale Funktion, *Rend. Circ. Mat. Palermo* **25**, 337–345.

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

Lebesgue, H. [1908] Sur la représentation approchée des fonctions,, *Rend. Circ. Mat. Palermo* **26**, 325–328.

Lebesgue, H. [1909] Sur les intégrales singulières, *Ann. Fac. Sci. Univ. Toulouse* **1**, 25–117.

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.

Levasseur, K. M. [1984] A probabilistic proof of the Weierstrass approximation theorem, *Amer. Math. Monthly* **91**, 249–250.

MacTutor. [2004] `http://www-groups.dcs.st-and.ac.uk/~history`

Mittag-Leffler, G. [1900] Sur la représentation analytique des functions d'une variable réelle, *Rend. Circ. Mat. Palermo* **14**, 217–224.

Natanson, I. P. [1964] "Constructive Function Theory, Volume I", Frederick Ungar, New York.

Ostrowski, A. [1951] "Vorlesungen über Differential-und Integralrechnung, Volume II", Birkhäuser, Zurich.

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

Picard, E. [1891b] "Traité D'Analyse", Tome I, Gauthier-Villars, Paris. (Many subsequent editions followed).

Picard, E. [1899] Lectures on Mathematics, in *Clark University 1880-1899 Decennial Celebration*, W. E. Story and L. N. Wilson, eds, Norwood Press, Norwood, Mass., 207–259.

Pinkus, A. [2000] Weierstrass and Approximation Theory, *J. Approx. Theory* **107**, 1-66.

Rivlin, T. J. [1974] "The Chebyshev Polynomial", John Wiley, New York.

Runge, C. [1885] Zur Theorie der eindeutigen analytischen Functionen, *Acta Math.* **6**, 229–244.

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

Schwarz, H. A. [1871] Zur Integration der partiellen Differentialgleichung $\partial^2 u/\partial x^2 + \partial^2 u/\partial y^2 = 0$, *J. für Reine und Angewandte Math.* **74**, 218–253.

Siegmund-Schultze, R. [1988] Der Beweis des Weierstrasschen Approximationssatzes 1885 vor dem Hintergrund der Entwicklung der Fourieranalysis, *Historia Math.* **15**, 299–310.

Skrasek, J. [1960] Le centenaire de la naissance de Matyas Lerch, *Czech. Math. J.* **10**, 631–635.

Stark, E. L. [1981] Bernstein-Polynome, 1912–1955, in *Functional Analysis and Approximation*, P. L. Butzer, B. Sz.-Nagy, and E. Görlich, eds, ISNM 60, Birkhäuser, Basel, 443–461.

Sz.-Nagy, B. [1965] "Introduction to Real Functions and Orthogonal Expansions", Oxford Univ. Press, New York.

Todd, J. [1961] "Introduction to the Constructive Theory of Functions", CalTech Lecture Notes, .

de la Vallée Poussin, Ch. J. [1908] Sur l'approximation des fonctions d'une variable réelle et leurs dérivées par des polynomes et des suites limitées de Fourier, *Bull. Acad. Royale Belgique* **3**, 193–254.

de la Vallée Poussin, Ch. J. [1918] L'approximation des fonctions d'une variable réelle, *L'Enseign. Math.* **20** , 5–29.

de la Vallée Poussin, Ch. J. [1919] "Leçons sur L'Approximation des Fonctions d'une Variable Réelle ", Gauthier-Villars, Paris. Also in "L'Approximation", Chelsea, New York, 1970.

Volterra, V. [1897] Sul principio di Dirichlet, *Rend. Circ. Mat. Palermo* **11**, 83–86.

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 Weierstrass' *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, I. I. T.
Haifa, 32000
Israel
pinkus@tx.technion.ac.il
http://www.math.technion.ac.il/~pinkus