Foundations of Mathematical Analysis, Master Program, UAM

Notes by Dragan Vukotić, 2022-23

Week 1: The Weierstrass approximation theorem

The great German mathematician Karl T. Weierstrass (1815-1897) is generally considered as the father of rigorous Mathematical Analysis, having created the $\varepsilon-\delta$ language and other formal definitions around 1860. In this chapter we will prove the basic approximation theorem discovered by him as late as 1885. This is one of the basic results in a vast area of Mathematics nowadays known as Approximation Theory. We note in passing that, just a few years later, his student Carl Runge (1856-1927) proved an important theorem on uniform approximation of holomorphic functions by polynomials on compact sets, relevant in Complex Analysis. However, this result will not be discussed here as it belongs to a completely different topic.

Notation. Throughout these lecture notes, $\mathbb{N}=\{1,2,3,\dots\}$ will denote the set of positive integers (also called natural numbers), $\mathbb{Q}$ the set of all rational numbers, $\mathbb{R}$ the set of all real numbers, and $\mathbb{C}$ the set of all complex numbers. Uniform convergence of the functions f_n to a function f on a set A will be denoted by writing $f_n \rightrightarrows_A f$.

Statement of Weierstrass' theorem and some consequences

In this section, $[a,b]$ will denote a compact (closed and bounded) interval in $\mathbb{R}$. In other words, it will be assumed that $-\infty<a<b<\infty$. (General compact sets will be reviewed briefly later on.) The space of all real-valued and continuous functions in $[a,b]$ will be denoted by $C[a,b]$. It is clearly a vector space over the reals $\mathbb{R}$ and, moreover, an algebra (meaning that the product of any two funcions in the space remains in it). Also, a natural norm can be defined on $C[a,b]$ by setting

$$\|f\|_\infty=\sup_{x\in[a,b]}|f(x)|=\max_{x\in[a,b]}|f(x)|$$

(we know from Calculus that the maximum is attained since $|f|\in C[a,b]$ as well). The convergence in this norm is the well-known uniform convergence; in other words, $\|f_n-f\|_\infty\to 0$ as $n\to\infty$ if and only if $f_n \rightrightarrows_{[a,b]} f$ (The readers are invited to check this simple equivalence.) Formally, this means that

$$\forall\varepsilon>0\ \exists N\in\mathbb{N}\ \forall n\ge N\ \forall x\in[a,b]\ |f_n(x)-f(x)|<\varepsilon\,.$$

(Just like in other similar definitions, the $<$ sign in the last inequality above can be replaced by $\le$ without harm.) The concepts of uniform convergence, metric spaces and compact sets may be reviewed briefly in an additional set of auxiliary notes for this course for those.

We say that a function f *can be uniformly approximated by polynomials* in $[a,b]$ if there exists a sequence of polynomials $(P_n)_{n=1}^{\infty}$ such that $P_n \rightrightarrows_{[a,b]} f$. Equivalently, this can be formulated by saying that for each $\varepsilon > 0$ there exists a polynomial P such that $\|P - f\| < \varepsilon$.

Theorem 1. *(Weierstrass). Every real-valued function $f \in C[a,b]$ can be uniformly approximated in $[a,b]$ by polynomials with real coefficients.*

The proof will be deferred for the time being. We first want to make a few comments about some relevant applications of this fundamental result.

Moments of a function. By definition, the *moments* of a function $f \in C[a,b]$ are the numbers

$$M_n(f) = \int_a^b x^n f(x)\,dx\,, \qquad n \in \{0,1,2,\ldots\}.$$

They can be defined in a similar fashion for integrable functions (in some sense) on other appropriate sets on the real line.

