Some approximation theorems in Math 522

Prelude: Basic facts and formulas for the partial sum operator for Fourier series.

Consider the partial sums of the Fourier series

\[ S_n f(x) = \sum_{k=-n}^{n} c_k e^{ikx} \]

where \( c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)e^{-ikt} \, dt \) are the Fourier coefficients.

We can write

\[ S_n f(x) = \sum_{k=-n}^{n} \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y)e^{-iky} \, dy e^{ikx} \]

\[ = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y)D_n(x-y) \, dy \quad \text{where } D_n(t) = \sum_{k=-n}^{n} e^{ikt}. \]

**Definition.** The convolution of two \( 2\pi \) periodic functions \( f, g \) is defined as

\[ f \ast g(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(y)g(x-y) \, dy. \]

Note that the convolution of \( 2\pi \) periodic continuous functions is well defined and is again a \( 2\pi \)-periodic continuous function. and we also have the commutativity property

\[ f \ast g(x) = g \ast f(x) \]

To see we first note that for a \( 2\pi \) periodic integrable function we have

\[ \int_{-\pi}^{\pi} F(t) \, dt = \int_{a-\pi}^{a+\pi} F(t) \, dt \]

for any \( a \). The commutativity property follows if in the definition of \( f \ast g \) we change variables \( t = x-y \) (with \( dt = -dy \)) and get

\[ 2\pi f \ast g(x) = \int_{-\pi}^{\pi} f(y)g(x-y) \, dy = \int_{-\pi}^{\pi} f(x-t)g(t)(-1) \, dt \]

\[ = \int_{-\pi}^{\pi} f(x-t)g(t) \, dt = \int_{-\pi}^{\pi} g(t)f(x-t) \, dt = 2\pi g \ast f(x) \]

where in the last formula we have used the \( 2\pi \)-periodicity of \( f \) and \( g \).

Going back to the partial sum of the Fourier series we have

\[ S_n f(x) = f \ast D_n(x) = D_n \ast f(x) \quad \text{where } D_n(t) = \sum_{k=-n}^{n} e^{ikt}. \]

Below we will need a more explicit expression for \( D_n \), namely

\[ D_n(t) = \frac{\sin(n + \frac{1}{2})t}{\sin \frac{t}{2}} \]
To see this we use $\sum_{k=0}^{n} e^{ikt} = \frac{e^{i(n+1)t} - 1}{e^{it} - 1}$ and $\sum_{k=-n}^{-1} e^{ikt} = \sum_{k=1}^{n} e^{-ikt} = \frac{e^{-i(n+1)t} - 1}{e^{-it} - 1} - 1$ and the second sum can be simplified to $\frac{e^{-int} - 1}{1 - e^{it}}$. Thus $D_n(t) = \frac{e^{i(n+1)t} - e^{-i(n+1)/2}t}{e^{it/2} - e^{-it/2}}$. Multiplying numerator and denominator with $e^{-it/2}$ yields $D_n(t) = e^{i(n+1)t/2} - e^{-i(n+1)/2}t e^{it/2} - e^{-it/2}$ and this yields the displayed formula.

I. Fejér’s theorem

We would like to prove that every continuous function can be approximated by trigonometric polynomials, uniformly on $[-\pi, \pi]$. One may think that, in view of Theorem 8.11 in Rudin’s book, the partial sums $S_n f$ of the Fourier series are good candidates for such an approximation. Unfortunately for merely continuous $f$, given $x$, the partial sums $S_n f(x)$ may not converge to $f(x)$ (and then of course $S_n f$ cannot converge uniformly).\footnote{The situation is even worse. Given $x \in [-\pi, \pi]$ one can show that in a certain sense the convergence of $S_n f(x)$ fails for typical $f$. I hope to return to this point later in the class.}

However instead of $S_n f$ we consider the better behaved arithmetic means (or Cesàro means) of the partial sums. Define

$$\sigma_N f(x) = \frac{1}{N+1} \sum_{n=0}^{N} S_n f(x).$$

The means $\sigma_N f$ are also called the Fejér means of the Fourier series, in tribute to the Hungarian mathematician Leopold Fejér who in 1900 published the following

**Theorem.** Let $f$ be a continuous $2\pi$-periodic function. Then the means $\sigma_N f$ converge to $f$ uniformly, i.e.

$$\max_{x \in \mathbb{R}} |\sigma_N f(x) - f(x)| \to 0, \text{ as } N \to \infty.$$ 

If we use the convolution formula $S_n f = D_n * f$ then it follows that

$$\sigma_N f(x) = K_N * f(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} K_N(x - y) f(y) dy = \frac{1}{2\pi} \int_{-\pi}^{\pi} K_N(t) f(x - t) dt$$

where

$$K_N(t) = \frac{1}{N+1} \sum_{n=0}^{N} D_n(t)$$

$K_N$ is called the $N$th Fejér kernel.

\footnote{Why is one allowed to write max here for sup?}
We need the following properties of \( K_N \).

