Journal of Approximation Theory **107**, 1–66 (2000)
doi:10.1006/jath.2000.3508, available online at http://www.idealibrary.com on IDEAL®

Weierstrass and Approximation Theory

Allan Pinkus

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

Communicated by Paul L. Butzer

Received November 10, 1999; accepted in revised form May 10, 2000;
published online October 18, 2000

We discuss and examine Weierstrass' main contributions to approximation theory.

1. WEIERSTRASS

This is a story about Karl Wilhelm Theodor Weierstrass (Weierstraß), what he contributed to approximation theory (and why), and some of the consequences thereof. We start this story by relating a little about the man and his life.

Karl Wilhelm Theodor Weierstrass was born on October 31, 1815, at Ostenfelde near Münster into a liberal (in the political sense) Catholic family. He was the eldest of four children, none of whom married. Weierstrass was a very successful gymnasium student and was subsequently sent by his father to the University of Bonn to study commerce and law. His father seems to have had in mind a government post for his son. However neither commerce nor law was to his liking, and he "wasted" four years there, not graduating. Beer and fencing seem to have been fairly high on his priority list at the time. The young Weierstrass returned home, and after a period of "rest", was sent to the Academy at Münster where he obtained a teacher's certificate. At the Academy he fortuitously came under the tutelage and personal guidance of C. Gudermann who was professor of mathematics at Münster and whose basic mathematical love and interest was the subject of elliptic functions and power series. This interest he was successful in conveying to Weierstrass. In 1841, Weierstrass received his teacher's certificate, and then spent the next 13 years as a teacher (for six years he was a teacher in a pregymnasium in the town of Deutsch-Krone (West Prussia), then for another seven years in a gymnasium in Braunsberg (East Prussia)). During this period he continued learning mathematics, mainly by studying the work of Abel. He also published some mathematical

0021-9045/00 $35.00

papers. However these appeared in school journals and were quite naturally not discovered at that time by any who could understand or appreciate them. (Weierstrass' collected works contain seven papers from before 1854, the first of which *On the development of modular functions* (49 pp.) was written in 1840.)

In 1854, Weierstrass published the paper *On the theory of Abelian functions* in *Crelle's Journal für die Reine und Angewandte Mathematik* (the first mathematical research journal, founded in 1826, and now referred to without Crelle's name in the formal title). It created a sensation within the mathematical community. Here was a 39 year old school teacher whom no one within the mathematical community had heard of. And he had written a masterpiece, not only in its depth, but also in its mastery of an area. Recognition was immediate. He was given a doctorate by the University of Königsberg, promoted by the Ministry of Education (of Prussia), and given a year's leave with pay. Eventually a temporary professorship was arranged for him at Berlin's Royal Polytechnic School (forerunner of the Berlin Technische Universität). Shortly thereafter he moved to the University of Berlin as an Associate Professor and was made a member of the Berlin Academy. From 1864, he was Professor of Mathematics at the University of Berlin.

There is a well-known much reproduced photograph of Weierstrass (see, for example, [60]) and in it he looks both old and tired. This is probably an unfair assessment. Weierstrass came to professional mathematics rather late in life. (In fact Weierstrass is probably the counterexample, par excellence, to the much overrated truism that mathematicians lose much of their creativity by the time they reach 40.) He was also never a healthy man from about his mid-40's. Nonetheless Weierstrass was not only very much admired and respected, but also liked. He was known as a popular, genial and approachable lecturer (a rarity at the time). In fact he was considered as one of the very best teachers of advanced students. As a consequence, but not least because he was undoubtedly one of the leading analysts of the nineteenth century, he had many formal and informal students (three of whom, Mittag–Leffler, Runge and Lerch, appear later in these pages).

Weierstrass did not publish much, and was in addition slow to publish. Nevertheless his collected works (Mathematische Werke) contain seven volumes of well over 2500 pages. However much of this Mathematische Werke is taken up with a great deal of previously unpublished lecture notes and similarly unpublished talks. Due to the nature of the material Weierstrass himself initially supervised the preparation of these volumes, and two volumes in fact appeared before his death in 1897. For further details on the life of Weierstrass see, for example, [4, 8, 16, 60] and references therein.

The areas of mathematics in which Weierstrass worked and contributed include elliptic functions, Abelian functions, the calculus of variations, the theory of analytic functions, the theory of periodic functions, bilinear and quadratic forms, differential equations and real variable function theory. Calculus students know Weierstrass' name because of the Bolzano–Weierstrass theorem, the two theorems of Weierstrass that state that every continuous real-valued function on a closed finite interval is bounded and attains its maximum and minimum, and the Weierstrass M-test for convergence of infinite series of functions. (What the students generally do not know is that Weierstrass also formulated the precise (ε, δ) definition of continuity at a point.)

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 in analysis was "nothing but power series"; see Bell [4, p. 462].

It is in this context that we should consider Weierstrass' contributions to approximation theory. In this paper we mainly consider two of Weierstrass' results. The first from 1872, see [105], is Weierstrass' example of a continuous nowhere differentiable function. It is a generally accepted fact that this was known and lectured upon by Weierstrass in 1861. Using function series (in this case cosines) Weierstrass constructs a function that is continuous but not in the least smooth. The second result, which appeared in 1885 in [107] is 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 these two papers, and these two facts, which very much lie at the heart of approximation theory.

2. CONTINUOUS NOWHERE DIFFERENTIABLE FUNCTIONS

I turn away with fear and horror from the lamentable plague of continuous functions which do not have derivatives...

—Hermite, letter to Stieltjes dated 20 May, 1893

The history of the proof of the existence of a continuous nowhere differentiable function is neither plain nor clear. Bolzano seems to have been the first to have constructed a function which is continuous but nowhere differentiable. Who was Bolzano? Bernard Placidus Johann Nepomuk Bolzano (1781–1848) was born in Prague (his father was from Italy). He was a priest and a scholar, and taught for some years at the University of Prague. However he was subsequently prohibited from teaching (and even placed for a while under house arrest) for expressing views that were not in the least acceptable to the authorities. His mathematical work went almost unnoticed and he never received the recognition he deserved until well after his death. Of course since some of it was unpublished this was not totally unwarranted. Bolzano was a contemporary of Cauchy, both chronologically and mathematically. He gave similar definitions of limits, derivatives, continuity, and convergence (see Grabiner [36]). He also made significant contributions to logic and set theory (see Bolzano [12]). Bolzano invented, sometime in the 1830's, it seems, a process for the construction of a continuous but not differentiable function. In fact, Bolzano only claimed non-existence of the derivative at a dense set of points (and such functions are very easily constructed). Nonetheless the derivative exists nowhere. This example of Bolzano was reported on by J. Masek in the early 1920's and Bolzano's manuscript containing this example was finally printed in 1930; see Kowalewski [49].

It seems to be an accepted fact (see, for example, [16, 60]) that Weierstrass gave an example of a continuous nowhere differentiable function in classroom lectures in 1861 (at the very latest). In Volume 2 of his Mathematische Werke (published in 1895) there appears the paper [105] wherein Weierstrass proves that the function

$$f(x) = \sum_{n=0}^{\infty} b^n \cos(a^n x\pi)$$

is continuous, but it is nowhere differentiable, if $b \in (0, 1)$, a is an odd integer, and $ab > 1 + (3\pi/2)$. Two facts should be stated regarding this paper. First, the paper is not a reprint of a previously published paper, but a record of a talk given to the Berlin Academy of Sciences on July 18, 1872, and it is unclear as to when exactly this "paper" was first formally written. (It finally appeared in the above Volume 2 which was published under Weierstrass' editorial supervision.) Second, in this paper Weierstrass himself specifically states that Riemann was the first to definitely assert (already in 1861 at the latest) that the infinite series

$$\sum_{n=1}^{\infty} \frac{\sin(n^2 x)}{n^2},$$

which is manifestly continuous, is not differentiable. Unfortunately it is far from evident that Riemann asserted or proved this fact. (See Ullrich [97] and Butzer, Stark [21] for a fascinating discussion of this whole question. Other sources are the many references therein, especially [69].) Work of Hardy [38] and Gerver [34, 35] eventually established the fact that this function is nondifferentiable at all but rational multiples of π where the rational number, in reduced form, is p/q with both p and q odd integers. At such points the derivative is $-1/2$. Again priority is here an issue; see Butzer, Stark [22, footnote on p. 57] and Ullrich [97, p. 246].

It is worth mentioning that there is a much simpler example of a continuous nowhere differentiable function. This example of Takagi [94] (see Yamaguti, Hata, Kigami [110, p. 11]) is given by

$$f(x) = \sum_{n=0}^{\infty} \frac{\psi(2^n x)}{2^n},$$

where $\psi(x) = \operatorname{dist}(x, \mathbb{Z})$. This example and variants thereof have often been rediscovered. When 2 is replaced by 10, this is generally referenced to van der Waerden [103] (see also Hildebrandt [39] and de Rham [77] where in "simplifying" van der Waerden's example they rediscovered the Takagi example). The proof of the nondifferentiability of this function is considered sufficiently elementary to be presented in the calculus text of Spivak [86], albeit with 10 rather than 2.

The discovery of continuous nowhere differentiable functions shocked the mathematical community. It also accentuated the need for analytic rigour in mathematics. Continuous nowhere differentiable functions may seem to some as pathological. One hundred years ago this was certainly an opinion expressed by many. (Note the quote at the beginning of this section.) Nonetheless, yesterday's pathologies are at times central in today's "cutting edge" theories and technologies. The existence of continuous nowhere differentiable functions is crucial to our proper understanding of mathematical analysis. Moreover, without nowhere differentiable functions we would not have Brownian motion, fractals, chaos, or wavelets, to mention only a few of the more popular modern theories (see e.g., [41, 63, 110] and references therein).

Continuous nowhere differentiable functions are also ubiquitous, in the sense of category. In the space of continuous functions on [0, 1] (with the uniform norm), the set of functions that at some point in [0, 1] have a one-sided derivative is of first category. That is, its complement is exceedingly large. This is one of the elegant applications of the Baire category theorem. This result, due to Mazurkiewicz [61] and generalized by Banach [3] may be found, for example, in Kuratowski [51] and in Oxtoby [71].

To return to Weierstrass' example, the first published proof of the nondifferentiability was given by du Bois-Reymond [10]. As du Bois-Reymond explicitly states, the proof of this is due to Weierstrass, and was given with his consent. It is word for word (except, it seems, for one misprint) Weierstrass' proof which "appears" in Volume 2 of his Mathematische Werke as [105].