We encounter them, for example, in Probability Theory: if X is continuous random variable with density function f, we define its *expectation* and *variance*, respectively, as

$$E[X] = \int_{\mathbb{R}} x f(x)\,dx\,, \quad V(X) = E[X^2] - E[X]^2 = \int_{\mathbb{R}} x^2 f(x)\,dx - \left(\int_{\mathbb{R}} x f(x)\,dx\right)^2.$$

That is,

$$E[X] = M_1(f)\,, \quad V(X) = M_2(f) - M_1(f)^2.$$

In different areas of Mathematical Analysis and Probability we can find questions called *moment problems.* In their different formulations, such problems consist in finding out how much information we need about the moments of a given function (or a measure) in order to be able to determine completely the function (or measure) in question. The following simple statement was a pioneering result in this direction.

Corollary 1. *(Hausdorff). If $f \in C[a,b]$ is a real-valued function whose moments are all zero:*

$$\int_a^b x^n f(x)\,dx = 0\,, \quad \forall n \in \{0,1,2,\ldots\},$$

then $f \equiv 0$ in $[a,b]$. Thus, if two continuous functions have the same moments, they are identically equal in $[a,b]$.

PROOF. Taking finite linear combinations of integrals of above type, we get that $\int_a^b P(x) f(x)\,dx = 0$ for every polynomial P with real coefficients. By Weierstrass' Theorem 1, f is a uniform limit of a certain sequence $(P_n)_n$ of polynomials in $[a,b]$. Since a continuous function is bounded on $[a,b]$, this readily implies that

$$P_n f \rightrightarrows_{[a,b]} f^2, \quad n \to \infty.$$

(Just estimate the difference $P_n f - f^2$ uniformly in $[a,b]$ to check this fact.) Due to uniform convergence, by a well-known theorem from Calculus we are allowed to exchange the limit and the integral, thus obtaining

$$\int_a^b f(x)^2\,dx = \lim_{n\to\infty} \int_a^b P_n(x) f(x)\,dx = 0.$$

Since $f^2 \in C[a,b]$ and $f(x)^2 \geq 0$ for all $x \in [a,b]$, we conclude that $f(x)^2 \equiv 0$ in $[a,b]$. (If you did not see a formal proof of this fact in your Analysis courses, it may be a good exercise to prove the statement.) Note that this is one place where we use in an essential way the assumptions that f is continuous and only takes real values. It follows immediately that $f(x) \equiv 0$ in $[a,b]$. ■

Separability of $C[a,b]$. The theorem of Weierstrass has another important consequence. Recall that a metric spaces is called *separable* if it contains a dense and countable set. For example, $\mathbb{R}$ (with its usual distance) is separable since it contains $\mathbb{Q}$ as a dense and countable subset. At this point, in addition to some metric/topological concepts, the reader is also expected to review the basic concepts and results from Set Theory needed here.

Corollary 2.. *The space $C[a,b]$, equipped with the usual norm, is separable.*

PROOF. Following a standard argument, we will show that the polynomials with rational coefficients form a dense countable set in $C[a,b]$.

First, note that the polynomials with coefficients in $\mathbb{Q}$ form a countable set. This can be shown as follows. First, for any fixed $n \in \{0,1,2,\ldots\}$, the set of all polynomials $Q(x) = b_0 + b_1 x + \ldots + b_n x^n$ of degree n, with b_0, b_1, $\ldots$, $b_{n-1} \in \mathbb{Q}$, $b_n \in \mathbb{Q}\setminus\{0\}$, has the same number of elements as $\mathbb{Q}^n \times (\mathbb{Q}\setminus\{0\})$. The latter set is countable because any finite Cartesian product of countable sets is countable. Then, finally, the set of all polynomials with coefficients in $\mathbb{Q}$ is countable by being a countable union of countable sets (by splitting the set of all polynomials into constant polynomials, linear polynomials, quadratic polynomials, and so on).