**Lemma.** (a) Explicit formulas for \( K_N \) on \([-\pi, \pi]\) are given by

\[
K_N(x) = \frac{1}{N+1} \frac{1 - \cos(N+1)x}{1 - \cos x} = \frac{1}{N+1} \left( \frac{\sin(\frac{N+1}{2}x)}{\sin \frac{x}{2}} \right)^2,
\]

if \( x \) is not an integer multiple of \( 2\pi \). Also \( K_N(0) = N+1 \).

(b) \( K_N(x) \geq 0 \) for all \( x \geq 0 \).

(c) \[
\frac{1}{2\pi} \int_{-\pi}^{\pi} K_N(t)dt = 1.
\]

(d) \[
K_N(x) \leq \frac{1}{N+1} \left( \frac{2}{1 - \cos \delta} \right) \text{ for } 0 < \delta \leq x \leq \pi.
\]

By (c), (d) most of \( K_N \) is concentrated near 0 for large \( N \). Properties (b), (c), (d) are important, the explicit expressions for \( K_N \) much less so.

**Proof of the Lemma.** We use and rewrite the above explicit formula for the Dirichlet kernel namely

\[
D_n(x) = \frac{\sin(n + \frac{1}{2}x)}{\sin \frac{x}{2}} = \frac{\sin \frac{x}{2} \sin(n + \frac{1}{2})x}{\sin^2 \frac{x}{2}}.
\]

Observe that \( 2 \sin a \sin b = \cos(a - b) - \cos(a + b) \) and apply this with \( a = (n + \frac{1}{2})x, b = \frac{x}{2} \) to get

\[
D_n(x) = \frac{\cos nx - \cos(n + 1)x}{2 \sin^2 \frac{x}{2}}.
\]

Thus

\[
K_N(x) = \frac{1}{N+1} \sum_{n=0}^{N} D_n(x)
\]

\[
= \frac{1}{N+1} \sum_{n=0}^{N} \frac{\cos nx - \cos(n + 1)x}{2 \sin^2 \frac{x}{2}}
\]

\[
= \frac{1}{N+1} \frac{1 - \cos(N + 1)x}{2 \sin^2 \frac{x}{2}}
\]

Now recall the formula \( \cos 2a = \cos^2 a - \sin^2 a = 1 - 2 \sin^2 a \). If we use this for \( a = x/2 \) we get the first claimed formula for \( K_N \), and if we use it for \( a = (N + 1)\frac{x}{2} \) then we get the second claimed formula. Compute the limit as \( x \to 0 \), this yields \( K_N(0) = N+1 \).

Property (d) is immediate from the first explicit formula. Estimate \(|1 - \cos(N + 1)x| \leq 2 \) and \((1 - \cos x) \geq 1 - \cos \delta \) for \( \delta \leq x \leq \pi \) and
also use that the cosine is an even function to get the same estimate for 
\(-\pi \leq x \leq -\delta\).

The nonnegativity of \(K_N\) is also clear from the explicit formulas.

The property (c) follows from \(\frac{1}{2\pi} \int_{-\pi}^{\pi} D_n(t)dt = 1\) (and taking the arithmetic mean of 1s gives a 1). \(\blacksquare\)

**Proof of Fejér’s theorem.** Given \(\varepsilon > 0\) we have to show that there is 
\(M = M(\varepsilon)\) so that for all \(N \geq M\),
\[
|\sigma_N f(x) - f(x)| \leq \varepsilon \quad \text{for all } x.
\]

Now we write
\[
\sigma_N f(x) - f(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} K_N(t)f(x-t)dt - f(x)
\]
\[
= \frac{1}{2\pi} \int_{-\pi}^{\pi} K_N(t)\left[f(x-t) - f(x)\right]dt;
\]
here we have used property (c).

\(f\) is continuous and therefore *uniformly continuous* on any compact interval. Since \(f\) is also \(2\pi\)-periodic, \(f\) is uniformly continuous on \(\mathbb{R}\). This means that there is a \(\delta > 0\) such that
\[
|f(x-t) - f(x)| \leq \frac{\varepsilon}{4} \quad \text{for } |t| \leq \delta, \text{ and all } x \in \mathbb{R}.
\]

We split the integral into two parts:
\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} K_N(t)\left[f(x-t) - f(x)\right]dt = I_N(x) + II_N(x)
\]
where
\[
I_N(x) = \frac{1}{2\pi} \int_{-\delta}^{\delta} K_N(t)\left[f(x-t) - f(x)\right]dt,
\]
\[
II_N(x) = \frac{1}{2\pi} \int_{[-\pi,\pi]\setminus[-\delta,\delta]} K_N(t)\left[f(x-t) - f(x)\right]dt.
\]