What Weierstrass proved for the above f is that it has no derivative at every point and that

$$\lim_{h \to 0} \frac{f(x+h)-f(x)}{h}$$

does not exist, even in the generalized sense, i.e., as ∞ or $-\infty$. The major generalization of this result is due to Hardy [38]. He showed that for every $b \in (0, 1)$ and $a > 1$, the above function has no finite derivative if (and only if) $ab \geqslant 1$. Hardy also proved additional facts concerning this function and the analogous

$$g(x) = \sum_{n=0}^{\infty} b^n \sin(a^n x\pi)$$

which exhibits much the same behaviour.

Let us now present Weierstrass' proof of his result. The fact that f is continuous follows immediately from the condition $b \in (0, 1)$ and

$$|b^n \cos(a^n x\pi)| \leqslant b^n,$$

by what we now call the Weierstrass M-test for convergence.

The proof of the nondifferentiability needs a bit more work. Fix $x \in \mathbb{R}$. For each positive integer m, let α_m be an integer closest to $a^m x$, i.e., $\alpha_m \in \mathbb{Z}$, and

$$x_m = a^m x - \alpha_m$$

satisfies $|x_m| \leqslant 1/2$. Define the two sequences $\{y_m\}$ and $\{z_m\}$ via

$$y_m = \frac{\alpha_m - 1}{a^m}; \qquad z_m = \frac{\alpha_m + 1}{a^m}.$$

Then

$$x - y_m = \frac{1 + x_m}{a^m}; \qquad z_m - x = \frac{1 - x_m}{a^m},$$

and, therefore,

$$y_m < x < z_m$$

and

$$\lim_{m \to \infty} y_m = \lim_{m \to \infty} z_m = x.$$

Let us now consider

$$\begin{aligned}\frac{f(x) - f(y_m)}{x - y_m} &= \sum_{n=0}^{\infty} b^n \frac{\cos(a^n x\pi) - \cos(a^n y_m \pi)}{x - y_m} \\ &= \sum_{n=0}^{m-1} b^n \frac{\cos(a^n x\pi) - \cos(a^n y_m \pi)}{x - y_m} \\ &\quad + \sum_{n=0}^{\infty} b^{n+m} \frac{\cos(a^{n+m} x\pi) - \cos(a^{n+m} y_m \pi)}{x - y_m}.\end{aligned}$$

We estimate the first sum as follows. For $n \in \{0, 1, ..., m-1\}$, it follows from the mean-value theorem that

$$\frac{\cos(a^n x\pi) - \cos(a^n y_m \pi)}{x - y_m} = -a^n \pi \sin c_n,$$

for some $c_n \in (a^n y_m \pi, a^n x\pi)$. Thus

$$\begin{aligned}\left| \sum_{n=0}^{m-1} b^n \frac{\cos(a^n x\pi) - \cos(a^n y_m \pi)}{x - y_m} \right| &\leqslant \sum_{n=0}^{m-1} (ab)^n \pi \, |\sin c_n| \\ &\leqslant \pi \sum_{n=0}^{m-1} (ab)^n = \pi \frac{(ab)^m - 1}{ab - 1} \leqslant \pi \frac{(ab)^m}{ab - 1}.\end{aligned}$$

To estimate the second sum, note that since a is an odd integer and α_m an integer, we have

$$\cos(a^{n+m} y_m \pi) = \cos(a^n \pi(\alpha_m - 1)) = (-1)^{\alpha_m - 1}.$$

In addition,

$$x - y_m = (1 + x_m)/a^m.$$

Thus

$$\sum_{n=0}^{\infty} b^{n+m} \frac{\cos(a^{n+m}x\pi) - \cos(a^{n+m}y_m\pi)}{x - y_m}$$
$$= (-1)^{\alpha_m} (ab)^m \sum_{n=0}^{\infty} b^n \frac{(-1)^{\alpha_m} \cos(a^{n+m}x\pi) + 1}{x_m + 1}.$$

Now

$$(-1)^{\alpha_m} \cos(a^{n+m}x\pi) + 1 \geqslant 0$$

for all $n = 1, 2, ...,$ while for $n = 0$

$$(-1)^{\alpha_m} \cos(a^m x\pi) + 1 = (-1)^{\alpha_m} \cos((\alpha_m + x_m)\,\pi) + 1 = \cos(x_m\pi) + 1 \geqslant 1$$

since $|x_m\pi| \leqslant \pi/2$. In addition,

$$\tfrac{1}{2} \leqslant x_m + 1 \leqslant \tfrac{3}{2}.$$

Thus

$$\sum_{n=0}^{\infty} b^n \frac{(-1)^{\alpha_m} \cos(a^{n+m}x\pi) + 1}{x_m + 1} \geqslant \frac{\cos(x_m\pi) + 1}{x_m + 1} \geqslant \frac{2}{3}.$$

This implies that

$$(-1)^{\alpha_m} \sum_{n=0}^{\infty} b^{n+m} \frac{\cos(a^{n+m}x\pi) - \cos(a^{n+m}y_m\pi)}{x - y_m} \geqslant (-1)^{\alpha_m} (ab)^m \frac{2}{3}.$$

From the above calculations we obtain

$$\frac{f(x) - f(y_m)}{x - y_m} = \varepsilon_m \pi \frac{(ab)^m}{ab - 1} + \eta_m (-1)^{\alpha_m} (ab)^m \frac{2}{3}$$

for some ε_m, η_m satisfying $|\varepsilon_m| \leqslant 1$ and $\eta_m > 1$. We can rewrite the right-hand-side as

$$\eta_m (-1)^{\alpha_m} (ab)^m \left[\frac{\varepsilon_m}{\eta_m} \frac{(-1)^{\alpha_m}\pi}{ab - 1} + \frac{2}{3} \right],$$

where $\eta_m > 1$ and $|\varepsilon_m/\eta_m| < 1$. The condition $ab > 1 + 3\pi/2$ is equivalent to

$$\frac{2}{3} > \frac{\pi}{ab - 1}.$$

Thus for such a, b we have

$$\lim_{m\to\infty} (-1)^{\alpha_m} \frac{f(x)-f(y_m)}{x-y_m} = \infty.$$

This suffices to prove that f has no derivative at x.

An analogous argument shows that

$$\lim_{m\to\infty} (-1)^{\alpha_m+1} \frac{f(z_m)-f(x)}{z_m-x} = \infty.$$

Thus

$$\lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$$

does not exist even in a generalized sense.

Years later in 1880 Weierstrass himself used this result, presenting

$$g(z) = \sum_{n=0}^{\infty} b^n z^{a^n}$$

as a function analytic in $|z|<1$, continuous in $|z|\leqslant 1$, but whose real part is nowhere differentiable on $|z|=1$. Thus g is not continuable as an analytic function anywhere beyond $|z|<1$ (see Weierstrass [106] and Hille [40, p. 91]).

3. THE FUNDAMENTAL THEOREM OF APPROXIMATION THEORY

The basis of the theory of approximation of functions of a real variable is a theorem discovered by Weierstrass which is of great importance in the development of the whole of mathematical analysis.

—A. F. Timan [95, p. 1]

In this section we review the contents of Weierstrass [107] 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 $\tilde{C}[a, b]$ the class of functions in $C[a, b]$ satisfying $f(a)=f(b)$. ($\tilde{C}[a, b]$ 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 [107]. It was published in 1885 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 papers appeared in translation (in French) one year later in Weierstrass [108]. 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 appears in Volume 3 which originally 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 was very interested 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,

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

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

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

where

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

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. As such the following is a consequence of Theorem A and a power series argument.

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 [107] 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 $f \in C[a, b]$, $-\infty < a < b < \infty$. 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 [107], 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

$$f(x) = \sum_{m=0}^{\infty} q_m(x). \tag{3.1}$$

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 [11]. Weierstrass' theorem was considered to be by many, and by Weierstrass himself, as a "representation theorem". The theorem was seen as a means of reconciling the "analytic" and "synthetic" viewpoints which had estranged late 19th century mathematics; see Gray [37]. Much of the remaining part of Weierstrass [107] is concerned with the construction (in some sense) of a good polynomial approximant or a good representation for f (as in (3.1)). 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 [107], Weierstrass proves the density of trigonometric polynomials in $\tilde{C}[0, 2\pi]$. His proof is interesting and proceeds via complex function theory.

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

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

where

$$\omega = \int_{0}^{\infty} \psi(x)\, dx,$$

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

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

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 it is an analytic function on $\mathbb{C}\setminus\{0\}$. 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 formally state.

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

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

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

As was stated at the beginning of this section, when [107] 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 (in meaning if not verbatim).

> The main result of this paper, restricted to the one variable case, can be summarized as follows:
>
> Let $f \in C(\mathbb{R})$. Then there exists a sequence $f_1, f_2, \ldots$ of entire functions for which
>
> $$f(x) = \sum_{i=1}^{\infty} f_i(x)$$
>
> for each $x \in \mathbb{R}$. In addition the convergence of the above sum is uniform on every finite interval.

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

$$F(x_1, ..., x_n, k) = \frac{1}{2^n k^n \omega^n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(u_1, ..., u_n) \times \psi\left(\frac{u_1 - x_1}{k}\right) \cdots \psi\left(\frac{u_n - x_n}{k}\right) du_1 \cdots du_n$$

and then essentially mimicking the proofs of Theorems A and B. However Picard [72] 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.

4. EARLY ADDITIONAL PROOFS OF THE FUNDAMENTAL THEOREM

If it were necessary to designate one theorem in approximation theory as being of greater significance than any other, that one would probably be the Weierstrass approximation theorem. The influence of this theorem has been felt not only in the obvious way through its use as a tool in analysis but also in the more far-reaching way of enticing mathematicians into generalizing it or providing it with alternative proofs.

—E. W. Cheney [25, p. 190]

In this section we present various alternative proofs of Weierstrass' theorems on the density of algebraic and trigonometric polynomials on finite intervals in $\mathbb{R}$. It is our belief that the echo of these proofs have an abiding value. Some of the papers we will quote 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 which 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. We discuss some of these in Section 5. For historical consistency we have chosen to present here these proofs in more or less chronological order. This lengthens the paper, but we hope the advantages of this approach offset this deficiency.