Next, we show that every polynomial with real coefficients can be uniformly approximated by polynomials with rational coefficients. Let $\varepsilon > 0$ and $P(x) = a_0 + a_1 x + \ldots + a_n x^n$ be an arbitrary polynomial with a_0, a_1,$\ldots$, $a_n \in \mathbb{R}$. Then we can find a polynomial $Q(x) = b_0 + b_1 x + \ldots + b_n x^n$ with b_0, b_1, $\ldots$, $b_n \in \mathbb{Q}$ and such that

$$|a_k - b_k||x|^k < \frac{\varepsilon}{n+1}, \quad \forall x \in [a,b], \quad \forall k \in \{0,1,\ldots,n\},$$

by the density of the rationals in $\mathbb{R}$ and also since $|x|^k$ can be bounded uniformly in $[a,b]$ (for example, by the constant $\max\{|a|^k, |b|\}^k$) for each k. It follows immediately that

$$|P(x) - Q(x)| \leq \sum_{k=0}^{n} |a_k - b_k||x|^k < (n+1)\frac{\varepsilon}{n+1} = \varepsilon, \quad \forall x \in [a,b].$$

Finally, it is easy to see that every function $f \in C[a,b]$ can be uniformly approximated by polynomials with rational coefficients. Given $\varepsilon > 0$, by Theorem 1 we can find a polynomial P with real coefficients with $\|P - f\|_\infty < \varepsilon/2$. Next, we know that we can also find a polynomial Q with rational coefficients such that $\|Q - P\|_\infty < \varepsilon/2$. From the triangle inequality for the norm we obtain

$$\|Q - f\|_\infty \le \|Q - P\|_\infty + \|P - f\|_\infty < \varepsilon .$$

which completes the proof. ■

Lebesgue's proof of Weierstrass' theorem: a sketch

We only outline an idea of a very direct proof due to the famous French mathematician Henry L. Lebesgue (1875-1941), the father of the Lebesgue integral.

His proofs begins by observing that any $f \in C[a,b]$ is uniformly continuous in $[a,b]$ by a theorem of Cantor from Calculus, which allows us to approximate f uniformly in $[a,b]$ by piecewise linear functions. These are functions whose graph is a polygonal path connecting finitely many points on the graph of f. Essentially, all such functions can be decomposed into pieces that look more or less like a letter "V" or an inverted letter "V". The pieces of this kind are obtained by a linear function from the function $u(x) = |x|$. An important point is to note that

$$|x - c| + (x - c) = \begin{cases} 0, & \text{if } x \le c \\ 2(x - c), & \text{if } x \ge c . \end{cases}$$

The problem of approximating f uniformly by polynomials then reduces to approximating uniformly by polynomials certain linear combinations of functions of the above type. The bottom line is that one needs to approximate uniformly by polynomials the the function $u(x) = |x|$ in the interval $[-1,1]$. (Why is the interval irrelevant will be explained in the next section on Landau's proof.)

Thus, the entire proof boils down to finding a sequence of polynomials that converges uniformly to u in $[-1,1]$ and this is done by extending and making more precise a well-known generalized binomial formula from Calculus:

$$|x| = \sqrt{x^2} = (1 - (1 - x^2))^{1/2} = \sum_{n=0}^{\infty} \binom{1/2}{n} (-1)^n (1 - x^2)^n , \quad -1 \le x \le 1 .$$

It can be proved that the partial sums of the above series (obviously, polynomials in x) converge to $|x|$ uniformly in $[-1,1]$. The result from Calculus is usually formulated only for the open interval $(-1,1)$ but holds also on the closed interval and, moreover, it can be proved that the convergence is uniform.

Further details can be consulted in Duren's book mentioned in the references (Sections 6.2 and 3.5; alternatively, to prove uniform convergence, one can use the Monotone convergence theorem from the Theory of Measure and Integration, to be reviewed later). The reader is advised to draw pictures and check carefully every detail.

Landau's proof of Weierstrass' theorem

In this section we describe in detail the proof of the theorem of Weierstrass given by Landau (Edmund G. Landau, 1877-1938, a well-known German number theorist and complex analyst). We begin by reviewing various relevant details and simplifications.

Reduction to a special case: the exact interval does not matter. We first observe that the theorem is equivalent to some of its special cases.