We give an estimate of \(I_N\) which holds for all \(N\). Namely
\[
|I_N(x)| \leq \frac{1}{2\pi} \int_{-\delta}^{\delta} |K_N(t)||f(x-t) - f(x)|dt
\]
\[
\leq \frac{1}{2\pi} \int_{-\delta}^{\delta} |K_N(t)|\frac{\varepsilon}{4}dt
\]
\[
\leq \frac{1}{2\pi} \int_{-\pi}^{\pi} |K_N(t)|\frac{\varepsilon}{4}dt = \frac{\varepsilon}{4},
\]
by (b) and (c). Since this estimate holds for all \(N\) we may now choose \(N\) large to estimate the second term \(II_N(x)\).
We use property (d) to estimate the integral for \( x \in [\delta, \pi] \cup [-\pi, -\delta] \). We crudely bound \(|f(x-t) - f(x)| \leq |f(x-t)| + |f(x)| \leq 2 \max |f| \). Thus

\[
|II_N(x)| \leq 2 \max |f| \frac{1}{2\pi} \int_{[-\pi, \pi] \setminus [-\delta, \delta]} \frac{1}{N+1} \left( \frac{2}{1 - \cos \delta} \right) dt
\]

As \( \frac{1}{N+1} \to 0 \) as \( N \to \infty \) we may choose \( N_0 \) so that for \( N \geq N_0 \) the quantity \( \frac{1}{N+1} \left( \frac{4 \max |f|}{1 - \cos \delta} \right) \) is less than \( \varepsilon/4 \). Thus for \( N \geq N_0 \) both quantities \(|I_N(x)|\) and \(|II_N(x)|\) are \( \leq \varepsilon/4 \) for all \( x \) and thus we conclude that

\[
\max_{x \in \mathbb{R}} |\sigma_N f(x) - f(x)| \leq \varepsilon/2 \text{ for } N \geq N_0.
\]

\( \square \)

**An application for the partial sum operator**

**Theorem.** Let \( f \) be a continuous \( 2\pi \)-periodic function. Then

\[
\lim_{n \to \infty} \left( \int_{-\pi}^{\pi} |S_n f(x) - f(x)|^2 dx \right)^{1/2} = 0
\]

i.e., \( S_n f \) converges to \( f \) in the \( L^2 \)-norm in the space of square-integrable functions. \(^3\)

**Proof.** By Theorem 8.11 in Rudin (which is linear algebra) we have \( S_N t_M = t_M \) for every trigonometric polynomial \( t_M(x) = \sum_{k=-M}^{M} \gamma_k e^{ikt} \) provided that \( N \geq M \).

Now let \( \varepsilon > 0 \). By Fejér’s theorem we can find such a trigonometric polynomial \( t_M \) (of some degree \( M \) depending on \( \varepsilon \)) so that \( \max |f(x) - t_M(x)| \leq \varepsilon \). Then for \( n > M \) we have \( S_n f - f = S_n(f - t_M) - (f - t_M) \).

We also have

\[
||S_N(f - t_M)||^2 \leq ||f - t_M||^2
\]

this is just (76) in 8.13 in Rudin. Thus

\[
||S_n f - f|| \leq ||S_n(f - t_M)|| + ||f - t_M|| \leq 2 ||f - t_M||.
\]

But we have

\[
\left( \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x) - t_M(x)|^2 dx \right)^{1/2} \leq \max |f - t_M| < \frac{\varepsilon}{2}
\]

and we are done. \( \square \)

\(^3\)Recall: This norm is given by \( ||f|| = (\frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)|^2 dx)^{1/2} \) and is derived from the scalar product \( \langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \overline{g(x)} dx \).
II. The Weierstrass approximation theorem

**Theorem.** Let \( f \) be a continuous function on an interval \([a, b]\). Then \( f \) can be uniformly approximated by polynomials on \([a, b]\).

In other words: Given \( \varepsilon > 0 \) there exists a polynomial \( P \) (depending on \( \varepsilon \)) so that
\[
\max_{x \in [a, b]} |f(x) - P(x)| \leq \varepsilon.
\]

Here \( f \) may be complex valued and then a polynomial is a function of the form \( \sum_{k=0}^{N} a_k x^k \) with complex coefficients \( a_k \) (considered for \( x \in [a, b] \)). If \( f \) is real-valued, the polynomial can be chosen real-valued.

A short proof relies on Fejér’s theorem and approximation of trigonometric functions by their Taylor polynomials.

**Proof.** We first consider the special case \([a, b] = [-\frac{\pi}{2}, \frac{\pi}{2}]\).

Extend the function \( f \) to a continuous function \( F \) on \([-\pi, \pi]\) so that \( F(x) = f(x) \) on \([-\frac{\pi}{2}, \frac{\pi}{2}]\) and \( F(-\pi) = F(\pi) = 0 \). Then we can extend \( F \) to a continuous \( 2\pi \) periodic function on \( \mathbb{R} \).