We start by formally stating certain facts which will be obvious to most readers, but perhaps not to everyone. The first two simple statements have to do with changes 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*

$$\operatorname{span}\{1, \sin x, \cos x, \sin 2x, \cos 2x, ...\}$$

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

$$\operatorname{span}\left\{1, \sin \frac{2\pi x}{b-a}, \cos \frac{2\pi x}{b-a}, \sin 2\,\frac{2\pi x}{b-a}, \cos 2\,\frac{2\pi x}{b-a}, ...\right\}$$

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

We now show that the density of algebraic polynomials in $C[a, b]$, and trigonometric polynomials in $\tilde{C}[0, 2\pi]$, are in fact equivalent statements. That is, we prove that each of the fundamental theorems follows from the other; see also Natanson [68, p. 16–19].

PROPOSITION 3. *If trigonometric polynomials are dense in* $\tilde{C}[0, 2\pi]$, *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 [72].

Assume, without loss of generality, that $0 \leqslant a < b < 2\pi$. Extend $f \in C[a, b]$ to some $\tilde{f} \in \tilde{C}[0, 2\pi]$. Since trigonometric polynomials are dense in $\tilde{C}[0, 2\pi]$, there exists a trigonometric polynomial t that is arbitrarily close to $\tilde{f}$ 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 $\tilde{f}$. Assume $f \in C[0, 2\pi]$, and define

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

Then $g \in \tilde{C}[0, 2\pi]$. We can 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. According to de la Vallée Poussin [100, 101] the idea in this proof is due to Bernstein.

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

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

Then $g \in \tilde{C}[-\pi, \pi]$ 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

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

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 [78]). Setting

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

we have

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

for all $x \in [0, 1]$. ∎

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

Proof. The first proof of this fact was the one given by Weierstrass in Section 3. 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 [100, 101] variation on a proof by Lebesgue [53].

Let $f \in \tilde{C}[0, 2\pi]$ 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

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

for all θ.

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

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

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

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

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 the 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 Weierstrass' theorem and in 1891 in Picard [72] gave the first in a series of different proofs of the Weierstrass theorems. This proof also appears in Picard's famous textbook [73]. 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 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

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

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.

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

Now

$$\begin{aligned}
&\frac{1}{2\pi} \int_{|x-\theta| < \delta} \frac{1 - r^2}{1 - 2r\cos(x - \theta) + r^2} |f(x) - f(\theta)|\, dx \\
&\quad < \frac{\varepsilon}{2\pi} \int_0^{2\pi} \frac{1 - r^2}{1 - 2r\cos(x - \theta) + r^2}\, dx = \varepsilon.
\end{aligned}$$

In addition

$$\begin{aligned}
&\frac{1}{2\pi} \int_{\delta \leqslant |x-\theta| \leqslant \pi} \frac{1 - r^2}{1 - 2r\cos(x - \theta) + r^2} |f(x) - f(\theta)|\, dx \\
&\quad \leqslant 2\,\|f\|\, \frac{1}{2\pi} \int_{\delta \leqslant |x-\theta| \leqslant \pi} \frac{1 - r^2}{1 - 2r\cos(x - \theta) + r^2}\, dx \leqslant \frac{\|f\|\,(1 - r^2)}{r(1 - \cos\delta)}.
\end{aligned}$$

This last inequality is a consequence of

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

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

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

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

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

$$P(r,\theta) = a_0/2 + \sum_{n=1}^{\infty} r^n[a_n \cos n\theta + b_n \sin n\theta].$$

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

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

for all θ. Set

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

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 [72] states, this proof is based on an inequality obtained by H. A. Schwarz (a student of Weierstrass) in his well-known paper [84]. In fact, as Cakon [23] points out, almost the entire Picard proof can be found in Schwarz [84]. 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 [60, 85]) who attended some of Weierstrass' lectures. Lerch wrote two papers [56, 57] that included proofs of the Weierstrass theorem for algebraic polynomials. Unfortunately the paper Lerch [56] written in 1892 is in Czech, difficult to procure, and I have found no reference to it anywhere in the literature except in Lerch [57] and in a footnote in Borel [13] (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 [102], although [56], written earlier, 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 [56]. We defer the discussion of Lerch [57] 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

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

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 in 1829; see, e.g., [93, 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, due to Volterra [102], was published in 1897. V. Volterra (1860–1940) proved only the density of trigonometric polynomials in $\tilde{C}[0, 2\pi]$. As he was aware of Picard [72], this should not detract from his proof.

Volterra was unaware of Lerch [56] but his proof is much the same. Let $f \in \tilde{C}[0, 2\pi]$. 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 [64] 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, [74] in 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 [102]. Mittag-Leffler [64] points out that Weierstrass' theorem also follows from work of Runge [81, 82] in 1885 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 [64] 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 which 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:

$$g(x) = g_1(x) + \sum_{i=1}^{m-1} [g_{i+1}(x) - g_i(x)]\, h(x - x_i), \tag{4.1}$$

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

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

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

What Runge did in [82] 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 $[-1, -\delta]\cup[\delta, 1]$ 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 (4.1) 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 [82]. It was published the same year as Weierstrass' theorem.

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 [81] 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 [64] (but according to Mittag-Leffler written in 1886), remarks that apparently Runge overlooked in [82] (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 [81].

What is the relevant result from [81]? 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 $a_0, ..., a_m$, where $a_0 = w_j$ and $a_m = w$, and the a_i are chosen so that

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

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 $\operatorname{Re} G$ 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 [81]. If we go back to Runge [82] 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 $r_2 = \overline{r_1}$ is a rational function, all of whose poles 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 $p_2 = \overline{p_1}$ 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 [81], 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 $\tilde{C}[0, 2\pi]$ is more complicated and we will not present it here.

In any case, as we have seen, the Weierstrass theorem is a fairly simple consequence of Runge's results from 1885. It is unfortunate and somewhat astonishing that Runge did not independently arrive at this theorem.

Lebesgue. Let us now give Lebesgue's proof of Weierstrass' theorem from 1898 as found in Lebesgue [53]. 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 (4.1) is as a spline. That is,

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

where

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

and $ax + b = g_1(x)$. (This easily follows from the form (4.1). 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

$$g(x) = Ax + B + \sum_{i=1}^{m-1} C_i \, |x - x_i| \tag{4.2}$$

for some real constants A, B, and C_i.

Lebesgue [53] considers the form (4.2) 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 $0 \leqslant x_i \leqslant 1$). By a judicious choice of η depending on the predetermined constants C_i in (4.2), 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

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

Thus

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

where $a_1=1/2$, and

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

This power series converges absolutely and uniformly to $(1-z)^{1/2}$ in $|z|\leqslant 1$. 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 [96]. This finishes Lebesgue's proof.

An alternative argument (see Ostrowski [70, p. 168] or Feinerman, Newman [28, 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 $0<\rho<1$

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

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 $|z| \leqslant 1$.

Bourbaki [15, p. 55] (see also Dieudonné [26, 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

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

$n=0, 1, 2, \ldots$. 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 $0 \leqslant p_n(t) \leqslant \sqrt{t}$. Then

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

since $\sqrt{t}+p_n(t) \leqslant 2\sqrt{t} \leqslant 2$ 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

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

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 proof may be found in Sz.-Nagy [93, p. 77]. He considers the series of polynomials defined by $p_0(x) \equiv 1$ and

$$p_{n+1}(x)=\tfrac{1}{2}\,[\,p_n^2(x)+(1-x^2)],$$

$n=0, 1, 2, \ldots$ on $[-1, 1]$. The $\{p_n\}$ monotonically (and uniformly) decrease to $1-|x|$. (Sz.-Nagy attributes this procedure to C. Visser.)

Mittag-Leffler. Mittag-Leffler presented his own proof in 1900 in [64]. He also considers g as given in (4.1). His proof then proceeds as follows. Let

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

It is easily checked that

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

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 any $\delta > 0$, 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 (4.1) on $[-1, -\delta] \cup [\delta, 1]$ 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 (4.1) 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 and thus a grandstudent of Weierstrass. What we will report on here is taken from [29] (he had just turned 20 when the paper appeared in 1900). This fundamental paper formed the basis for Fejér's doctoral thesis obtained in 1902 from the University of Budapest. This 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 $f \in \tilde{C}[0, 2\pi]$, 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

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

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

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

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 $f \in \tilde{C}[0, 2\pi]$. 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. The functions s_m do not necessarily converge uniformly, or pointwise, to f as $m \to \infty$. However let us now set

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

Explicitly the S_n are given by

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

Surprisingly (at the time) the S_n always converge uniformly to f.

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

Proof. From the above

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

Since $f \in \tilde{C}[0, 2\pi]$, f may be considered to be uniformly continuous on all of $\mathbb{R}$. 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

$$\begin{aligned} S_n(x) - f(x) &= \frac{1}{\pi} \int_0^{2\pi} [f(y) - f(x)]\, G_n(y-x)\, dy \\ &= \frac{1}{\pi} \int_{|y-x|<\delta} [f(y) - f(x)]\, G_n(y-x)\, dy \\ &\quad + \frac{1}{\pi} \int_{\delta \leqslant |y-x| \leqslant \pi} [f(y) - f(x)]\, G_n(y-x)\, dy. \end{aligned}$$

We estimate each of the above two integrals.

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

$$\left|\frac{1}{\pi}\int_{|y-x|<\delta}[f(y)-f(x)]\,G_n(y-x)\,dy\right|<\frac{\varepsilon}{2}\frac{1}{\pi}\int_{|y-x|<\delta}G_n(y-x)\,dy$$
$$<\frac{\varepsilon}{2}\frac{1}{\pi}\int_0^{2\pi}G_n(y-x)\,dy=\frac{\varepsilon}{2}.$$

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 $|f(y)-f(x)|\leqslant 2\,\|f\|$ we have

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

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

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

For n sufficiently large

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

Thus for such n

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

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

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

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

Lerch II. The paper Lerch [57] published in 1903 contains yet another proof of the density of algebraic polynomials in $C[0, 1]$. In his previous proof, in [56], 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 [56] that it suffices to be able to arbitrarily approximate the polygonal line g as given in (4.1). Lerch rewrites (4.1) in the form

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

where

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

(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

$$\frac{1}{2} - x = \sum_{n=1}^{\infty} \frac{\sin 2n\pi x}{n\pi}, \qquad 0 < x < 1, \tag{4.3}$$

and

$$x^2 - x + \frac{1}{6} = \sum_{n=1}^{\infty} \frac{\cos 2n\pi x}{n^2\pi^2}, \qquad 0 \leqslant x \leqslant 1. \tag{4.4}$$

There is a problem with the convergence of the Fourier series in (4.3). 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 (4.4) does converge absolutely and uniformly to the given function on $[0, 1]$. 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 $x_1, ..., x_{m-1}$. There remains the problem of convergence at $x_0 = 0$ and $x_m = 1$. (However if $g \in \tilde{C}[0, 1]$, 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 (4.3) to obtain

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

(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 [57], that Lerch was aware of any of the other previous proofs of the Weierstrass theorem. A more careful consideration of this proof shows that it is essentially a quasi-constructive version of Lebesgue's proof.