First reduction. Proving Theorem 1 for the space $C[a,b]$ (with arbitrary $-\infty < a < b < \infty$) is equivalent to proving it for $C[0,1]$; hence, it suffices to consider only the latter case. The reason is that, starting from an arbitrary interval $[a,b]$, we can establish a bijective linear map

$$L : [0,1] \to [a,b], \quad L(t) = a + (b-a)t\,.$$

Obviously, L is continuous and $Q(t)$ is a polynomial of t with real coefficients if and only if $Q(L(t)) = Q(a+(b-a)t)$ is one, thanks to the Newton binomial formula. Hence, $f \in C[a,b]$ can be approximated by polynomials uniformly in $[a,b]$ if and only if $f \circ L$ can be uniformly approximated by polynomials in $[0,1]$.

Second reduction. Instead of proving Theorem 1 for all functions in $C[0,1]$, it suffices to prove it only for those functions in $C[0,1]$ that satisfy the additional condition $f(0) = f(1) = 0$. Indeed, given $f \in C[0,1]$, define the function g by

$$g(x) = f(x) - f(0) - x(f(1) - f(0))\,.$$

Clearly, $g(0) = 0 = g(1)$ and $g \in C[0,1]$ since $g - f$ es a linear function. Moreover, if a sequence of polynomials (P_n) converges to g, then the sequence of polynomials

$$Q_n(x) = P_n(x) + f(0) + x(f(1) - f(0))\,,$$

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

Approximate identity (summability kernel). By the previous discussion, if Weierstrass' theorem is valid for one compact interval, it is also valid for any other such interval (passing from $[0,1]$ to that interval). As we shall see later, it is convenient to work in symmetric intervals $[-a,a]$, where $0 < a < \infty$.

Landau's proof is based only on a few properties that are usually required from what is called an approximate identity or a summability kernel. The argument can be adapted to other proofs in the context of approximation (uniforme, pointwise, or in some mean value). We will likely see examples of this in Fourier Analysis.

Definition. A sequence of functions $Q_n : [-a,a] \to \mathbb{R}$ is called an *approximate identity* (or *summability kernel*) if it satisfies the following conditions:

(1) $Q_n(t) \ge 0$, for all $n \in \mathbb{N}$ and every $t \in [-a,a]$;

(2) $\int_{-a}^{a} Q_n(t)\,dt = 1$, for all $n \in \mathbb{N}$;

(3) $\forall \delta \in (0,a)$, $Q_n \rightrightarrows 0$ in $\{x : \delta \le |x| \le a\} = [-a,-\delta] \cup]\delta, a]$.

Sometimes it is convenient to require an additional condition:

(4) Q_n is an even function in $[-a,a]$.

Later on we will explain why is the last condition useful and we will see that there exist different relevant families of functions that have the properties of an approximate identity, among them the Fejér and Poisson kernels.

Proposition 1. *Let $f \in C[0,a]$, $f(0) = f(a) = 0$, and extend f continuously to the rest of the real line by defining $f(x) = 0$ in $(-\infty,0) \cup (a,\infty)$. Let $(Q_n)_n$ be an approximate identity in $[-a,a]$. If we define the functions P_n according to the formula*

$$P_n(x) = \int_{-a}^{a} f(x+t)Q_n(t)\,dt, \tag{1}$$

then $P_n \rightrightarrows f$ in $[0,a]$.

Proof. By Cantor's theorem, f is uniformly continuous in $[0,a]$. Thus,

$$\forall \varepsilon > 0\ \exists \delta > 0\ \forall x, y \in [0,a]\ \ |x-y| < \delta \Rightarrow |f(x) - f(y)| < \frac{\varepsilon}{3}.$$

Since $f(x) = 0$ in $(-\infty,0) \cup (a,\infty)$, we can actually conclude that

$$\forall \varepsilon > 0\ \exists \delta > 0\ \forall x, y \in \mathbb{R}\ \ |x-y| < \delta \Rightarrow |f(x) - f(y)| < \frac{\varepsilon}{3}.$$

(For example, if $|x-y| < \delta$ and $x \in [0,a]$ but $y > a$, then $|x-a| < \delta$, hence $|f(x) - f(y)| = |f(x) - f(a)| < \frac{\varepsilon}{3}$. The discussion is similar if $|x-y| < \delta$, $x < 0$ and $y \in [0,a]$, etc.)