Let \( \varepsilon > 0 \). By Fejér’s theorem we can find a trigonometric polynomial \( T(x) = a_0 + \sum_{k=1}^{N} [a_k \cos kx + b_k \sin kx] \) so that
\[
\max_{x \in \mathbb{R}} |F(x) - T(x)| < \varepsilon/2.
\]

Now the Taylor series for \( \cos \) and \( \sin \) converge uniformly on every compact interval. Thus we can find a polynomial \( P \) so that
\[
\max_{x \in [-\pi, \pi]} |T(x) - P(x)| < \varepsilon/2.
\]

Combining the two estimates (and using that \( f = F \) on \([-\frac{\pi}{2}, \frac{\pi}{2}]\)) yields
\[
\max_{|x| \leq \pi/2} |f(x) - P(x)| = \max_{|x| \leq \pi/2} |F(x) - P(x)| < \varepsilon.
\]

**Arbitrary compact intervals.** Consider an interval \([a, b]\) and let \( g \in C([a, b]) \). We wish to approximate \( g \) by polynomials on \([a, b]\). Let \( \ell(t) = Ct + D \) so that \( \ell(-\pi/2) = a \) and \( \ell(\pi/2) = b \) (you can compute that \( C = \frac{b-a}{\pi}, D = \frac{b+a}{2} \)).

The inverse of \( \ell \) is given by \( \ell^{-1}(x) = \frac{x}{b-a}(x - \frac{b+a}{2}) \).

The function \( g \circ \ell \) is in \( C([-\frac{\pi}{2}, \frac{\pi}{2}] ) \). Thus by what we have already done, there exists a polynomial \( P \) such that
\[
\max_{t \in [-\frac{\pi}{2}, \frac{\pi}{2}]} |g(\ell(t)) - P(t)| < \varepsilon
\]
and therefore if we set \( Q(x) = P(\ell^{-1}(x)) = P(\frac{x}{b-a}(x - \frac{b+a}{2})) \) then \( Q \) is a polynomial and we have
\[
\max_{x \in [a, b]} |g(x) - Q(x)| < \varepsilon.
\]
III. Approximations of the identity

In this section we leave the subject of polynomial approximation and try to approximate continuous functions vanishing at $\pm \infty$ by smooth functions.

In a previous homework problem a $C^\infty$-function $\phi$ was constructed with the property that $\phi$ is positive on $(-1, 1)$ and $\phi(t) = 0$ for $|t| \geq 1$. If we divide by a suitable constant we may achieve and assume

$$\int_{-1}^{1} \phi(t) dt = 1$$

and we may also write $\int_{-\infty}^{\infty} \phi(t) dt = 1$ since $\phi$ vanishes off $[-1, 1]$.

Now for $s > 0$ define

$$\phi_s(t) = \frac{1}{s} \phi\left(\frac{t}{s}\right).$$

Then we also have $\int \phi_s(t) dt = 1$, by the substitution $u = t/s$. Graph the function $\phi_s$ for small values of the parameter $s$.

**Definition.** For continuous $f \in C(\mathbb{R})$ we define

$$A_s f(x) = \int_{-\infty}^{\infty} \phi_s(x-t) f(t) dt.$$ 

We shall be interested in the behavior of $A_s f$ for $s \to 0$. Note that the $t$-integral extends over a compact interval depending on $x, s$. The integral is also called a convolution of the functions $\phi_s$ and $f$.

**Exercise:** Let $f \in C(\mathbb{R})$. Show that for every $s > 0$ the function $x \mapsto A_s f$ is a $C^\infty$ function on $(-\infty, \infty)$. If $\lim_{|x| \to \infty} |f(x)| = 0$ then show also that $\lim_{|x| \to \infty} |A_s f(x)| = 0$.

**Theorem.** (a) Let $f \in C(\mathbb{R})$ and let $J$ be any compact interval. Then, as $s \to 0$, $A_s f$ converges to $f$ uniformly on $J$.

(b) Let $f$ be as in (a) and assume in addition that $\lim_{|x| \to \infty} |f(x)| = 0$. Then $A_s f$ converges to $f$ uniformly on $\mathbb{R}$.

**Proof.** We shall only prove part (b). As an exercise you can prove part (a) in the same way, or alternatively, deduce it from part (b).

One may change variables to write

$$A_s f(x) = \int_{-\infty}^{\infty} \phi_s(t) f(x-t) dt.$$ 

\[\text{The convolution of two functions defined on } \mathbb{R} \text{ is given by } f \ast g(x) = \int_{-\infty}^{\infty} f(y) g(x-y) dy \text{ whenever this makes sense; again one checks } f \ast g = g \ast f. \text{ We will not go into details here.}\]
Since \( \int \phi_s(t)dt = 1 \) we see that
\[
A_s f(x) - f(x) = \int_{-\infty}^{\infty} \phi_s(t)[f(x-t) - f(x)]dt.
\]
Note that, since \( \phi_s(t) = 0 \) for \( |t| > s \), the \( t \) integral is really an integral over \([-s, s]\).