Landau. The 1908 proof of E. Landau (1877–1938) in [52] 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

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

and set

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

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 $0 < \delta < 1$,

$$\int_{\delta \leqslant |u| \leqslant 1} (1-u^2)^n \, du \leqslant \int_{\delta \leqslant |u| \leqslant 1} (1-\delta^2)^n \, du < 2(1-\delta^2)^n.$$

Similarly

$$\begin{aligned} k_n = \int_{-1}^{1} (1-u^2)^n \, du &\geqslant \int_{|u| \leqslant 1/\sqrt{n}} (1-u^2)^n \, du \\ &\geqslant \int_{|u| \leqslant 1/\sqrt{n}} \left(1-\frac{1}{n}\right)^n du = \frac{2}{\sqrt{n}} \left(1-\frac{1}{n}\right)^n. \end{aligned}$$

Thus

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

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 $n \geqslant N$

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

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

$$\begin{aligned} |p_n(x) - f(x)| &= \left| \frac{1}{k_n} \int_0^1 f(y)[1 - (x - y)^2]^n \, dy - f(x) \right| \\ &\leqslant \frac{1}{k_n} \int_0^1 |f(y) - f(x)| \, [1 - (x - y)^2]^n \, dy \\ &\quad + |f(x)| \left| 1 - \frac{1}{k_n} \int_0^1 [1 - (x - y)^2]^n \, dy \right|. \end{aligned}$$

We separate the integral

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

into

$$\begin{aligned} &\frac{1}{k_n} \int_{|x - y| < \delta} |f(y) - f(x)| \, [1 - (x - y)^2]^n \, dy \\ &\quad + \frac{1}{k_n} \int_{\substack{\delta \leqslant |x - y| \\ 0 \leqslant y \leqslant 1}} |f(y) - f(x)| \, [1 - (x - y)^2]^n \, dy. \end{aligned}$$

Now

$$\begin{aligned} &\frac{1}{k_n} \int_{|x - y| < \delta} |f(y) - f(x)| \, [1 - (x - y)^2]^n \, dy \\ &\quad < \frac{\varepsilon}{3} \frac{1}{k_n} \int_{|x - y| < \delta} [1 - (x - y)^2]^n \, dy < \frac{\varepsilon}{3}. \end{aligned}$$

Furthermore

$$\frac{1}{k_n} \int_{\substack{\delta \leqslant |x-y| \\ 0 \leqslant y \leqslant 1}} |f(y) - f(x)| \, [1-(x-y)^2]^n \, dy$$

$$\leqslant \frac{2 \, \|f\|}{k_n} \int_{\delta \leqslant |u| \leqslant 1} [1-u^2]^n \, du$$

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

Finally

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

$$\leqslant \frac{\|f\|}{k_n} \left| \int_1^1 [1-u^2]^n \, du - \int_{-x}^{1-x} [1-u^2]^n \, du \right|.$$

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

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

$$\leqslant \frac{\|f\|}{k_n} \int_{\delta \leqslant |u| \leqslant 1} [1-u^2]^n \, du$$

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

This proves the result.

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

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

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 [45]. 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

$$p_n(x) = \frac{1}{k_n} \int_0^1 f(y)[1-(x-y)^2]^n \, dy$$
$$= \frac{1}{k_n} \int_{-1}^1 f(x+u)(1-u^2)^n \, du$$

and thus we get the simpler

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

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

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

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

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

Set

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

Integration by parts yields

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

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

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

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 [21] and the many references therein.

A few months after the appearance of Landau [52], Lebesgue “responded” with [54] 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 [55]). 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 [55]. We will consider this approach later in Section 5. 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. In addition to the above claim of Lebesgue, the 1908 treatise of de la Vallée Poussin [98] 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. So it seems that three outstanding mathematicians almost simultaneously discovered this method of proving Weierstrass' theorem. As Landau points out, this integral had in fact already been introduced by Stieltjes in a letter to Hermite dated September 12, 1893 (see [2]).

De la Vallée Poussin, in the second half of [98], introduced what he regarded as the periodic analogues of the Landau integrals. These are

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

where

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

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 $f \in \tilde{C}[-\pi, \pi]$ is very similar to the proof of the analogous result for the Landau integrals. 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 [20].

Bernstein. What we will arbitrarily call the last of the early proofs of the Weierstrass theorem is due to S. N. Bernstein (1880–1968) and appeared in 1912/13 in [7]. (The thesis advisor of Bernstein's first doctorate was Picard.) This paper is reproduced in Stark [87]. A translation into Russian appears in his somewhat more accessible collected works. This proof is very much 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

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

and

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

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

$$0 \leqslant \bar{f}(x) - f(x) < \frac{\varepsilon}{2},$$

and

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

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

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

From the decomposition

$$\begin{aligned} B_n(x) &= \sum_{m=0}^{n} f\left(\frac{m}{n}\right) \binom{n}{m} x^m (1-x)^{n-m} \\ &= \sum_{\{m\colon |x-(m/n)| \leqslant \delta\}} f\left(\frac{m}{n}\right) \binom{n}{m} x^m (1-x)^{n-m} \\ &\quad + \sum_{\{m\colon |x-(m/n)| > \delta\}} f\left(\frac{m}{n}\right) \binom{n}{m} x^m (1-x)^{n-m}, \end{aligned}$$

it easily follows that

$$\underline{f}(x)[1-\eta_n(x)] - \|f\|\,\eta_n(x) \leqslant B_n(x) \leqslant \bar{f}(x)[1-\eta_n(x)] + \|f\|\,\eta_n(x).$$

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

$$\begin{aligned}f(x)+[\underline{f}(x)-f(x)]-\eta_n(x)[\,\|f\|+\underline{f}(x)]\\ \leqslant B_n(x)\leqslant f(x)+[\bar{f}(x)-f(x)]+\eta_n(x)[\,\|f\|-\bar{f}(x)],\end{aligned}$$

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 [58].) 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

$$\begin{aligned}\eta_n(x)&=\sum_{\{m:\,|x-(m/n)|>\delta\}}\binom{n}{m}x^m(1-x)^{n-m}\\ &\leqslant\sum_{\{m:\,|x-(m/n)|>\delta\}}\left(\frac{x-\frac{m}{n}}{\delta}\right)^2\binom{n}{m}x^m(1-x)^{n-m}\end{aligned}$$

$$\leqslant \frac{1}{\delta^2} \sum_{m=0}^{n} \left(x - \frac{m}{n}\right)^2 \binom{n}{m} x^m (1-x)^{n-m}$$

$$= \frac{1}{\delta^2} \left[x^2 - 2x \cdot x + x^2 + \frac{x(1-x)}{n} \right]$$

$$= \frac{x(1-x)}{n\delta^2} \leqslant \frac{1}{4n\delta^2}$$

for all $x \in [0, 1]$. Thus for each fixed $\delta > 0$ we can in fact choose N such that for all $n \geqslant N$ 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 values. 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 [13, pp. 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

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

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.

5. GENERALIZATIONS AND ADDITIONAL PROOFS

A time came when there was no longer any distinction in inventing a proof of Weierstrass's theorem, unless the new method could be shown to possess some specific excellence, in the way of simplicity, for example, or rapidity of convergence.
—D. Jackson [43, p. 418]

Most great theorems are significant not only in the questions they answer, but also in their influence on the development of a field. This is particularly valid in the case of the Weierstrass approximation theorems. If we were to consider here all consequences or developments from the Weierstrass theorems, then this article would be an immense book. We will not do that. We will rather consider various results which provide different perspectives, insights and generalizations of the Weierstrass theorems. The topics we will touch upon in this section are (again in chronological order) the Müntz theorem, Hermite–Fejér interpolation, Carleman's theorem, the Stone–Weierstrass theorem, the Bohman–Korovkin theorem, and finally, a strikingly elementary proof of the Weierstrass theorem due to Kuhn.

Müntz's theorem. The three principal mathematicians who led the development of approximation theory in the early decades of the twentieth century were S. N. Bernstein, D. Jackson and Ch. J. de la Vallée Poussin. The predominant of these was undoubtedly S. N. Bernstein. In his paper Bernstein [5] in the proceedings of the 1912 International Congress of Mathematicians held at Cambridge, Bernstein wrote the following: *It will be very interesting to know if the conditions* $\sum \frac{1}{p_k}=\infty$ *are necessary and sufficient for the system* $\{x^{p_k}\}$ *to be complete. However it is not completely certain that such necessary and sufficient conditions will exist*. Bernstein also addressed this question in Bernstein [6].

The Weierstrass theorems can and should be viewed as density theorems. In fact they were the first significant density theorems. Thus it is natural to search for other "complete" systems of functions, i.e., other systems whose linear span would be dense. This is the question being posed by Bernstein, who himself had obtained some partial results.

It was just two years later in 1914 that Ch. H. Müntz (1884–1956), see [65], was able to provide a solution confirming Bernstein's qualified guess.

MÜNTZ'S THEOREM. *The system*

$$x^{p_0}, x^{p_1}, x^{p_2}, \ldots,$$

where $0 \leqslant p_0 < p_1 < p_2 < \cdots$ *is dense in* $C[0, 1]$ *if and only if* $p_0 = 0$ *and*

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

Müntz's proof of his theorem contains all the elements of the proof which may be found in many of the classic texts on approximation theory, see e.g. Achieser [1, pp. 43–46], Cheney [25, pp. 193–198], and Borwein, Erdélyi [14, pp. 171–205]. (The last reference presents many generalizations of Müntz's theorem and also surveys the literature on this topic.) Müntz's basic idea was to prove that one can approximate each x^n in $L^2[0, 1]$ ($n \in \mathbb{N}$) from the above system iff $\sum \frac{1}{p_k} = \infty$. This was done via "lemmas" due to Cauchy and to Gram. One then uses a simple trick bounding the uniform norm of the function with the $L^2[0, 1]$ norm of its derivative, and finally one applies Weierstrass' theorem. The proof in Müntz [65], although it contains all these ideas, is rather clumsy. Two years later Szász [92] generalized Müntz's results and also put Müntz's argument into a more elegant form. We mention all this in order to justify the fact that we will not reprove this result here.

