2016

Best Approximations, Lethargy Theorems and Smoothness

Caleb Case
Claremont McKenna College

Claremont McKenna College

Best Approximations, Lethargy Theorems and Smoothness

Submitted to
Professor Asuman Guven Aksoy
and
Dean Peter Uvin

by
Caleb Case

for
Senior Thesis
Fall 2015
11/30/15

Contents

Abstract. In this paper we consider sequences of best approximation. We first examine the ρ best approximation function and its applications, through an example in approximation theory and two new examples in calculating n-widths. We then further discuss approximation theory by examining a modern proof of Weierstrass's Theorem using Dirac sequences, and providing a new proof of Chebyshev's Equioscillation Theorem, inspired by the de La Vallee Poussin Theorem. Finally, we examine the limits of approximation theorem by looking at Bernstein Lethargy theorem, and a modern generalization to infinite-dimensional subspaces. We all note that smooth functions are bounded by Jackson's Inequalities, but see a newer proof that a single non-differentiable point can make functions again susceptible to lethargic rates of convergence.

1 Introduction

1.1 Context

Approximation theory has a formal beginning in Weierstrass's 1885 paper *On the possibility of giving an analytic representation to an arbitrary function of a real variable.* As the title suggests, Weierstrass was initially interested in complex functions, and in representing them with power series. However, this pursuit led him to discover and prove the Weierstrass Theorems, showing that both algebraic polynomials and trigonometric polynomials are dense in the set of continuous functions.

These conclusions were what led his paper to immediate and lasting recognition. His paper was reprinted in French within a year of original publication, and his theorems have been called "of great importance in the development of the whole of mathematical analysis"[12] and "one theorem in approximation theory as being of greater significance than any other"[5]. Weierstrass's theorem remains the foundation of Approximation Theory, and has inspired numerous alternate proofs, by other well-known names such as Lebesque, Bernstein, De la Vallee Poussin, and Landau.[10]

A counterpoint to his Approximation Theory, Weierstrass was also interested in continuous, nowhere differentiable functions. While not the first to construct such functions, Weierstrass did serve to bring previous discoveries to light through his teaching and published works. In 1872, Weierstrass proved that the function

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

is continuous but nowhere differentiable when $b \in (0, 1)$, a is an odd integer and $ab > 1 + (3\pi/2)$. Weierstrass also claimed that Riemann had proved an earlier example, that

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

is also continous but nowhere differentiable. In light of this research, it is no surprise that Weierstrass was later so enthralled in the discovery of entire approximating functions. After showing that some continuous functions are nowhere differentiable, it is certainly a compelling discovery that they can be approximated perfectly by a series of the most analytic and well behaved functions, polynomials.

After it had been established that the set of polynomials is dense in continuous functions, further questions arose. The main concern became the question of approximating functions with polynomials of a limited degree. This is driven by practical concerns, as an infinite series is often no easier to compute than the original function. Instead, if we limit ourselves to polynoials of degree at most n, what can be said of the best approximation? There is no unified answer to this question. In fact, Bernstein proved that there exist functions whose best approximation converges arbitrarily slowly as the degree of the polynomial rises.
The next step then, is determining best approximations for functions with specific properties. Jackson found a bound on the error of approximation for all smooth functions, while more recently Almira[1] has proved that for functions not differentiable at a single point no such bound can be created. We also see limits on specific functions, methods of determining when a polynomial is the best approximation, and methods for creating approximations on a finite number of points instead of an interval.

2 Sequences of Best Approximation

2.1 Measuring Distances

Before we can compare approximations, we need a metric to determine how close an approximation is to the desired target. In order to maintain generality, we can define a distance-measure for all metric spaces. We will call this the ρ-function and define it as follows:

Definition 2.1. Let M be a metric space with measure d, and let $S \subset M$ be a subset of that space. Then, for any point $x \in M$, we can define the distance from x to S as

$$dist(x, S) = \rho(x, S) = \inf_{y \in S} d(x, y)$$

There are a few properties we can derive from ρ to show that it functions well as a distance metric.

(1.) $\rho(x_1 + x_2, S) \leq \rho(x_1, S) + \rho(x_2, S) \forall x_1, x_2 \in M$

This is equivalent to saying that ρ maintains the Triangle Inequality.

We can see that this is true by first letting y_1 be the point at the limit of the distances to x_1, such that $d(y_1, x_1) = \inf_{y \in S} d(x, y)$. Note that y_1 is not necessarily in S. Define y_2 similarly, and from the Triangle Inequality we can see that $d(x_1 + x_2, y_1) \leq d(x_1, y_1) + d(x_2, y_1)$, and that $d(x_1 + x_2, y_2) \leq d(x_1, y_2) + d(x_2, y_2)$. Because S is a subspace, we see that $y_1 + y_2 \in \bar{S}$ as well. Thus,

$$\begin{aligned}
\rho(x_1 + x_2, S) &\leq d(x_1 + x_2, y_2) + d(y_2, y_1 + y_2) \\
&= d(x_1, y_2) + d(x_2, y_2) + d(y_2, y1 + y2) \\
&\leq d(x_1, y_1) + d(x_2, y_2) + d(y_1, y_1 + y_2) - d(y_2, y_1 + y_2) \\
&= \rho(x_1, S) + \rho(x_2, S)
\end{aligned}$$

(2.) $\rho(x + y, S) = \rho(x, S) \forall y \in S$.