The assumptions that \( f \) is continuous and that \( \lim_{|x|\to\infty} |f(x)| = 0 \) imply that \( f \) is uniformly continuous on \( \mathbb{R} \) (prove this!). Thus given \( \varepsilon > 0 \) there is a \( \delta > 0 \) so that \( |f(x-t) - f(x)| < \varepsilon/2 \) for all \( t \) with \( |t| \leq \delta \) and for all \( x \in \mathbb{R} \). If \( 0 < s < \delta \) we have by the nonnegativity of \( \phi_s \)
\[
|A_s f(x) - f(x)| \leq \int_{-s}^{s} \phi_s(t)|f(x-t) - f(x)|dt \leq \frac{\varepsilon}{2} \int_{-s}^{s} \phi_s(t)dt = \frac{\varepsilon}{2}
\]
for all \( x \in \mathbb{R} \).

\[\Box\]

**Terminology:** The linear transformations (aka as linear operators) \( A_s \) are called approximations of the identity. The identity operator \( \text{Id} \) is simply given by \( \text{Id}(f) = f \), and the above Theorem says that the operators \( A_s \) approximate in a certain sense the identity operator as \( s \to 0 \).

One can use other approximations of the identity defined like the one above where \( \phi \) is replaced by a not necessarily compactly supported function. If one drops the compact support the proofs get slightly more involved.

Other types of approximations of the identity (with a parameter \( n \to \infty \)) are given by the families of linear operators \( L_n \) in §IV below and \( B_n \) in §V below. For each \( f \) these linear operators will produce families of polynomials depending on \( f \).

### IV. The Landau polynomials:
#### A second proof of Weierstrass’ theorem

Let \( f \) be continuous on the interval \([-1/2, 1/2]\). Define
\[
Q_n(x) = c_n(1 - x^2)^n
\]
where \( c_n = (\int_{-1}^{1}(1 - s^2)^n ds)^{-1} \) so that \( \int_{-1}^{1} Q_n(t)dt = 1 \). The sequence of *Landau polynomials* associated to \( f \) is defined by
\[
L_n f(x) = \int_{-1/2}^{1/2} f(t)Q_n(t-x)dt.
\]
Verify that \( L_n f \) is a polynomial of degree at most \( 2n \).

By a change of variables one can use the following theorem to prove the Weierstrass approximation theorem on any compact interval \([a, b]\).

**Theorem.** Let \( \gamma > 0 \) and let \( I_\gamma = [-1/2 + \gamma, 1/2 - \gamma] \). The sequence \( L_n f \) converges to \( f \), uniformly on the interval \( I_\gamma \), i.e.
\[
\max_{x \in I_\gamma} |L_n f(x) - f(x)| \to 0, \text{ as } n \to \infty.
\]
**Proof.** We first need some information about the size of the polynomials $Q_n$. Consider $c_n^{-1} = \int_{-1}^{1}(1-s^2)^n ds$. We use the inequality

$$(1-x^2)^n \geq 1 - nx^2, \text{ for } 0 \leq x \leq 1.$$ 

To see this let $h(x) = (1-x^2)^n - 1 + nx^2$. The derivative of $h$ is $h'(x) = -2nx(n-1) + 2nx = 2nx(1-(1-x^2)^{n-1})$ which is positive for $x \in [0,1]$. Thus $h$ is increasing on $[0,1]$ and since $h(0) = 0$ we see that $h(x) \geq 0$ for $x \in [0,1]$. Since $h$ is even we have $h(x) \geq 0$ for $x \in [-1,1]$. We use the last displayed inequality in the integral defining the constant $c_n$ and get

$$c_n^{-1} = \int_{-1}^{1}(1-x^2)^n dx = 2 \int_{0}^{1}(1-x^2)^n dx \geq 2 \int_{0}^{n^{-1/2}}(1-x^2)^n dx \geq 2 \int_{0}^{n^{-1/2}}(1-nx^2)dx > n^{-1/2}$$ 

and from this we obtain

$$(*) \quad Q_n(x) \leq \sqrt{n}(1-x^2)^n.$$ 

Given $\varepsilon > 0$ the goal is to show that $\max_{x \in I, |L_n f(x) - f(x)| < \varepsilon}$ for sufficiently large $n$.

Let $\varepsilon > 0$. Since $f$ is uniformly continuous on $[-1/2,1/2]$ we can find $\delta > 0$ so that $\delta < \gamma$ and so that for all $x \in I$ and all $t$ with $|t| \leq \delta$ we have that $|f(x+t) - f(x)| < \varepsilon/4$.