Let $\varepsilon > 0$. By the previous discussion, with the value of δ chosen as above (depending on ε), if $t \in (-\delta,\delta)$, $x \in \mathbb{R}$, we obtain that

$$|(x+t) - x| = |t| < \delta \Rightarrow |f(x+t) - f(x)| < \frac{\varepsilon}{3}.$$

For the same value of δ, recalling that $Q_n \geq 0$ by property (1) of approximate identities, for all $x \in [0,a]$ we have

$$\begin{aligned} |P_n(x) - f(x)| &= \left| \int_{-a}^{a} f(x+t)Q_n(t)\,dt - \int_{-a}^{a} f(x)Q_n(t)\,dt \right| \leq \int_{-a}^{a} |f(x+t) - f(x)|Q_n(t)\,dt \\ &= \int_{[-a,-\delta]} + \int_{(-\delta,\delta)} + \int_{[\delta,a]}, \end{aligned}$$

the integrand being repeated in each integral in the last line. By what was said earlier, we estimate the second of the three integrals as follows:

$$\int_{(-\delta,\delta)} |f(x+t) - f(x)|Q_n(t)\,dt \leq \frac{\varepsilon}{3} \int_{(-\delta,\delta)} Q_n(t)\,dt \leq \frac{\varepsilon}{3} \int_{-a}^{a} Q_n(t)\,dt = \frac{\varepsilon}{3},$$

using property (2) of approximate identities.

Let $M = \max\{|f(x)| : x \in [0,a]\}$; note that it exists by continuity of f and compactness of $[0,a]$). Then

$$|f(x+t) - f(x)| \le 2M, \quad \text{para todo } x, t \in \mathbb{R}.$$

Property (3) of approximate identities allows us to intechange the limit and the integral to obtain

$$\lim_{n\to\infty} \int_{-a}^{-\delta} Q_n(t)\,dt = 0 = \lim_{n\to\infty} \int_{\delta}^{a} Q_n(t)\,dt.$$

Thus, given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that for all $n \ge N$ we have

$$\int_{-a}^{-\delta} Q_n(t)\,dt < \frac{\varepsilon}{6M}, \quad \int_{\delta}^{a} Q_n(t)\,dt < \frac{\varepsilon}{6M}.$$

Hence,

$$\int_{-a}^{-\delta} |f(x+t) - f(x)| Q_n(t)\,dt < 2M \cdot \frac{\varepsilon}{6M} = \frac{\varepsilon}{3}, \qquad \int_{\delta}^{a} |f(x+t) - f(x)| Q_n(t)\,dt < \frac{\varepsilon}{3}.$$

Finally, putting together the three estimates with $\frac{\varepsilon}{3}$, we obtain that $|P_n(x) - f(x)| < \varepsilon$, for all $n \ge N$. ■

The above Proposition 1 essentially tells us that convolutions of a function with an approximate identity converge uniformly to the given function. However, the reader familiar with convolutions may recognize that the definition of P_n given above does not exactly coincide with that of a convolution. So why do we talk about convolutions? Here is the reason.

Observation. If we assume the additional condition (4): the function Q_n is even, then P_n is exactly the convolution of f with Q_n (in the sense of the usual definition seen in mathematical texts). This can be seen as follows.

Note that $f(x+t) \ne 0 \Rightarrow 0 \le x+t \le a \Leftrightarrow -x \le t \le a-x$, meaning that in the definition of P_n the interval of integration gets reduced to $[-x, a-x]$ since the integrand is zero in the rest. After the simple change of variable $x+t = s$, using the fact that $Q_n(s-x) = Q_n(x-s)$, we conclude that

$$P_n(x) = \int_{-x}^{a-x} f(x+t) Q_n(t)\,dt = \int_0^a f(s) Q_n(s-x)\,ds = \int_0^a Q_n(x-s) f(s)\,ds = (Q_n * f)(x),$$

which corresponds exactly to the usual definition of a convolution of two functions.