This follows immediately from the previous property, and from noting that $\rho(y, S) = 0$ when $y \in S$. Thus, $\rho(x + y, S) \leq \rho(x, Y) + 0$, and we know the distance cannot be strictly less than, or by adding an infinite series of other $y_i \in S$ points we could reduce the distance arbitrarily low.

(3.) $|\rho(x_1, S) - \rho(x_2, S)| \leq |x_1 - x_2|$.

First we note that $\rho(x_2, S) \leq \rho(x_1, S) + |x_1 - x_2|$. The same applies to $\rho(x_2, S)$, and so we see that $|\rho(x_1, S) - \rho(x_2, S)| \leq \rho(x_1, S) + |x_1 - x_2| + \rho(x_2, S) + |x_1 - x_2$. Some rearrangement gives us $2||\rho(x_1, S) - \rho(x_2, S)| \leq 2|x_1 - x_2|$, and thus proves the property.

(4.) If S is compact, then $\rho(x, S)$ is realized in S.

Namely, there exists some $s \in S$ such that $d(x, s) = \rho(x, S)$. This is due to the nature of compact sets, where every limit point is contained within the set. We see that because ρ is the infemum of some series of points in S, the ρ is realized at the limit point of this series, and this point is thus contained in S.

The third property of ρ gives us a continuous map $M \to \mathbb{R}^+$ defined as $x \to \rho(x, S)$. This is continuous because for any $\varepsilon > 0$ and $x_0 \in M$, we see that the neighborhood $|x - x_0| < \varepsilon$ gurantees that $|\rho(x_0, S) - \rho(x, S)| < \varepsilon$.

2.2 Applications for ρ

2.2.1 Approximation Theorem

ρ gives us an extremely useful measure for the amount a polynomial approximates a function. If we let the space M be the space of all continuous functions on $[0, 1]$, and we let P be the space of polynomials, then for any function x we can see that $\rho(x, P)$ gives us the difference betwee x and the best approximating polynomial. Given that polynomials are dense in this space, this gives us an intuitive proof that $\rho(x, P) = 0$, meaning that all functions can be approximated by a polynomial.
If we want to restrict our subset to specific polynomials, say those of degree n or less, we find that the minimum degree of approximation is still equal to $\rho(x, P')$. As long as we choose a compact subset, we even know that there is a polynomial $p \in P'$ such that $d(p, x) = \rho(x, P')$. Thus, there exists a polynomial that is the closest possible approximation of f given our constraints.

So, to prove the existence of such a polynomial, we must show that the set of polynomials of at most degree n, P_n, is closed in $C[a, b]$. We define the distance between functions as $d(f, g) = \sup(|f(x) - g(x)|)$ for all $x \in \mathbb{C}$. We can note that the space of continuous functions is also normed, using the norm $||.||$ such that $||f(z)|| = \sup |f(x)|$. We also note that the set of n-degree polynomials can be represented as all vectors $v_i = \{c_0, c_1 \cdots c_n\}$, giving the space dimension n. This implies that the space is a finite-dimensional subspace of a normed space, which we know is closed and so can conclude that P_n is closed.

This implies that P_n contains all of its limit points. Thus, if we can define a convergent series of n-degree polynomials, we know that the convergence of this series will also be an n-degree polynomial. This gives us the existence of a best approximating polynomial in P_n.

We can also note that the best approximation is unique. Consider when $\rho(f, P_n) = c > 0$. Let there be two polynomials of degree less than or equal to n, p_1, p_2 such that $c = ||f - p_1|| = ||f - p_2||$. Consider the construct $\frac{p_1 + p_2}{2}$, which is also a polynomial of degree n. For any point x, note that $|p_1(x) - f(x)| \leq c$, and $|p_2(x) - f(x)| \leq c$. Thus, $|f(x) - \frac{p_1 + p_2}{2}| \leq c$. Now, however, consider the points where $|p_1(x) - f(x)| = c$, of which at least one must exist. If $p_1 \neq p_2$, then at least one of these points, $p_1(x) \neq p_2$, and thus $|f(x) - |p_1(x) - f(x)| < c$, meaning a better approximating polynomial exists. However, we already assumed p_1 and p_2 were best approximations. Thus, $p_1 = p_2$, and we see that the best approximation of f of n degree is unique. Lorentz [7] gives an example of an approximating polynomial. First, we restate Chebyshev's equioscillation theorem. A proof of the forward direction of this theorem can be found later, on page 24.