Write (with a change of variables)

$$\int_{-1/2}^{1/2} f(s)Q_n(s-x)ds = \int_{-\frac{1}{2}+x}^{\frac{1}{2}+x} f(t+x)Q_n(t)dt$$

Since $x \in I = [-\frac{1}{2} + \gamma, 1/2 - \gamma]$ and since $\delta < \gamma$ we have $-1/2 + x < -\delta < \delta < 1/2 + x$. We may thus split the integral as

$$\int_{-1/2+x}^{-\delta} + \int_{-\delta}^{\delta} + \int_{\delta}^{1/2+x} f(t+x)Q_n(t)dt.$$ 

The idea is that the first and the third term will be small for large $n$. We modify the middle integral further to write

$$\int_{-\delta}^{\delta} f(t+x)Q_n(t)dt = \int_{-\delta}^{\delta} [f(t+x) - f(x)]Q_n(t)dt + f(x) \int_{-\delta}^{\delta} Q_n(t)dt$$

$5$The proof here is essentially the same as the proof of Weierstrass’ theorem in Theorem 7.26 of W. Rudin’s book.
and finally (using \( \int_{-1}^{1} Q_n(t) dt = 1 \))

\[ f(x) \int_{-\delta}^{\delta} Q_n(t) dt = f(x) - f(x) \int_{-1}^{-\delta} Q_N(t) dt - f(x) \int_{\delta}^{1} Q_N(t) dt. \]

Putting it all together we get

\[ L_n f(x) - f(x) = I_n(x) + II_n(x) + III_n(x) \]

where

\[ I_n(x) = \int_{-\delta}^{\delta} [f(t + x) - f(x)] Q_n(t) dt \]

\[ II_n(x) = \int_{-1/2}^{-1/2 + x} f(t + x) Q_n(t) dt + \int_{\delta}^{1/2 + x} f(t + x) Q_n(t) dt \]

\[ III_n(x) = -f(x) \int_{-\delta}^{\delta} Q_N(t) dt - f(x) \int_{\delta}^{1} Q_N(t) dt. \]

Estimate

\[ |I_n(x)| = \int_{-\delta}^{\delta} |f(t + x) - f(x)| Q_n(t) dt \leq \frac{\varepsilon}{4} \int_{-\delta}^{\delta} Q_N(t) dt \leq \frac{\varepsilon}{4} \int_{-1}^{1} Q_N(t) dt = \frac{\varepsilon}{4}; \]

this estimate is true for all \( n \).

Now let \( M = \max_{x \in [-1/2, 1/2]} |f(x)| \). Then by our estimate (*) for \( Q_n \) we see that

\[ |II_n(x)| + |III_n(x)| \leq 2M \max_{t \in [-1, -\delta] \cup [\delta, 1]} Q_n(t) \leq 2M \sqrt{n} (1 - \delta^2)^n \]

and since \( 2M \sqrt{n} (1 - \delta^2)^n \) tends to 0 as \( n \to \infty \) we see that there is \( N \) so that for \( n \geq N \) we have \( \max_{x \in I_n} |II_n(x) + III_n(x)| < \varepsilon/2 \) for \( n \geq N \). If we combine this with the estimate for \( I_n(x) \) we see that \( |L_n f(x) - f(x)| < \varepsilon \) for \( n > N \) and all \( x \in I_n \).

\[ \Box \]

V. The Bernstein polynomials:

A third proof of Weierstrass’ theorem

Here we consider the interval \([0, 1]\). For \( n = 1, 2, \ldots \) define

\[ B_n f(t) = \sum_{k=0}^{n} f\left( \frac{k}{n} \right) \binom{n}{k} t^k (1-t)^{n-k}, \]

the sequence of Bernstein polynomials associated to \( f \). Here \( \binom{n}{k} = \frac{n!}{k!(n-k)!} \), the binomial coefficients. For each \( n, B_n f \) is a polynomial of degree at most \( n \).
Theorem. If \( f \in C([0,1]) \) then the polynomials \( B_n f \) converge to \( f \) uniformly on \([0,1]\).

For the proof we will use the following auxiliary Lemma.

\[
\sum_{0 \leq k \leq n} \left( \frac{k}{n} - t \right)^2 \binom{n}{k} t^k (1-t)^{n-k} \leq \frac{1}{4n}.
\]

We shall first prove the Theorem based on the Lemma and then give a proof of the Lemma. There is also a probabilistic interpretation of the Lemma which is appended below.

Proof of the theorem. By the binomial theorem

\[
1 = (t + (1-t))^n = \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k}
\]

and thus we may write

\[
B_n f(t) - f(t) = \sum_{k=0}^{n} f\left( \frac{k}{n} \right) \binom{n}{k} t^k (1-t)^{n-k} - f(t) \cdot 1
\]

* = \[
= \sum_{k=0}^{n} \left[ f\left( \frac{k}{n} \right) - f(t) \right] \binom{n}{k} t^k (1-t)^{n-k}
\]

Given \( \varepsilon > 0 \) find \( \delta > 0 \) so that \( |f(t + h) - f(t)| \leq \varepsilon/4 \) if \( t, t + h \in [0,1] \) and \( |h| < \delta \). For the terms with \( |\frac{k}{n} - t| \leq \delta \) we will exploit the smallness of \( f\left( \frac{k}{n} \right) - f(t) \) and for the terms with \( |\frac{k}{n} - t| > \delta \) we will exploit the smallness of the term in the Lemma, for large \( n \). We thus split \( B_n f(t) - f(t) = I_n(t) + II_n(t) \) where

\[
I_n(t) = \sum_{0 \leq k \leq n} \left[ f\left( \frac{k}{n} \right) - f(t) \right] \binom{n}{k} t^k (1-t)^{n-k}
\]

\[
II_n(t) = \sum_{0 \leq k \leq n} \left[ f\left( \frac{k}{n} \right) - f(t) \right] \binom{n}{k} t^k (1-t)^{n-k}
\]

the decomposition depends of course on \( \delta \) but \( \delta \) does not depend on \( n \). We show that \( |I_n(t)| \leq \varepsilon/4 \) for all \( n = 2,3,\ldots \).