An alternative proof of Müntz's theorem and its numerous generalizations is via duality and the possible sets of uniqueness for analytic functions, see e.g. Rudin [80, pp. 304–307], Luxemburg, Korevaar [59], and Feinerman, Newman [28, Chap. X]. This method of proof is not based on the Weierstrass theorems. As such it provides us with yet another proof, albeit far from simple or elementary, of the Weierstrass theorems. For some different approaches see, for example, Rogers [79], Burckel, Saeki [19].

Hermite–Fejér interpolation. Lagrange interpolation by algebraic polynomials has a long and distinguished history. One topic which has evoked much interest over the years has been the question of the convergence of the interpolation process.

To be more specific, given a triangular array $\{x_{nj}\}_{j=0}^{n}{}_{n=0}^{\infty}$ of points in $[a, b]$, $a \leqslant x_{n0} < \cdots < x_{nn} \leqslant b$, then to each $f \in C[a, b]$ and $n \in \mathbb{N}$ there exists a unique algebraic polynomial p_n of degree at most n for which

$$p_n(x_{nj}) = f(x_{nj}), \qquad j = 0, 1, \ldots, n.$$

It is natural to ask if there exists an array, as above, for which the associated polynomial sequence $\{p_n\}$ uniformly converges to f for every $f \in C[a, b]$. That is, does there exist a fixed triangular array of points for which the Weierstrass theorem follows by interpolation?

For more than a century it was known that for some reasonable triangular arrays, with more or less equally spaced points, there exist continuous functions for which the associated polynomial sequence diverges, see e.g. Méray [62] or the better known example from Runge [83]. Nonetheless it was somewhat surprising when Faber [27] proved in 1914 that for every triangular array of points there exists a continuous function for which the associated polynomial sequence diverges. (For much more on this subject see the book of Szabados and Vértesi [91].) This result of Faber should be compared with Bernstein's 1912/13 proof of the Weierstrass theorem. Bernstein constructs a polynomial of degree n based on the values of the function (defined on $[0, 1]$) at the equally spaced points $\{j/n\}$, $j = 0, 1, ..., n$. These polynomials converge uniformly to the function, but they do not interpolate the function at these points.

Based on this result of Faber, it was all the more surprising when Fejér proved in 1916 in [30] (see also [31]) that Weierstrass' theorem may be obtained via interpolation. The difference was that Fejér used Hermite interpolation rather than Lagrange interpolation. Hermite interpolation is the term applied to the generalizations of Lagrange interpolation which are based not only on function values, but also on consecutive derivative values. Fejér considered a rather specific Hermite type interpolation scheme. (He actually considered two schemes, but we will detail only one.) This interpolation scheme is today called *Hermite–Fejér interpolation.*

Let $f \in C[-1, 1]$ and $x_j = \cos(2j-1)\,\pi/2n$, $j = 1, ..., n$, be the zeros of the Chebyshev polynomial T_n of degree n (see Section 4). There exists a unique polynomial H_n of degree at most $2n-1$ which satisfies

$$\begin{aligned} H_n(x_j) &= f(x_j), \qquad j = 1, ..., n \\ H_n'(x_j) &= 0, \qquad j = 1, ..., n. \end{aligned} \tag{5.1}$$

The following is contained in [30, Theorem XI].

HERMITE–FEJÉR INTERPOLATION THEOREM. *For every $f \in C[-1, 1]$ the sequence of polynomials H_n, as defined above, converges uniformly to f on $[-1, 1]$.*

Proof. For any given distinct $\{x_j\}_{j=1}^n$, set

$$\omega(x) = \prod_{j=1}^{n} (x - x_j)$$

and

$$\ell_i(x) = \frac{\omega(x)}{\omega'(x_i)(x - x_i)}. \tag{5.2}$$

The ℓ_i are (fundamental) polynomials of exact degree $n-1$ that satisfy

$$\ell_i(x_j) = \delta_{ij}, \qquad i, j = 1, ..., n.$$

For each $i = 1, ..., n$, set

$$h_i(x) = [1 - 2(x - x_i)\, \ell_i'(x_i)](\ell_i(x))^2.$$

It is readily verified from (5.2) (and L'Hôpital's rule) that

$$h_i(x) = \left[1 - \frac{\omega''(x_i)(x - x_i)}{\omega'(x_i)}\right] (\ell_i(x))^2. \tag{5.3}$$

Each h_i is a polynomial of exact degree $2n-1$, and

$$h_i(x_j) = \delta_{ij}, \qquad j = 1, ..., n$$

$$h_i'(x_j) = 0, \qquad j = 1, ..., n.$$

Thus the polynomial of degree at most $2n-1$

$$H_n(x) = \sum_{i=1}^{n} f(x_i)\, h_i(x)$$

satisfies (5.1).

We now assume, as in the statement of the theorem, that $x_j = \cos(2j-1)\pi/2n$, $j = 1, ..., n$. Then $\omega(x) = aT_n(x)$ for some known constant a. (ω is monic, while T_n is normalized to have norm one.) In this case we show how we can further refine formula (5.3) for the h_i. The polynomial ω satisfies the second order differential equation

$$(1 - x^2)\, \omega''(x) - x\omega'(x) + n^2\omega(x) = 0$$

(see e.g. Rivlin [78, p. 31]). At the points x_i we have $\omega(x_i) = 0$ and therefore

$$(1 - x_i^2)\, \omega''(x_i) = x_i\omega'(x_i)$$

and

$$\frac{\omega''(x_i)}{\omega'(x_i)} = \frac{x_i}{(1 - x_i^2)}. \tag{5.4}$$

Furthermore it is easily verified from the formula $T_n(x) = \cos(n \arccos x)$ that

$$\omega'(x_i) = aT'_n(x_i) = \frac{an(-1)^i}{\sqrt{1-x_i^2}}.$$

Thus from (5.2)

$$(\ell_i(x))^2 = \left[\frac{\omega(x)}{\omega'(x_i)(x-x_i)}\right]^2 = \frac{(1-x_i^2)}{(x-x_i)^2}\frac{T_n^2(x)}{n^2}. \tag{5.5}$$

Substituting (5.4) and (5.5) into (5.3) we obtain

$$h_i(x) = \frac{(1-xx_i)}{(x-x_i)^2}\frac{T_n^2(x)}{n^2}.$$

Note that since $|x_i| < 1$ it follows that

$$h_i(x) \geqslant 0 \tag{5.6}$$

for all $x \in [-1, 1]$. Furthermore

$$\sum_{i=1}^{n} h_i(x) \equiv 1 \tag{5.7}$$

since the right hand side is the unique polynomial of degree at most $2n-1$ which assumes the value 1 and has derivative 0 at each x_j, $j = 1, ..., n$.

Finally, before proving the convergence result, we note that since $|T_n(x)| \leqslant 1$ and $|1 - xx_i| \leqslant 2$ for all $x \in [-1, 1]$, we have the inequality

$$h_i(x) \leqslant \frac{2}{n^2(x-x_i)^2} \tag{5.8}$$

for all $x \in [-1, 1]$ and $i = 1, ..., n$.

The remaining steps of the convergence proof are now similar to what we have seen in previous proofs. Given $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $x, y \in [-1, 1]$ satisfying $|x - y| < \delta$ we have $|f(x) - f(y)| < \varepsilon$. Now from (5.7)

$$f(x) - H_n(x) = \sum_{i=1}^{n} [f(x) - f(x_i)]\, h_i(x).$$

We divide the sum on the right hand side into

$$\sum_{\{i:\, |x-x_i|<\delta\}} [f(x)-f(x_i)]\, h_i(x) + \sum_{\{i:\, |x-x_i|\geqslant\delta\}} [f(x)-f(x_i)]\, h_i(x).$$

Applying both (5.6) and (5.7) we have

$$\left| \sum_{\{i:\, |x-x_i|<\delta\}} [f(x)-f(x_i)]\, h_i(x) \right| < \varepsilon \sum_{\{i:\, |x-x_i|<\delta\}} h_i(x) \leqslant \varepsilon \sum_{i=1}^{n} h_i(x) = \varepsilon.$$

We estimate the second sum by an application of (5.8).

$$\begin{aligned} \left| \sum_{\{i:\, |x-x_i|\geqslant\delta\}} [f(x)-f(x_i)]\, h_i(x) \right| &\leqslant 2\, \|f\| \left[\sum_{\{i:\, |x-x_i|\geqslant\delta\}} h_i(x) \right] \\ &\leqslant \frac{4\, \|f\|}{n^2} \sum_{\{i:\, |x-x_i|\geqslant\delta\}} \frac{1}{(x-x_i)^2} \\ &\leqslant \frac{4\, \|f\|}{n^2} \frac{n}{\delta^2} = \frac{4\, \|f\|}{n\delta^2}. \end{aligned}$$

Choosing n sufficiently large it follows that

$$|f(x) - H_n(x)| < 2\varepsilon$$

for all $x \in [-1, 1]$. ∎

Carleman's Theorem. In 1927 T. Carleman (1892–1949), see [24], proved a direct generalization of Weierstrass' original Theorem A (see Section 3). This result would have undoubtedly pleased Weierstrass. It is the following.

CARLEMAN'S THEOREM. *Let $\eta \in C(\mathbb{R})$, $\eta(x) > 0$ for all x. To each $f \in C(\mathbb{R})$ there exists an entire function g for which*

$$|f(x) - g(x)| < \eta(x)$$

for all $x \in \mathbb{R}$.

Proof. In what follows we assume $z \in \mathbb{C}$ and $x \in \mathbb{R}$. Furthermore, let $\alpha_0 > \alpha_1 > \cdots$ satisfy

$$0 < \alpha_n < \min_{n \leqslant |x| \leqslant n+1} \eta(x)$$

and

$$\beta_n = \alpha_{n+1} - \alpha_{n+2}$$

$n = 0, 1, 2, ...,$ and $\beta_{-1} = 0$.