Observation. If each $Q_n(x)$ is a polynomial with real coefficients, then so is $P_n(x)$.

This is also easy to check. If $Q_n(x)$ is a polynomial, then $Q_n(x-s)$ is also such for every value of s. Computing the integrals that define P_n (integrating respect to the variable t), we see that $P_n(x)$ is also a polynomial with real coefficients.

Theorem 2. *(Bernoulli's inequality). If $0 \le a \le 1$, then $(1-a)^n \ge 1 - na$, for every positive integer n.*

PROOF. Various simple proofs are possible (see if you can give your own!), one of them by induction on n, for example.

Another typical way of proving the inequality is by using Calculus as follows. Let

$$u(a) = (1-a)^n - (1-na), \quad 0 \le a \le 1.$$

It is immediate that

$$u'(a) = n - n(1-a)^{n-1} \ge 0, \forall a \in [0,1],$$

so it follows that the function u is non-decreasing in $[0,1]$. Hence $u(a) \ge u(0) = 0$ for all $a \in [0,1]$. ■

Now we are finally ready to give a proof of the theorem of Weiestrass, using the tools developed above and following Landau's idea.

PROOF. By the two reductions seen earlier, it suffices to consider only the functions $f \in C[0,1]$ such that $f(0) = f(1) = 0$. Any such function can be extended to a continuous function on all of $\mathbb{R}$ by defining its values on the set $(-\infty,0) \cup (1,\infty)$ to be 0 (drawing a picture will be helpful here). This means that we can apply the case $a = 1$ of Proposition 1 already proved. To this end, we need only show the existence of approximate identity consisting of even polynomials. Consider the following even polynomials:

$$Q_n(x) = c_n(1-x^2)^n,$$

where the constants $c_n > 0$ are chosen (after integration) in such a way that $\int_{-1}^{1} Q_n(x)\,dx = 1$. Obviously, we have $Q_n(x) \ge 0$ in $[-1,1]$. In order to see that the sequence $(Q_n)_n$ is an approximate identity, we need only check its uniform convergence to zero on the set $\{x : \delta \le |x| \le 1\}$. Starting from

$$\frac{1}{c_n} = \int_{-1}^{1} (1-x^2)^n\,dx = 2\int_0^1 (1-x^2)^n\,dx \ge 2\int_0^{1/\sqrt{n}} (1-x^2)^n\,dx \ge 2\int_0^{1/\sqrt{n}} (1-nx^2)\,dx = \frac{4}{3\sqrt{n}} > \frac{1}{\sqrt{n}},$$

(using Lemma 2 in the second inequality), it follows that $c_n < \sqrt{n}$. Hence, for $\delta \le |x| \le 1$ we obtain

$$0 \le Q_n(x) \le \sqrt{n}\,(1-x^2)^n \le \sqrt{n}\,(1-\delta^2)^n.$$

The upper bound in the last inequality does not depend on x as long as $\delta \le |x| \le 1$. It is well known from Calculus that, for any fixed $r \in (0,1]$, the sequence $\sqrt{n}r^n$ tends to zero as $n \to \infty$. Taking $r = 1-\delta^2$, we obtain that $Q_n(x) \rightrightarrows 0$ on $\{x : \delta \le |x| \le 1\}$.

Now we can apply Proposition 1 to conclude that f can be approximated uniformly by the functions P_n which, in this case, are polynomials since the functions Q_n are polynomials. ■

Recommended bibliography for this lecture

- P.L. Duren: *Invitation to Classical Analysis*, American Mathematical Society, Providence, Rhode Island 2012 (Chapter 6).

- W. Rudin: *Principles of Mathematical Analysis*, McGraw-Hill, Nueva York 1976. Third edition (Chapter 7).