Indeed, since \( |f\left( \frac{k}{n} \right) - f(t)| \leq \varepsilon/4 \) for \( |\frac{k}{n} - t| \leq \delta \) we compute

\[
|I_n(t)| \leq \sum_{0 \leq k \leq n} |f\left( \frac{k}{n} \right) - f(t)| \binom{n}{k} t^k (1-t)^{n-k}
\]

\[
\leq \frac{\varepsilon}{4} \sum_{0 \leq k \leq n} \binom{n}{k} t^k (1-t)^{n-k} = \frac{\varepsilon}{4}
\]
where we have used again the binomial theorem.

Concerning $II_n$ we observe that $1 \leq \delta^{-2} \left( \frac{k}{n} - t \right)^2$ for $|\frac{k}{n} - t| \geq \delta$ and estimate $|f\left(\frac{k}{n}\right) - f(t)| \leq 2 \max |f|$. Thus

$$II_n(t) \leq \sum_{0 \leq k \leq n} \delta^{-2} \left( \frac{k}{n} - t \right)^2 |f\left(\frac{k}{n}\right) - f(t)| \binom{n}{k} t^k (1 - t)^{n-k}$$

$$\leq \delta^{-2} 2 \max |f| \sum_{0 \leq k \leq n} \left( \frac{k}{n} - t \right)^2 \binom{n}{k} t^k (1 - t)^{n-k}$$

By the Lemma $|II_n(t)| \leq (4n)^{-1} \delta^{-2} 2 \max |f|$ and for sufficiently large $n$ this is $\leq \varepsilon/2$ and we are done.

\[\square\]

**Proof of the Lemma.** We set $\psi_0(t) = 1$, $\psi_1(t) = t$ and $\psi_2(t) = t^2$, etc. Then we can explicitly compute the polynomials $B_n \psi_0, B_n \psi_1, B_n \psi_2$ for $n = 1, 2, \ldots$.

First, by the binomial theorem (as used before)

$$B_n \psi_0(t) = \sum_{k=0}^{n} \binom{n}{k} t^k (1 - t)^{n-k} = 1$$

thus $B_n \psi_0 = \psi_0$. Next for $n \geq 1$

$$B_n \psi_1(t) = \sum_{k=0}^{n} \frac{k}{n} \binom{n}{k} t^k (1 - t)^{n-k}$$

$$= \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)! (n-k)!} t^k (1 - t)^{n-k}$$

$$= t \sum_{k=1}^{n} \binom{n-1}{k-1} t^{k-1} (1 - t)^{n-1-(k-1)}$$

$$= t \sum_{j=0}^{n-1} \binom{n-1}{j} t^j (1 - t)^{n-1-j} = t$$

which means $B_n \psi_1 = \psi_1$ for $n \geq 1$. 

To compute \( B_n \psi_2 \) we observe that
\[
B_1 \psi_2(t) = \psi_2(0)(1 - t) + \psi_2(1)t = \psi_1(t)
\]
and, for \( n \geq 2 \)
\[
B_n \psi_2(t) = \sum_{k=0}^{n} \frac{k^2}{n^2} \binom{n}{k} t^k (1 - t)^{n-k}
\]
\[
= \frac{1}{n} \left( t^2(n-1) \sum_{k=2}^{n} \frac{n-2}{k-2} t^{k-2} (1 - t)^{n-2-(k-2)}
+ t \sum_{k=1}^{n} \frac{n-1}{k-1} t^{k-1} (1 - t)^{n-1-(k-1)} \right)
\]
\[
= \frac{(n-1)t^2 + t}{n} = t^2 + \frac{t - t^2}{n}.
\]
We summarize: For \( n \geq 2 \) we have
\[
B_n \psi_0 = \psi_0, \quad B_n \psi_1 = \psi_1, \quad B_n \psi_2 = \psi_2 + \frac{1}{n} (\psi_1 - \psi_2).
\]

To prove the assertion in the Lemma lets multiply out
\[
\left( \frac{k}{n} - t \right)^2 = \left( \frac{k}{n} \right)^2 - 2t \frac{k}{n} + t^2
\]
and use that the transformation \( f \mapsto B_n f(t) \) is linear (i.e we have \( B_n [c_1 f_1 + c_2 f_2](x) = c_1 B_n f_1(x) + c_2 B_n f_2(x) \) for functions \( f_1, f_2 \) and scalars \( c_1, c_2 \)). We compute, for \( n \geq 2 \)
\[
\sum_{0 \leq k \leq n} \left( \frac{k}{n} - t \right)^2 \binom{n}{k} t^k (1 - t)^{n-k}
\]
\[
= B_n \psi_2(t) - 2tB_n \psi_1(t) + t^2
\]
\[
= t^2 + \frac{t - t^2}{n} - 2t \cdot t + t^2 = \frac{t - t^2}{n}
\]
and since \( \max_{0 \leq t \leq 1} t - t^2 = 1/4 \) we get the assertion of the Lemma.