We will construct a sequence of polynomials $\{p_n\}$ in the following manner. The polynomial p_0 is chosen, by Weierstrass' theorem, to satisfy

$$|f(x) - p_0(x)| < \beta_0$$

for $|x| \leqslant 1$. Now set

$$h_1(z) = p_0(z), \quad |z| \leqslant 1 \qquad \text{and} \qquad h_1(x) = f(x), \quad 3/2 \leqslant |x| \leqslant 2$$

and extend h_1 to $\{x\colon 1 < |x| < 3/2\}$ so that it is continuous on $\{x\colon 1 \leqslant |x| \leqslant 2\}$ and also satisfies

$$|f(x) - h_1(x)| < \beta_0$$

on $\{x\colon 1 \leqslant |x| \leqslant 3/2\}$. This is possible. Set

$$A_1 = \{z\colon |z| \leqslant 1\} \cup \{x\colon 1 \leqslant |x| \leqslant 2\}.$$

By a theorem of Walsh [104, p. 47, Theorem 15] (a Runge type theorem) the function h_1 can be uniformly approximated on A_1 by polynomials. Thus there exists a polynomial p_1 satisfying

$$|h_1(z) - p_1(z)| < \beta_1$$

for all $z \in A_1$.

The general form of the construction is the following. For $n \in \mathbb{N}$ set

$$A_n = \{z\colon |z| \leqslant n\} \cup \{x\colon n \leqslant |x| \leqslant n+1\}.$$

Assume we have chosen the polynomial p_n satisfying

$$|f(\pm(n+1)) - p_n(\pm(n+1))| < \beta_n.$$

Set

$$h_{n+1}(z) = \begin{cases} p_n(z), & |z| \leqslant n+1 \\ f(x), & n+3/2 \leqslant |x| \leqslant n+2 \end{cases}$$

and extend h_{n+1} to $\{x\colon n+1 < |x| < n+3/2\}$ so that it is continuous on A_{n+1} and also satisfies

$$|f(x) - h_{n+1}(x)| < \beta_n$$

on $\{x: n+1 \leqslant |x| \leqslant n+3/2\}$. This is possible. By the above-mentioned theorem of Walsh, there exists a polynomial p_{n+1} satisfying

$$|h_{n+1}(z) - p_{n+1}(z)| < \beta_{n+1}$$

for all $z \in A_{n+1}$ (and thus our "assumption" also holds at $\pm(n+2)$).

Let

$$g(z) = \lim_{n \to \infty} p_n(z) = p_0(z) + \sum_{k=0}^{\infty} [p_{k+1}(z) - p_{k+2}(z)].$$

We claim that g is an entire function which satisfies the claim of the theorem.

The function g is entire since

$$|p_{n+1}(z) - p_n(z)| < \beta_{n+1}$$

on $\{z: |z| \leqslant n+1\}$, and $\sum_{k=0}^{\infty} \beta_k < \infty$.

To prove the approximation property note that on $\{x: n \leqslant |x| \leqslant n+1\}$ we have

$$|f(x) - p_n(x)| \leqslant |f(x) - h_n(x)| + |h_n(x) - p_n(x)| < \beta_{n-1} + \beta_n$$

(which also holds for $n=0$ since we have set $\beta_{-1}=0$). Furthermore, on $\{x: n \leqslant |x| \leqslant n+1\}$

$$|g(x) - p_n(x)| = \left| \sum_{k=n}^{\infty} [p_{k+1}(z) - p_k(z)] \right| < \sum_{k=n}^{\infty} \beta_{k+1}.$$

Thus on $\{x: n \leqslant |x| \leqslant n+1\}$ we have

$$|f(x) - g(x)| \leqslant |f(x) - p_n(x)| + |p_n(x) - g(x)| < \sum_{k=n-1}^{\infty} \beta_k.$$

Recalling that $\beta_k = \alpha_{k+1} - \alpha_{k+2}$, $k = 0, 1, 2, ...$, and $\beta_{-1}=0$, it follows that

$$\sum_{k=n-1}^{\infty} \beta_k \leqslant \begin{cases} \alpha_n, & n \geqslant 0 \\ \alpha_1 < \alpha_0, & n = 0 \end{cases}$$

and thus

$$|f(x) - g(x)| < \eta(x)$$

for all $x \in \mathbb{R}$. ∎

We have presented here a variation on Carleman's original proof, although the basic structure of the proof is much the same. The major difference is that Carleman does not reference Walsh, but constructs the desired $\{p_n\}$. The proof as given here may be found in Kaplan [46]. He ascribes it to Marcel Brelot. This is essentially the same proof as appears in Gaier [33], where numerous extensions are discussed. In addition, in Carleman's original formulation η was taken as a positive constant. However from the method of proof it easily follows that the positive constant can be replaced by any η as above. Note that η can tend to 0 as $x \to \pm\infty$.

H. Whitney, in his seminal paper [109] on the analytic extension of differentiable functions, proves an extension of this result of Carleman to open sets in $\mathbb{R}^n$ and also simultaneously approximates the function and any finite set of derivatives. Narasimhan [67, p. 34] contains an elegant proof along the lines of both Whitney's proof and Weierstrass' original proof. Unfortunately Whitney's paper contains no reference to Carleman. As a consequence there seem to have been two streams of papers which discuss and generalize these results, each stream referencing one author but not the other. Frih and Gauthier [32] have some interesting extensions to both results.

Stone–Weierstrass theorem. In [88], written in 1937, M. H. Stone (1903–1989) generalized Weierstrass' theorem proving a result which, as stated in Buck [18, p. 4], *represents 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 of the original theorem and various proofs have been given. See, for example, Nachbin [66] and Prolla [75], and references therein. Stone himself reworked relevant portions of [88] in [89], which was reprinted in the more accessible Stone [90]. According to Stone, the proof in [89, 90], was *much improved by Kakutani, with the aid of suggestions made by Chevalley.* (He is referring to the double compactness argument given below.) The importance of the theorem and the insight it provides into the Weierstrass theorems is such that we feel it imperative that we present and prove a form of this theorem here. Our proof will follow closely the essential ideas contained in [89, 90].

THEOREM. *Let X be a compact space and let $C(X)$ be the space of continuous real-valued functions defined on X. Assume A is a subalgebra of $C(X)$ which contains the constant function and separates points. Then A is dense in $C(X)$ in the uniform norm.*

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 $x, y \in C(X)$ there exists a g in the set for which $g(x) \neq g(y)$.

Proof. First some preliminaries. From the Weierstrass theorem, or more explicitly from Lebesgue's proof thereof and its variations as given in Section 4, there exists a sequence of algebraic polynomials $\{p_n\}$ which uniformly approximates the function $|t|$ on $[-c, c]$, every $c > 0$. As such, if f is in $\bar{A}$, the closure of A in the uniform norm, then so is $p_n(f)$ for each n which implies that $|f|$ is also in $\bar{A}$. 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 $f, g \in \bar{A}$, then $\max\{f, g\}$ and $\min\{f, g\}$ are also in $\bar{A}$. 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 $\alpha, \beta \in \mathbb{R}$. There exists a $g \in A$ for which $g(x) \neq g(y)$, and the constant function is also in A. Thus

$$h(w) = \beta + (\alpha - \beta)\,\frac{g(w) - g(y)}{g(x) - g(y)}$$

is in A and satisfies the interpolation conditions $h(x) = \alpha$ and $h(y) = \beta$.

We can now present a proof of this 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 neighbourhood V_y of y for which $h_y(w) \geqslant f(w) - \varepsilon$ for all $w \in V_y$. The $\bigcup_{y \in X} V_y$ cover X. As X is a compact metric space, it has a finite subcover, i.e., there are points $y_1, ..., y_n$ in X such that

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

Let $g = \max\{h_{y_1}, ..., h_{y_n}\}$. Then $g \in \bar{A}$ and $g(w) \geqslant f(w) - \varepsilon$ 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 $g_x(w) \geqslant f(w) - \varepsilon$ for all $w \in X$. As f and g_x are continuous there exists a neighbourhood U_x of x for which $g_x(w) \leqslant f(w) + \varepsilon$ for all

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

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

Let

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

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

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

for all $w \in X$. Thus

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

This implies that $f \in \bar{A}$. ∎

Bohman–Korovkin theorem. As we noted in Section 4, many of the proofs contained therein are based on sequences of singular integrals, and in fact on positive singular integrals. In his famous treatise [55] of 1909, Lebesgue considered the subject of singular integrals. This paper was largely motivated by the various above-mentioned proofs. As Lebesgue states in [54] in reference to the methods of proof of Weierstrass, Picard, Fejér and Landau: ...*the study of these diverse integrals is done by the same process and evidently depends on those properties relative to singular integrals of positive functions.*

The paper [55] is lengthy and contains many diverse results on integrals, different forms of convergence of sequences of singular integrals, and upper and lower bounds on the orders of approximation by various approximation processes. With respect to convergence of sequences of singular integrals, it is perhaps easiest to formulate some of these main concepts in the periodic case.

THEOREM. *Assume that for each $n \in \mathbb{N}$ we have $K_n \in \tilde{C}[-\pi, \pi]$, $K_n(y) \geqslant 0$ for all $y \in [-\pi, \pi]$, and*

$$\int_{-\pi}^{\pi} K_n(y)\, dy = 1.$$

Further assume that for every $\delta > 0$

$$\lim_{n\to\infty} \int_{\delta<|y|\leqslant\pi} K_n(y)\,dy = 0.$$

For each $f \in \tilde{C}[-\pi, \pi]$ *set*

$$I_n(f; x) = \int_{-\pi}^{\pi} f(y)\,K_n(x-y)\,dy.$$

Then

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

and the convergence is uniform on $[-\pi, \pi]$.

The proof of this result is elementary. We have essentially proven it repeatedly in this and the previous section.

If K_n is a trigonometric polynomial, as in the proofs of Fejér and de la Vallée Poussin, then I_n is a trigonometric polynomial and this immediately implies Weierstrass' theorem. The singular integral of Jackson (contained in his thesis [42] and also in the more accessible [44])

$$J_n(f; x) = \int_{-\pi}^{\pi} f(y)\,j_n(x-y)\,dy,$$

where

$$j_n(y) = a_n \left[\frac{\sin(ny/2)}{n\sin(y/2)}\right]^4$$

with a_n chosen so that j_n integrates to 1, is another example thereof. If $I_n(f; x)$ is either a polynomial or suitably analytic, in which case it can be replaced by a truncated power series approximant, then we also obtain the Weierstrass theorem. The proofs of Weierstrass and Landau fall within these categories. This framework and these results can also be generalized to include Bernstein's proof and the proof via Hermite–Fejér interpolation.