Lemma 2.2. *Let f be a continuous function on $[a, b]$. Among all polynomials of degree $\leq n$, the polynomial g minimizes the degree of approximation if and only if there are $n + 2$ distinct points in $[a, b]$ such that $f(x_i) - g(x_i) = \sigma(-1)^i ||f - g||$ where $\sigma = \pm 1$.*

Theorem 2.3. *Let $f(x) = (x - a)^{-1}, x \in [-1, 1], a > 1$. Then for $c = a - \sqrt{a^2 - 1} < 1$,*

$$E_n((x - a)^{-1}) = \frac{4c^{n+2}}{(1 - c^2)^2}$$

Proof. Consider the 1-1 map $x = \frac{1}{2}(w + w^{-1})$. Any $x \in \mathbb{C}$ can be written in this form for some w, and more specificaly the disk $|w| = 1$ is mapped to $[-1, 1]$. This map gives us easy access to oscillating functions when we represent w as $e^{i\theta}$. We see that this makes w periodic, and more importantly if we take $w^n = e^{in\theta}$, we see that this is equal to 1 or -1 exactly $n + 2$ times as θ goes from 0 to π. Thus, if we could find some function $\Psi(w)$ such that $\arg \Psi(w)$ goes from 0 to $(n + 2)k\imath$ on $|w| = 1$, and use it to construct a polynomial, we could meet Chebyshev's requirements.
In fact, such a function exists, which we will call $\Psi(w) = w^n(c - w)(1 - cw)^{-1}$. We note that both Ψ and Ψ^{-1} oscillate together $n + 2$ times, then we see that a function of the form

$$\Phi(x) = \frac{M}{2}(\Psi(w) + \Psi(w)^{-1})$$

achieves a value of M or $-M$ exactly $n + 2$ times. We will now show first, that a polynomial in x can be extracted from this equation, and be of the form

$$\Phi(x) = \frac{A}{x - a} - P_n(x)$$

By taking the limit as $x \to a$, we can then isolate A as $\Phi(x)(x - a)$. By choosing an appropriate value of M, namely $\frac{4c^{n+2}}{(1 - c^2)^2}$, we can ensure that

$A = 1$, and thus Φ is of the form $\dfrac{1}{x-a} - P_n(x)$, and that this function is equal to $\pm M$ exactly $n+2$ times, thus satisfying Chebyshev's theorem and proving that M is the degree of approximation of functions of this form.

We first note that $c = c = a - \sqrt{a^2 - 1} < 1$ is equivalent to $a = \dfrac{1}{2}(c + c^{-1})$. Let us prove an important property of this map, that $(1+c^2)(1-\dfrac{x}{a}) = (\dfrac{1}{w} - c)(w-c)$.
We can prove this property by first expanding the left side.

$$\begin{aligned}
(1+c^2)(1-\frac{x}{a}) &= (1+c^2)(1-\frac{w+w^{-1}}{c+c^{-1}}) \\
&= 1 + c^2 - \frac{w+w^{-1}}{c+c^{-1}}(1+c^2) \\
&= 1 + c^2 - \frac{(1+c^2)w}{c+c^{-1}} - \frac{(1+c^2)w^{-1}}{c+c^{-1}} \\
&= 1 + c^2 - cw - \frac{c}{w} \\
&= (\frac{1}{w} - c)(w-c)
\end{aligned}$$

We reach the second to last step by noting that $\dfrac{(1+c^2)}{c+c^{-1}} = c$.

We now define a function that will oscillate on the unit circle,

$$\Phi(x) = \frac{M}{2}\left(w^n \frac{c-w}{1-cw} + w^{-n}\frac{1-cw}{c-w}\right)$$

on $\mathbb{C}$. Note that $w^k + w^{-k}$ for $k \in \mathbb{Z}$ is a polynomial of x of degree k, because of the definition of our map φ. Note also that $(1+c^2)(1-\dfrac{x}{a}) = (\dfrac{1}{w} - c)(w-c)$.

Therefore we can expand Φ to

$$\begin{aligned}\Phi(x) &= \frac{M}{2}\left(w^n\frac{c-w}{1-cw}+w^{-n}\frac{1-cw}{c-w}\right)\\ &= \frac{M}{2}\left((-1)w^{n-1}\frac{w^2-cw}{1-cw}+(-1)w^{1-n}\frac{\frac{1}{w}-c}{w-c}\right)\\ &= -\frac{M}{2}\left(w^{n-1}\frac{w-c}{w^{-1}-c}+w^{1-n}\frac{w^{-1}-c}{w-c}\right)\end{aligned}$$

Recalling that $w^{-1}-c = \dfrac{(1+c^2)(1-\frac{x}{a})}{w-c}$, we get

$$\begin{aligned}\Phi &= -\frac{M}{2}\left(w^{n-1}\frac{(w-c)^2}{(1+c^2)(1-\frac{x}{a})}+w^{1-n}\frac{(w^{-1}-c)^2}{(1+c^2)(1-\frac{x}{a})}\right)\\ &= -\frac{M}{2}\left(w^{n-1}(w-c)^2+w^{1-n}(w^{-1}-c)^2\right)\frac{1}{(1-\frac{x}{a})(1+c^2)}\end{aligned}$$

We can now extract a polynomial from this function. First we expand to

$$-\frac{M}{2(1-\frac{x}{a})(1+c^2)}(w^{n-1}(w-c)^2)+-\frac{M}{2(1-\frac{x}{a})(1+c^2)}w^{1-n}(w^{-1}-c)^2$$

Then convert the denominator to get

$$-\frac{M}{2}\frac{w^{n-1}(w-c)^2}{(w^{-1}-c)(w-c)}-\frac{M}{2}\frac{w^{1-n}(w^{-1}-c)^2}{(w^{-1}-c)(w-c)}$$

$$=-\frac{M}{2}\frac{w^{n-1}(w-c)}{w^{-1}-c}-\frac{M}{2}\frac{w^{1-n}(w^{-1}-c)}{w-c}$$