**Remark.** Let’s consider an arbitrary compact interval \([a, b]\) and let \( f \in C([a, b]) \). Then the polynomials
\[
P_n f(x) = \sum_{k=0}^{n} f\left(a + \frac{k}{n}(b - a)\right) \binom{n}{k} \frac{(x - a)^k (b - x)^{n-k}}{(b - a)^n}
\]
converge to \( f \) uniformly on \([a, b]\).

Using a change of variable derive this statement from the above theorem.
Addendum:

Probabilistic interpretation of the Bernstein polynomials. You might have seen the expressions $B_n f(t)$ in a course on probability. In what follows the parameter $t$ is a parameter for a probability (between 0 and 1).

Let’s consider a series of trials of an experiment. Each trial may is supposed to have two possible outcomes (either success or failure). Each integer in $X_n := \{1, \ldots, n\}$ represents a trial; we label the $j$th trial as $T_j$. Let $t \in [0, 1]$ be fixed. In each trial the probability of success is assumed to be $t$, and the probability of failure is then $(1 - t)$. The trials are supposed to be independent.

Let $A$ be a specific subset of $\{1, \ldots, n\}$ which is of cardinality $k$, i.e. $A$ is of the form $\{j_1, j_2, \ldots, j_k\}$ for mutually different integers $j_1, \ldots, j_k$; if $k = 0$ then $A = \emptyset$. Then the event $\Omega_A$ that for each $j \in A$ the trial $T_j$ results in a success and for each $j \in X_n \setminus A$ the trial $T_j$ results in a failure has probability $t^k (1 - t)^{n - k}$. There are exactly $\binom{n}{k}$ subsets $A$ of $X_n$ which have cardinality $k$ and they represent mutually exclusive (aka disjoint) events. Let $\Omega_k$ be the event that the $n$ trials result in $k$ successes, then the probability of $\Omega_k$ is

$$\mathbb{P}(\Omega_k) = \binom{n}{k} t^k (1 - t)^{n - k}.$$  

The probabilities of the mutually exclusive events $\Omega_k$ add up to 1;

$$\sum_{k=0}^{n} \mathbb{P}(\Omega_k) = 1;$$

(cf. the binomial theorem).

Let now $X$ be the number of successes in a series of $n$ trials ($X$ is a “random variable” which depends on the outcome of each trial). The event $\Omega_k$ is just the event that $X$ assumes the value $k$ (one writes $\mathbb{P}(\Omega_k)$ also as $\mathbb{P}(X = k)$). The random variable $X/n$ is the ratio of successes and total number of trials, and it takes values in $[0, 1]$ (more precisely in $\{0, \frac{1}{n}, \ldots, \frac{n}{n}\}$).

The expected value of $X/n$ is by definition

$$\mathbb{E}[X/n] = \sum_{k=0}^{n} k \cdot \mathbb{P}(\Omega_k)$$

and in the proof of the Lemma we computed it to

$$\mathbb{E}[X/n] = \sum_{k=0}^{n} k \binom{n}{k} t^k (1 - t)^{n - k} = B_n \psi_1(t) = t.$$  

Generally, if $f$ is a function of $t$, the expected value of $f(X/n)$ is equal to

$$\mathbb{E}[f(X/n)] = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \mathbb{P}(\Omega_k) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} t^k (1 - t)^{n - k};$$


that gives the probabilistic interpretation of the Bernstein polynomials evaluated at $t$;

$$B_n f(t) = \mathbb{E}[f(X/n)].$$

The Variance of $X/n$ is given by

$$\mathbb{E}[(X/n - \mathbb{E}[X/n])^2] = \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k} = \frac{t - t^2}{n},$$

as computed in the proof of the lemma.

Let $\delta > 0$ be a small number. The probability that the number of successes deviates from the expected value $tn$ by more than $\delta n$ is given by

$$\sum_{0 \leq k \leq n} \mathbb{P}(\Omega_k) = \sum_{0 \leq k \leq n} \binom{n}{k} t^k (1-t)^{n-k}.$$

The smallness of this quantity (uniformly in $t$) played an important role in the Bernstein proof of Weierstrass’ theorem. It was estimated by

$$\mathbb{E} \left[ \frac{(X - \mathbb{E}[X])^2}{\delta n} \right] = \delta^{-2} \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k} = \frac{t - t^2}{\delta^2 n}.$$

Thus, by the statement of the Lemma, the event that the number of successes deviates from the expected value $tn$ by more than $\delta n$ has probability no more than $(4\delta^2 n)^{-1}$. 