In the above we sought conditions verifying that a sequence of singular integrals appropriately approaches the identity. Positive singular integrals give rise to positive linear operators. It so happens that there are easily checked properties guaranteeing the convergence of a sequence of positive linear operators to the identity operator. The major result in this context is the following which can be applied to simplify many of the methods of proof of the previous section.

BOHMAN–KOROVKIN THEOREM. *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

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

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

Since the L_n are positive linear operators, this implies that

$$(L_n\, p_\ell)(x) \leqslant (L_n f)(x) \leqslant (L_n\, p_u)(x) \tag{5.9}$$

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 which 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 in $[a, b]$, it follows that there exists an N such that for all $n \geqslant N$, and every choice of $y \in [a, b]$

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

and similarly

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

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 $n \geqslant N$ and every $y \in [a, b]$ we have from (5.9)

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

This proves the theorem. ▌

H. Bohman (1920–1996) was a Swedish actuary and statistician. In [9], published in 1952, he proved the above mentioned result but only for positive linear operators of the form

$$(L_n\, f)(x) = \sum_{i=0}^{n} f(\xi_{i,n})\, \psi_{i,n}(x),$$

where the $\psi_{i,n}$ are non-negative functions, and the points $\xi_{i,n}$ are in $[a, b]$, $i = 0, 1, ..., n$. His proof, and the main idea of his approach, was a generalization of Bernstein's proof of the Weierstrass theorem (see Section 4).

P. P. Korovkin (1913–1985) one year later proved the same theorem for integral type operators. Korovkin's original proof, as found in [47], is based on positive singular integrals (à la Lebesgue). Korovkin was probably unaware of Bohman's result. Korovkin subsequently much extended his theory, major portions of which can be found in [48]. The proof we have presented here is taken from [48].

Kuhn's proof. There are many elegant proofs of Weierstrass' theorem. For those comfortable with either power series or Fourier series or singular integrals, then the previous sections contain many simple proofs. But perhaps the most elementary proof (of which we are aware) is Kuhn [50]. Kuhn's proof uses one basic inequality, namely Bernoulli's inequality

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

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

We present Kuhn's proof except that we save a step by recalling from Section 4 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

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

This form was used in the proofs of Runge/Phragmén, Mittag-Leffler and 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. They may be given as

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

(Note that the polynomials $\{x[2p_n(x) - 1]\}$ uniformly converge to $|x|$ on $[-1, 1]$. See Lebesgue's proof as given in Section 4.)

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

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

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

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

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

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

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

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

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 the proof.

Kuhn's proof motivated Brosowski and Deutsch [17] and subsequently Ransford [76], to provide elementary proofs of the Stone–Weierstrass Theorem.

ACKNOWLEDGMENTS

Numerous people have helped by either locating or translating source material or by constructive criticism, or any combination thereof. I thank J. M. Almira, Hubert Berens, Carl de Boor, Bruno Brosowski, Frank Deutsch, Christine Kellogg, Vera Kurkova, Vladimir Lin, Rolf Nessel, Paul Nevai, Lee Rudolf, Robert Schaback, Karl Scherer, Itai Shafrir, and Bronislaw Wajnryb, and apologise to anyone I may have inadvertently left off the list.

REFERENCES

1. N. I. Achieser, "Theory of Approximation," Ungar, New York, 1956. [Originally published in Russian in 1947.]
2. B. Baillaud and H. Bourget, "Correspondance d'Hermite et de Stieltjes," Gauthier-Villars, Paris, 1905.
3. S. Banach, Über die Baire'sche Kategorie gewisser Funktionenmengen, *Stud. Math.* **3** (1931), 174–179.
4. E. T. Bell, "Men of Mathematics," Scientific, London, 1936.
5. S. N. Bernstein, Sur les recherches récentes relatives à la meilleure approximation des fonctions continues par les polynomes, *Proc. of 5th Inter. Math. Congress* **1** (1912), 256–266. [Also appears in Russian translation in Bernstein's Collected Works.]
6. S. N. Bernstein, Sur l'ordre de la meilleure approximation des fonctions continues par les polynomes de degré donné, *Mém. Cl. Sci. Acad. Roy. Belg.* **4** (1912), 1–103.
7. S. N. Bernstein, Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités, *Comm. Soc. Math. Kharkow* **13** (1912/13), 1–2. [Also appears in Russian translation in Bernstein's Collected Works.]
8. K. R. Biermann, Karl Weierstrass, *in* "Dictionary of Scientific Biography" (C. C. Gillispie, Ed.), Scribner's, New York, 1976.
9. H. Bohman, On approximation of continuous and of analytic functions, *Ark. Mat.* **2** (1952), 43–56.
10. P. du Bois-Reymond, Versuch einer Classification der willkürlichen Functionen reeller Argumente nach ihren Aenderungen in den kleinsten Intervallen, *J. Reine Angew. Math.* **79** (1875), 21–37.
11. P. du Bois-Reymond, Untersuchungen über die Convergenz und Divergenz der Fourierschen Darstellungsformeln, *Abh. Math.-Phys. Classe K. Bayerische Akad. Wissenschaften* **12** (1876), 1–13.
12. B. Bolzano, "Paradoxes of the Infinite," Routledge and Kegan Paul, London, 1950.
13. É. Borel, "Lecons sur les Fonctions de Variables Réelles et les Développements en Séries de Polynomes," Gauthier-Villars, Paris, 1905. [2nd edition, 1928]
14. P. Borwein and T. Erdélyi, "Polynomials and Polynomial Inequalities," Springer-Verlag, New York, 1995.
15. N. Bourbaki, "Topologie Générale (Livre III). Espaces Fonctionnels Dictionnaire (Chapitre X)," Hermann & Cie, Paris, 1949.
16. C. B. Boyer and U. C. Merzbach, "A History of Mathematics," 2nd ed., Wiley, New York, 1989.
17. B. Brosowski and F. Deutsch, An elementary proof of the Stone–Weierstrass theorem, *Proc. Amer. Math. Soc.* **81** (1981), 89–92.
18. R. C. Buck, "Studies in Modern Analysis," Vol. 1, Mathematical Association of America, Washington, D.C., 1962.
19. R. B. Burckel and S. Saeki, An elementary proof of the Müntz–Szász theorem, *Expo. Math.* **4** (1983), 335–341.
20. P. L. Butzer and R. J. Nessel, Aspects of de la Vallée Poussin's work in approximation and its influence, *Arch. Hist. Exact Sci.* **46** (1993), 67–95.
21. P. L. Butzer and E. L. Stark, 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" (P. T. Bateman, L. Mirsky, H. L. Montgomery, W. Schall, I. J. Schoenberg, W. Schwarz, and H. Wefelscheid, Eds.), Vol. 3, pp. 83–111, Thales-Verlag, Essen, 1986.

22. P. L. Butzer and E. L. Stark, “Riemann's Example” of a continuous nondifferentiable function in the light of two letters (1865) of Christoffel to Prym, *Bull. Soc. Math. Belgique* **38** (1986), 45–73.
23. R. Cakon, “Alternative Proofs of Weierstrass Theorem of Approximation: An Expository Paper,” Master's Thesis, Department of Mathematics, The Pennsylvania State University, 1987.
24. T. Carleman, Sur un théorème de Weierstrass, *Ark. Mat., Ast. Fysik B* **20** (1927), 1–5.
25. E. W. Cheney, “Introduction to Approximation Theory,” McGraw–Hill, New York, 1966.
26. J. Dieudonné, “Foundations of Modern Analysis,” Academic Press, New York, 1969.
27. G. Faber, Über die interpolatorische Darstellung stetiger Funktionen, *Jahresber. Deut. Math. Verein* **23** (1914), 190–210.
28. R. P. Feinerman and D. J. Newman, “Polynomial Approximation,” Williams and Wilkins, Baltimore, 1974.
29. L. Fejér, Sur les fonctions bornées et intégrables, *C.R. Heb. Séances Acad. Sci. Paris* **131** (1900), 984–987.
30. L. Fejér, Ueber Interpolation, *Nachr. Gesell. Wiss. Göttingen Math. Phys. Kl.* (1916), 66–91.
31. L. Fejér, Über Weierstrasssche Approximation, besonders durch Hermitesche Interpolation, *Math. Ann.* **102** (1930), 707–725.
32. E. M. Frih and P. M. Gauthier, Approximation of a function and its derivatives by entire functions of several variables, *Canad. Math. Bull.* **31** (1988), 495–499.
33. D. Gaier, “Lectures on Complex Approximation,” Birkhäuser, Boston, 1987.
34. J. Gerver, The differentiability of the Riemann function at certain rational multiples of π, *Amer. J. Math.* **92** (1970), 33–55.
35. J. Gerver, More on the differentiability of the Riemann function, *Amer. J. Math.* **93** (1970), 33–41.
36. J. V. Grabiner, “The Origins of Cauchy's Rigorous Calculus,” MIT Press, Cambridge, MA, 1981.
37. J. D. Gray, The shaping of the Riesz representation theorem: A chapter in the history of analysis, *Arch. Hist. Exact Sci.* **31** (1984), 127–187.
38. G. H. Hardy, Weierstrass's non-differentiable function, *Trans. Amer. Math. Soc.* **17** (1916), 301–325.
39. T. H. Hildebrandt, A simple continuous function with a finite derivative at no point, *Amer. Math. Monthly* **40** (1933), 547–548.
40. E. Hille, “Analytic Function Theory,” Vol. II, Ginn and Co., Boston, 1962.
41. B. R. Hunt, The Hausdorff dimension of graphs of Weierstrass functions, *Trans. Amer. Math. Soc.* **126** (1998), 791–800.
42. D. Jackson, “Über die Genauigkeit der Annäherung stetiger Funktionen durch ganz rationale Funktionen gegebenen Grades und trigonometrische Summen gegebener Ordnung,” Preisschrift und Dissertation, Univ. Göttingen, 1911.
43. D. Jackson, The general theory of approximation by polynomials and trigonometric sums, *Bull. Amer. Math. Soc.* **27** (1921), 415–431.
44. D. Jackson, “The Theory of Approximation,” Amer. Math. Soc., Colloquium Publ., Vol. XI, New York, 1930.
45. D. Jackson, A proof of Weierstrass's theorem, *Amer. Math. Monthly* **41** (1934), 309–312.
46. W. Kaplan, Approximation by entire functions, *Michigan Math. J.* **3** (1955/56), 43–52.
47. P. P. Korovkin, On convergence of linear positive operators in the space of continuous functions, *Dokl. Akad. Nauk SSSR* **90** (1953), 961–964.
48. P. P. Korovkin, “Linear Operators and Approximation Theory,” Hindustan, Delhi, 1960. [The Russian original appeared in 1959.]