Examining just the first term, multiply by $\frac{1+c^2}{1+c^2}$ and expand the top to get

$$\begin{aligned}\Phi(x) &= \frac{-M}{2(1+c^2)} \frac{w^n - cw^{n-1} + c^2 w^n - c^3 w^{n-1}}{w^{-1} - c} \\ &= \frac{-M}{2(1+c^2)} \frac{w^n - cw^n(w^{-1} - c) - c^3 w^{n-1}}{w^{-1} - c} \\ &= -w^n \frac{cM}{2(1+c^2)} + \frac{M}{2(1+c^2)} \frac{w^n - c^3 w^{n-1}}{\frac{1}{w} - c}\end{aligned}$$

Using a parallel process with the second term, we get $\frac{-M}{2(1+c^2)} \frac{w^{-n} - cw^{-n}(w-c) - c^3 w^{1-n}}{w-c} = -w^{-n} \frac{cM}{2(1+c^2)} + \frac{M}{2(1+c^2)} \frac{w^{-n} - c^3 w^{1-n}}{w-c}$. This allows us to rewrite the equation as

$$\frac{M}{2(1+c^2)} \frac{w^n - c^3 w^{n-1}}{\frac{1}{w} - c} + \frac{M}{2(1+c^2)} \frac{w^{-n} - c^3 w^{1-n}}{w-c} - \frac{cM}{2(1+c^2)}(w^n + w^{-n})$$

$$\frac{M}{2(1+c^2)} w^{n-2} \frac{w^2 - wc^3}{\frac{1}{w} - c} + \frac{M}{2(1+c^2)} w^{2-n} \frac{w^{-2} - c^3 w^{-1}}{w-c} - \frac{cM}{2(1+c^2)}(w^n + w^{-n})$$

We notice here that the left terms are now very similar to the inital state of the function, except that the degree of w has grown 1 step closer to 0. Thus, we can repeat this process of isolating the highest degree, giving us a series of the form $q_k(w^k + w^{-k})$ where q_k is some constant, and leaving us with a leftover term composed of w and c. We notice that in the leftover term, the denominator always retains the term $1 - \frac{x}{a}$, and the rest of the term never includes this. Thus, by taking what is left after all the polynomials have been filtered out, and then multiplying by $\frac{-a}{-a}$, we get some A such that the leftover term is $\frac{A}{x-a}$. Note that the function is equal to this leftover plus the polynomial of degree n, call it $P_n(x)$, and thus we can write Φ as

$$\varphi(x) = \frac{A}{x-a} - P_n(x)$$

Now, calculating A by hand would be an intensive process, and potentially different for every n. However, if we recall that Φ is defined on all of $\mathbb{C}$, then we can simplify the process by consider the limit when $x \to a$. We see that the polynomial term simple becomes $P_n(a)$, some constant, finite value. This becomes insignificant when compared to $\lim_{x\to a} \frac{A}{x-a}$, and so we see that $\lim_{x\to a} \Phi(x) = \lim_{x\to a} \frac{A}{x-a}$, and thus $A = \lim_{x\to a}(x-a)(\Phi(x))$. We can use this equality to solve for A, and see that

$$\begin{aligned}
A &= \lim_{x\to a} 1/2(w + w^{-1} - c - c^{-1})\frac{M}{2}(w^n \frac{c-w}{1-cw} + w^{-n}\frac{1-cw}{c-w}) \\
&= \frac{M}{4}\lim_{w\to c}(w + w^{-1} - c - c^{-1})(w^n \frac{c-w}{1-cw}) + w^{-n}\frac{1-cw}{c-w}) \\
&= \frac{M}{4}\lim_{w\to c}(w + w^{-1} - c - c^{-1})(w^{-n}\frac{1-cw}{c-w}) \\
&= \frac{M}{4}\lim_{w\to c}((w^{1-n}\frac{1-cw}{c-w}) + (w^{-n-1}\frac{1-cw}{c-w}) - (w^{-n}\frac{c-c^2w}{c-w}) - ((w^{-n}\frac{c^{-1}-w}{c-w})) \\
&= \frac{M}{4}\lim_{w\to c} w^{-n-1}\frac{1-cw}{c-w} - w^{-n}\frac{c^{-1}-w}{c-w} \\
&= \frac{M}{4}\lim_{w\to c} w^{-n}\frac{w^{-1}-c}{c-w} - \frac{c^{-1}-w}{c-w} \\
&= \frac{M}{4}\lim_{w\to c} w^{-n}\frac{w^{-1}-c-c^{-1}+w}{c-w} \\
&= \frac{M}{4}\lim_{w\to c} w^{-n}\frac{c-c^2w-w+w^2c}{cw(c-w)} \\
&= \frac{M}{4}\lim_{w\to c} w^{-n}\frac{(w-c)(cw-1)(1-cw)}{cw(c-w)} \\
&= \frac{M(1-c^2)^2}{4c^{n+2}}
\end{aligned}$$

Thus, if we want A to be 1, we choose $M = \frac{4c^{n+2}}{(1-c^2)^2}$, and get $\Phi(x) = \frac{1}{x-a} - P_n(x)$.