49. G. Kowalewski, Über Bolzanos nichtdifferenzierbare stetige Funktion, *Acta Math.* **44** (1923), 315–319.
50. H. Kuhn, Ein elementarer Beweis des Weierstrassschen Approximationssatzes, *Arch. Math.* **15** (1964), 316–317.
51. C. Kuratowski, “Topologie,” Vol. 1, 4th ed., Panstwowe Wydawnictwo Naukowe, Warszawa, 1958.
52. E. Landau, Über die Approximation einer stetigen Funktion durch eine ganze rationale Funktion, *Rend. Circ. Mat. Palermo* **25** (1908), 337–345.
53. H. Lebesgue, Sur l’approximation des fonctions, *Bull. Sci. Math.* **22** (1898), 278–287.
54. H. Lebesgue, Sur la représentation approchée des fonctions, *Rend. Circ. Mat. Palermo* **26** (1908), 325–328.
55. H. Lebesgue, Sur les intégrales singulières, *Ann. Fac. Sci. Univ. Toulouse* **1** (1909), 25–117.
56. M. Lerch, About the main theorem of the theory of generating functions (in Czech), *Rozpravy Ceske Akad.* **33** (1892), 681–685.
57. M. Lerch, Sur un point de la théorie des fonctions génératrices d’Abel, *Acta Math.* **27** (1903), 339–351.
58. K. M. Levasseur, A probabilistic proof of the Weierstrass approximation theorem, *Amer. Math. Monthly* **91** (1984), 249–250.
59. W. A. J. Luxemburg and J. Korevaar, Entire functions and Müntz–Szász type approximation, *Trans. Amer. Math. Soc.* **157** (1971), 23–37.
60. MacTutor, Available at http://www-groups.dcs.st-and.ac.uk/history.
61. S. Mazurkiewicz, Sur les fonctions non dérivables, *Stud. Math.* **1** (1929), 92–94.
62. C. Méray, Nouveaux exemples d’interpolations illusoires, *Bull. Sci. Math.* **20** (1986), 266–270.
63. Y. Meyer, “Wavelets: Algorithms and Applications,” Chap. 9, SIAM, Philadelphia, 1993.
64. G. Mittag-Leffler, Sur la représentation analytique des functions d’une variable réelle, *Rend. Circ. Mat. Palermo* **14** (1900), 217–224.
65. C. H. Müntz, Über den Approximationssatz von Weierstrass, *in* “H. A. Schwarz’s Festschrift,” pp. 303–312, Berlin, 1914.
66. L. Nachbin, “Elements of Approximation Theory,” Krieger, Huntington, NY, 1976.
67. R. Narasimhan, “Analysis on Real and Complex Manifolds,” North-Holland, Amsterdam, 1968.
68. I. P. Natanson, “Constructive Function Theory,” Vol. I, Frederick Ungar, New York, 1964.
69. E. Neuenschwander, Riemann’s example of a continuous ‘nondifferentiable’ function, *Math. Intelligencer* **1** (1978), 40–44.
70. A. Ostrowski, “Vorlesungen über Differential-und Integralrechnung,” Vol. II, Birkhäuser, Basel, 1951.
71. J. C. Oxtoby, “Measure and Category,” GTM 2, 2nd ed., Springer-Verlag, New York, 1986.
72. E. Picard, Sur la représentation approchée des fonctions, *C. R. Heb. Séances Acad. Sci. Paris* **112** (1891), 183–186.
73. E. Picard, “Traité D’Analyse,” Tome I, Gauthier-Villars, Paris, 1891. [Many subsequent editions followed.]
74. E. Picard, Lectures on mathematics, *in* “Clark University 1880–1899 Decennial Celebration” (W. E. Story and L. N. Wilson, Eds.), pp. 207–259, Norwood Press, Norwood, MA, 1899.
75. J. B. Prolla, “Weierstrass–Stone, the Theorem,” Peter Lang, Frankfurt, 1993.
76. T. J. Ransford, A short elementary proof of the Bishop–Stone–Weierstrass theorem, *Math. Proc. Cambridge Philos. Soc.* **96** (1984), 309–311.

77. G. de Rham, Sur un exemple de fonction continue sans dérivée, *Enseign. Math.* **3** (1957), 71–72.
78. T. J. Rivlin, "The Chebyshev Polynomial," Wiley, New York, 1974.
79. L. C. G. Rogers, A simple proof of Müntz's theorem, *Math. Proc. Cambridge Philos. Soc.* **90** (1981), 1–3.
80. W. Rudin, "Real and Complex Analysis," McGraw–Hill, New York, 1966.
81. C. Runge, Zur Theorie der eindeutigen analytischen Functionen, *Acta Math.* **6** (1885), 229–244.
82. C. Runge, Über die Darstellung willkürlicher Functionen, *Acta Math.* **7** (1885/86), 387–392.
83. C. Runge, Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten, *Zeit. Math. Physik* **46** (1901), 224–243.
84. H. A. Schwarz, Zur Integration der partiellen Differentialgleichung $\partial^2 u/\partial x^2 + \partial^2 u/\partial y^2 = 0$, *J. Reine Angew. Math.* **74** (1871), 218–253.
85. J. Skrasek, Le centenaire de la naissance de Matyas Lerch, *Czech. Math. J.* **10** (1960), 631–635.
86. M. Spivak, "Calculus," 3rd ed., Publish or Perish, 1994.
87. E. L. Stark, Bernstein–Polynome, 1912–1955, *in* "Functional Analysis and Approximation" (P. L. Butzer, B. Sz.-Nagy, and E. Görlich, Eds.), pp. 443–461, ISNM 60, Birkhäuser, Basel, 1981.
88. M. H. Stone, Applications of the theory of Boolean rings to general topology, *Trans. Amer. Math. Soc.* **41** (1937), 375–481.
89. M. H. Stone, A generalized Weierstrass approximation theorem, *Math. Magazine* **21** (1948), 167–184, 237–254.
90. M. H. Stone, A generalized Weierstrass approximation theorem, *in* "Studies in Modern Analysis" (R. C. Buck, Ed.), Vol. 1, pp. 30–87, Mathematical Association of America, Washington, D.C., 1962. [This is a reprint of [89].]
91. J. Szabados and P. Vértesi, "Interpolation of Functions," World Scientific, Singapore, 1990.
92. O. Szász, Über die Approximation stetiger Funktionen durch lineare Aggregate von Potenzen, *Math. Ann.* **77** (1916), 482–496.
93. B. Sz.-Nagy, "Introduction to Real Functions and Orthogonal Expansions," Oxford Univ. Press, New York, 1965.
94. T. Takagi, A simple example of a continuous function without derivative, *Proc. Physico-Math. Soc. Japan* **1** (1903), 176–177.
95. A. F. Timan, "Theory of Approximation of Functions of a Real Variable," Pergamon Press, Oxford, 1963.
96. J. Todd, "Introduction to the Constructive Theory of Functions," CalTech Lecture Notes, 1961.
97. P. Ullrich, Anmerkungen zum "Riemannschen Beispiel" $\sum_{n=1}^{\infty} \frac{\sin n^2 x}{n^2}$ einer stetigen, nicht differenzierbaren Funktion, *Result. Math.* **31** (1997), 245–265.
98. Ch. J. de la Vallée Poussin, 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** (1908), 193–254.
99. Ch. J. de la Vallée Poussin, Sur la meilleure approximation des fonctions d'une variable réelle par des expressions d'ordre donné, *C.R. Acad. Sci. Paris* **166** (1918), 799–802.
100. Ch. J. de la Vallée Poussin, L'approximation des fonctions d'une variable réelle, *L'Enseign. Math.* **20** (1918), 5–29.
101. Ch. J. de la Vallée Poussin, "Leçons sur L'Approximation des Fonctions d'une Variable Réelle," Gauthier-Villars, Paris, 1919. [Also in "L'Approximation", Chelsea, New York, 1970.]

102. V. Volterra, Sul principio di Dirichlet, *Rend. Circ. Mat. Palermo* **11** (1897), 83–86.
103. B. L. van der Waerden, Ein einfaches Beispiel einer nicht-differenzierbaren stetigen Funktion, *Math. Z.* **32** (1930), 474–475.
104. J. L. Walsh, "Interpolation and Approximation by Rational Functions in the Complex Domain," Amer. Math. Soc., Providence, RI, 1935.
105. K. Weierstrass, Über continuierliche Functionen eines reellen Arguments, die für keinen Werth des letzteren einen bestimmten Differentialquotienten besitzen, Königliche Akademie der Wissenschaften, 18 Juli 1872. [Also in "Mathematische Werke," Vol. 2, pp. 71–74, Mayer & Müller, Berlin, 1895.]
106. K. Weierstrass, Zur Functionenlehre, Monatsber. Königl. Akad. Wiss., 1880. [Also in "Mathematische Werke," Vol. 2, pp. 210–223, Mayer & Müller, Berlin, 1895.]
107. K. Weierstrass, Über die analytische Darstellbarkeit sogenannter willkürlicher Functionen einer reellen Veränderlichen, Sitzungsber. Akad. Berlin, 1885, pp. 633–639, 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.]
108. K. Weierstrass, Sur la possibilité d'une représentation analytique des fonctions dites arbitraires d'une variable réelle, *J. Math. Pure Appl.* **2** (1886), 105–113, 115–138. [This is a translation of [107] 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*.]
109. H. Whitney, Analytic extensions of differentiable functions defined in closed sets, *Trans. Amer. Math. Soc.* **36** (1934), 63–89.
110. M. Yamaguti, M. Hata, and J. Kigami, "Mathematics of Fractals," AMS Transl. Math. Monographs, Amer. Math. Soc., Providence, RI, 1997.