We can now examine how this function maps between $[-1, 1]$ and $|w| = 1$. We see that as w moves on the circle from angle 0 to π counterclockwise, x

moves from -1 to 1. For simplicity, we define $\Psi(w) = w^n(c-w)(1-cw)^{-1}$. We can now represent $\varphi(x) = \frac{M}{2}(\Psi(w) + \Psi(w)^{-1}$. We now note that if $|w| = 1$, then $|\Psi(w)| = 1$. Thus, $|\Phi(x)| \leq \frac{M}{2}(1+1) = M$. We see that $varPsi(w)$ has $n+1$ roots, and note that $\Psi(1) = 1$ and $\Psi(-1) = -1$, thus $\Phi(1) = M$ and $\Phi(-1) = -M$. Because Ψ is periodic, if we represent $w = e^{i\theta}$ then at least n times $\theta = 0$ and $\Phi(x) = \pm M$. Thus $\Phi(x) = M$ or $-M$ exactly $n+2$ times, and by Chebyshev's Equioscillation Theorem, P_n is the polynomial of best approximation, and the error of approximation is M. □

We see from this example that $\rho(f, P_n) = \inf |f(x) - P_i| = M$, satisfying that the given polynomial really is the best approximation. We also see that as $n \to \infty$, $M \to 0$ and thus the sequence of polynomials we get as we increase n converges to f. We can also begin to see the form of the actual best approximation, by noting that $P_n(x) = \frac{1}{x-a} + \Phi(x)$.

2.2.2 Jackson Inequality

Given what we have shown regarding continuous functions, a natural next step is to ask if the smoothness of the function also impacts the error of approximation. In fact, the derivative of a function, if it exists, can be used to create an upper bound on the degree of error. More specifically, the approximation can be bounded of some function of the norm of the derivative. This is intuitively consistent, as we imagine that a function that had no large changes (points of high magnitude of derivative) would be easier to approximate than one with large changes. Also, as the norm of the derivative approaches 0, the function approaches a linear one and we would expect the error of approximation to thus approach 0 as well. Jackson provided the first proofs of this, for trigonometric polynomials.

We begin by narrowing our metric space. Let us consider the space of continuous 2π-periodic functions, using the standard supremum norm. Let us also consider the subspace of trigonometric polynomials, which are of the form $p(\theta) = \sum_k^n (a_k \cos(k\theta) + b_k \sin(k\theta)$. These are clearly 2π periodic as well, and form a subspace of all such continuous functions. For any function f, we find that $\rho(f, P_n)$ exists, and in fact because P_n is closed in this space as well, we find that there always exists an optimal trigonometric polynomial of degree

n.
We can even put a maximum and minimum bound on $\rho(f, P_n)$, using Jackson's first theorem. Jackson's proof of this can be found in Cheney's book [5]. We will be dealing with integration of trigonometric polynomials, but we do not want integrals in our final bound, so we first need a way to force the integral of a sine function to be 0. We find this is possible by multiplying it with the sign of the max term of the polynomial.

Lemma 2.4. *If $k < n$, then $\int_0^\pi \sin(kx) \operatorname{sgn} \sin nx \; dx = 0$*

Proof. We can first simplify what we must demonstrate. Because the integrand is simply *sin* times a scalar value, it is even, and so we can also prove that $\int_{-\pi}^{\pi} \sin(kx) \operatorname{sgn} \sin nx dx = 0$. Recall that $\sin kx = \dfrac{e^{ikx} - e^{-ikx}}{2i}$. Since both these values of x will always have the same sign, we know that both must integrate to 0, and so we must only show that for $|m| < n$

$$\int_{-\pi}^{\pi} e^{imx} \operatorname{sgn} \sin nx \; dx = 0$$

This integrand is periodic, and we will show that by shifting the limits of integration, we can extract a non-1 constant and end with an integral of the same form. The integral times a constant is itself, and so must be 0. We now change the variable $x = y + \dfrac{pi}{n}$, and call the integral in this form I, such that

$$I = \int_{-\pi - \frac{\pi}{n}}^{\pi - \frac{\pi}{n}} e^{im(y + \frac{\pi}{n})} \operatorname{sgn} \sin(ny + \pi) dy$$

We can factor out the constant $e^{im\pi/n}$ and noting that the integrand is still 2π periodic, and the length of integration is exatly 2π, we shift the limits of integration by $\dfrac{pi}{n}$ to get

$$I = e^{im\pi/n} \int_{-\pi}^{\pi} e^{imy} \operatorname{sgn} \sin(ny + \pi) dy$$

We also note that $\sin ny + \pi = -\sin(ny)$, and thus we can also factor out a -1 from the integrand. We note that this leaves us with an integrand identical to the first, times a constant. Namely,

$$I = -e^{im\pi/n} \int_{-\pi}^{\pi} e^{imy} \operatorname{sgn} \sin(ny) dy = -e^{im\pi/n} I$$

If $I = -e^{im\pi/n}I$, then either $-e^{im\pi/n} = 1$ or $I = 0$. We see that $|m| < n$ implies that $|m\pi/n| < \pi$, and thus cannot be a multiple of π. This means that $-e^{im\pi/n} \neq 1$, and thus we conclude that $I = 0$. □

We now have a way of removing the integral from a function, and we will use this to establish a lower bound on the integral of the differences between a function and a trigonometric polynomial. Essentially, by multiplying by the sign of $\sin nx$ we get a strictly smaller value, and can then simplify this value so that it does not need the integral.

Lemma 2.5. *The minimum value for $\int_0^\pi |x - \sum_{k=1}^n a_k \sin kx| dx$ is $\pi^2/2n$ for all choices of a_k.*

Proof. First, we note that if we multiply the inside of the integrand by $\operatorname{sgn} \sin nx$, the absolute value of the whole integral will decrease. If we take the absolute value of the integrand, it is of course only positive. But if we introduce $\operatorname{sgn} \sin nx$, we see that it is sometimes negative, and so the abolute value of the whole integral will be strictly less. Rigorously, this means

$$\int_0^\pi |x - \sum_{k=1}^n a_k \sin kx| dx \geq \left| \int_0^\pi (x - \sum_{k=1}^n a_k \sin kx) \operatorname{sgn} nx \; dx \right|$$

Distributing and noting from 2.4 that all terms except the one with x will be 0, we get

$$= \left| \int_0^\pi x \operatorname{sgn} nx \; dx \right|$$

We can then break this up into the smallest sections where $\operatorname{sgn} \sin nx$ is constant, giving us

$$= \left| \sum_{k=0}^{n-1} (-1)^k \int_{k\pi/n}^{(k+1)\pi/n} x \; dx \right|$$

We can now evaluate the integral to

$$= \left| \sum_{k=0}^{n-1} (-1)^k \frac{1}{2} \left[\left(\frac{k+1}{n} \pi \right)^2 - \left(\frac{k}{n} \pi \right)^2 \right] \right|$$

$$= \frac{\pi^2}{2n^2} | \sum_{k=0}^{n-1} (-1)^k (2k+1) |$$

We see that the sum is $1 - 2 - 1 + 4 + 1 - 6 - 1 \cdots + (-1)^{n-1}(2(n-1)+1)$. We see that the 1's cancel out if n is even, or evaluate to 1 if n is odd. The other terms evaluate to $\pm\lceil (n/2) \rceil$. Thus, $|\sum_{k=0}^{n-1}(-1)^k(2k+1)| = n$, and we arrive at

$$\int_0^\pi |x - \sum_{k=1}^n a_k \sin kx| dx \geq \frac{\pi^2}{2n}$$

We now must show that we can actually achieve this lower bound. Recall where we introduced the inequality originally. If we can choose a_k such that $x - \sum a_k \sin kx = 0$ exactly when $\sin nx = 0$, then the inequality will simply become an equality. Let us define the function $\phi(x) = x - \sum a_k \sin kx$. Namely, for all $i \in \{1 \cdots n-1\}$,

$$\sum_{k=0}^{n-1} a_k \sin kx_i = x_i$$

We can prove that $\phi(x)$ changes sign at each x_i by first noting that $\{\sin(x), \cdots, \sin(n-1)x\}$ is a set of linearly independent vectors. Thus, the system of equations represented above has a solution. □

We now require a way to represent an approximating function, so that we can evaluate the difference between this approximation and f. We define

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

where A_k is up to our choice, but a_k, b_k are the Fourier coefficients of f. We show in the below lemma that the difference between Lf and f can be represented as a combination of $||f'||$ and the integral of a sum. In fact, it's the integral of the same sum we addressed in 2.5, and so we will have a method of minimizing the difference.

Lemma 2.6. *If f is 2π periodic and if f' is continuous, then*

$$(Lf - f)(x) = \frac{1}{\pi}\int_{-\pi}^{\pi} \left[\frac{1}{2}t + \sum_{k=1}^n \frac{(-1)^k}{k} A_k \sin kt\right] f'(x + \pi - t)dt$$

.

Proof. We work backwards, to show that the given integral is equal to the difference in equations. For simplicity, let us define $\phi(t) = \frac{1}{2}t + \sum_{k=1}^{n} \frac{(-1)^k}{k} A_k \sin kt$. We can then integrate the equation by parts, letting $\phi(t) = u$ and $f(x+\pi-t) = v$ to give us

$$\frac{-1}{\pi}\phi(t)f(x+\pi-t)\Big|_{-\pi}^{\pi} + \frac{1}{\pi}\int_{\pi}^{\pi}\phi'(t)f(x+\pi-t)dt$$

We can evaluate these functions, noting that $\phi(\pi) = \frac{\pi}{2} + \sum_{k=1}^{n} \frac{(-1)^k}{k} A_k \sin k\pi = \frac{\pi}{2}$, and symmetrically $\phi(-\pi) = \frac{-\pi}{2}$. Recall also that, because f is periodic, $f(x) = f(x+2\pi)$. Thus, we can evaluate to

$$-f(x) + \frac{1}{\pi}\int_{-\pi}^{\pi}[\frac{1}{2} + \sum_{k=1}^{n}(-1)^k A_k \cos kt]f(x+\pi-t)dt$$

We now put t in terms of a new parametric variable s so that we can simplify the argument of the function. Let $t = x+\pi-s$. We immediately see that $f(x+\pi-t) = f(s)$. We also note that, by using the difference in angles identity of cosine, we get that $\cos(kt) = \cos k(x+\pi-s) = \cos k(x+\pi) - ks = \cos k(x+\pi)\cos ks + \sin k(x+\pi) + \sin ks$. Noting that $k\pi$ will shift the value of each trig function by $(-1)^k$, we get $(-1)^k(\cos kx \cos ks + \sin kx \sin ks)$. We can now substitute these results into the equation to get

$$-f(x) + \frac{1}{\pi}\int_{-\pi}^{\pi}[\frac{1}{2} + \sum_{k=1}^{n}[A_k(\cos kx \cos ks + \sin kx \sin ks)]f(s)ds$$

Recall that the Fourier coefficients used in Lf are $a_k = \frac{1}{\pi}\int_{\pi}^{\pi} f(s)\cos ks\ ds$ and $b_k = \frac{1}{\pi}\int_{\pi}^{\pi} f(s)\sin ks\ ds$. We see that the expression becomes

$$\begin{aligned}
&- f(x) + \frac{1}{\pi}\int_{-\pi}^{\pi}\frac{1}{2}f(s)\ ds + A_k\sum_{k=1}^{n}\left[\int_{-\pi}^{\pi} f(s)\cos kx \cos ks + \int_{-\pi}^{\pi} f(s)\sin kx \sin ks\ ds\right] \\
&= -f(x) + \frac{a_0}{\pi} + A_k\sum_{k=1}^{n} a_k \cos ks + b_k \sin ks\ ds \\
&= Lf(x) - f(x)
\end{aligned}$$

□

We can now prove Jackson's Theorem using these lemmas. We see that lemma 3 gives us an integral representing the error of approximation of Lf to f, and lemma 2 then gives us the minimum value of this integral.

Theorem 2.7 (Jackson's Theorem 1)**.** *For all 2π-periodic and continuously differentiable functions f,*

$$E_n(f) \leq \frac{\pi}{2(n+1)}||f'||$$

and the constant $\pi/(2(n+1))$ is best possible.

Proof. From Lemma 3, we see that

$$\begin{aligned} E_n(f) &\leq ||Lf - f|| \\ &= ||\frac{1}{\pi}\int_{-\pi}^{\pi}\left[\frac{1}{2}t + \sum_{k=1}^{n}\frac{(-1)^k}{k}A_k \sin kt\right] f'(x+\pi-t)dt|| \end{aligned}$$

We see that if we choose the maximum absolute value of f', which is $||f'(x)||$, and leave the rest of the equation as an absolute value, then we can only increase the value of the product. Thus, we get

$$\leq ||f'||\frac{1}{\pi}\int_{-\pi}^{\pi}\left|\frac{t}{2} + \sum_{k=1}^{n}\frac{(-1)^k}{k}A_k \sin kt\right| dt$$

We now note that the integrand is symmetric over 0, so we can simplify the limits of integration and double the integrand to get

$$||f'||\frac{1}{\pi}\int_{0}^{\pi}\left|t + \sum_{k=1}^{n}\frac{2(-1)^k}{k}A_k \sin kt\right| dt$$

By Lemma 2, we know that for some choice of A_k the minimum is achieved, and so the value of the integral becomes $\pi^2/2(n+1)$. Thus, we see that

$$E_n(f) \leq ||f'||\frac{1}{\pi}\pi^2/2(n+1) = \frac{\pi}{2(n+1)}||f'||$$

and the upper bound is proven.

□

2.2.3 N-Widths

We can also use ρ in the comparison of subspaces to other subspaces. Given two subspaces in a metric space, we define the distance between space X and Y as $d(X, Y) = \sup_{x \in X} \rho(x, Y)$. Essentially, the distance between the two spaces is the distance of the point in the first space that maximizes the distance to the second space. This definition is used to determine the **n-widths** of a subspace. We define n-width for some metric space M and subspace S as

$$d_n(S, M) = \inf_{D \in \varsigma_n} d(S, D)$$

where ς_n is the set of all n-dimensional subspaces in M. There are a few notable properties of n-widths. To show these, we must simple recall some properties of the infemum and supremum. Namely, if $A \subseteq B$, then $\inf(A) \geq \inf(B)$, and $\sup(A) \leq \sup(B)$.

1. **Monotone in n:** For a space X and subspace S, then $n > m \implies d_n(S, X) \leq d_m(S, x)$. This can be seen by noting that, for any m-dimensional subspace A, there exists an n-dimensional subspace B such that $A \subset B$. This is simply the subspace A where for all dimensions $> m$, the value in B is just the additive identity, 0. Thus, the set of all n-dimensional subspaces contains all the m-dimensional ones. We see that $d(S, A) = \inf\{\rho(s, A) \forall s \in S\}$ and $d(S, B) = \inf\{\rho(s, B) \forall s \in S\}$. Since $A \subseteq B$, we see that $\{\rho(s, A)\} \subseteq \{\rho(s, B)\}$, because any point in A that defined the ρ would also exist in B. Thus, $d(S, A) \geq d(S, B)$, and because A and B were chosen generally to be any m- and n-simensional subspace respectively, we see that $d_n(S, X) \leq d_m(S, X)$.

2. **Monotone in X:** If Z is a linear subspace such that $Z \subset X$, then $d_n(S, X) \leq d_n(S, Z)$. Intuitively, we can see that Z is more constrained than X, and so our best approximation to Z should be higher. We can show this rigorously by noting that any n-dimensional subspace $A \subset Z$ is also contained in X. Recalling that $d(S, Z) = \sup_{N \in Z} \inf_{s \in S}(\rho(s, N))$ and $d(S, X) = \sup_{N \in X} \inf_{s \in S}(\rho(s, N)))$ where N are n-dimensional subspaces, we see that because $Z \subset X$, we can say $\inf_{s \in S}(\rho(s, N))N \in Z \subseteq \inf_{s \in S}(\rho(s, N))N \in X$, and thus by the property of supremums we get that $d_n(S, X) \leq d_n(S, Z)$.

For an example, let us consider the subspace $A = \{x \in \mathbb{R}^3 \mid |x_1|+|x_2|+|x_3|\} \leq 1$. We will show that $d_1(A, \mathbb{R}^3) = \sqrt{\frac{2}{3}}$.

First, for any 1-dimensional subspace Y we note that Y is a line through the origin, and call it $(y_1, y_2, y_3)t$. Consider the subspace $t(1,1,1)$. We see that for any point p in A,$d(p,Y) = \sqrt{(x_1-t)^2+(x_2-t)^2+(x_3-t)^2}$. Using the bounds on A, we get that $|x_3| \leq 1-|x_1|-|x_2|$, $x_1^2+x_2^2 \leq 1$ and if we consider only positive t and x_i, we see that

$$\begin{aligned} d(p,Y) \leq & \sqrt{(x_1-t)^2+(x_2-t)^2+(1-x_1-x_2-t)^2} \\ = & \sqrt{x_1^2-2x_1t+t^2+x_2^2-2x_2t+t^2+1+x_1^2+x_2^2+t^2-2x_1- 2x_2-2t+2x_1x_2+2x_1t+2x_2t} \\ = & \sqrt{3t^2-2t+2x_1^2+2x_2^2-2x_1-2x_2+2x_1x_2+1} \\ = & \sqrt{3t^2-2t+x_1(2x_1+x_2-2)+x_2(2x_2+x_1-2)+1} \\ \leq & \sqrt{3t^2-2t+1} \end{aligned}$$

We see that for the choice of $t = \frac{1}{3}$ the resulting distance is less than $\sqrt{3(1/3)^2-2/3+1} = \sqrt{\frac{2}{3}}$. Also, if we consider any negative x_i we can also use a negative t and arrive at the same conclusion.

We also note that for Y, there exists a point $p \in A$ such that $\rho(p,Y) \geq \sqrt{\frac{2}{3}}$. Consider the line through the origin $X = t\{y_1^{-1}, -y_2^{-1}, 0\}$. We see that this is orthogonal to Y, and intersects it at the origin. Therefore, the distance from any point $x \in X$ to Y is simply the distance from the origin. Let us then consider the point when $t = \sqrt{\frac{2}{3}}\frac{y_1y_2}{\sqrt{(y_1^2+y_2^2)}}$. Then $x = \{\sqrt{\frac{2}{3}}\frac{y_2}{\sqrt{(y_1^2+y_2^2)}}, \sqrt{\frac{2}{3}}\frac{y_1}{\sqrt{(y_1^2+y_2^2)}}, 0\}$, and $d(x,Y) = |x| = \sqrt{\frac{2}{3}\frac{y_1^2+y_2^2}{y_1^2+y_2^2}} =$

$\sqrt{\frac{2}{3}}$. We also note that

$$\begin{aligned} |x_1| + |x_2| + |x_3| &= \left|\sqrt{\frac{2}{3}}\frac{y_2}{\sqrt{(y_1^2 + y_2^2)}}\right| + \left|\sqrt{\frac{2}{3}}\frac{y_1}{\sqrt{(y_1^2 + y_2^2)}}\right| + 0 \\ &= \sqrt{\frac{2}{3}}\frac{y_1 + y_2}{\sqrt{(y_1^2 + y_2^2)}} \\ &\leq 1 \end{aligned}$$

Thus we have demonstrated that $\sqrt{\frac{2}{3}} \leq d_1(A, \mathbb{R}^3) \leq \sqrt{\frac{2}{3}}$, and so $d_1(A, \mathbb{R}^3) = \sqrt{\frac{2}{3}}$.

N-widths can apply to any metric space, not just the traditional Euclidean 3D space. Let us consider a second example. Let $X = \mathbb{R}^3$ with the maximum norm; $||x|| = \max\{|x_1|, |x_2|, |x_3|\}$, let $A = \{x : x_1^2 + x_2^2 + x_3^2 \leq 1\}$. Note that A is unit ball in the Euclidean norm. We will show that $d_1(A, X) = \frac{\sqrt{2}}{2}$.

First, we note that 1-dimensional subspaces are still lines in X. We see for any point $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$, $d(x, y) = \max\{|y_1 - x_1|, |y_2 - x_2|, |y_3 - x_3|\}$. Consider any point $x = (x_1, x_2, x_3)$ and line $Y = t(y_1, y_2, y_3)$. If we imagine the point $y \in Y$ that minimizes $d(x, Y)$, a few properties must hold. If $\max\{|y_1 - x_1|, |y_2 - x_2|, |y_3 - x_3|\} = |ty_i - x_i|$, then $|ty_i - x_i|$ must be at a local minimum with respect to t and be larger than the distance in the other dimensions, or $|ty_i - x_i| = |ty_j - x_j|$ for some dimension j, and the derivatives must be opposite signs. We see the first case is only minimized if $ty_i - x_i = 0$, but for this to be the maximum difference all three dimensions must be 0, and thus $x \in Y$. Otherwise, we find that $|y_i t - x_i| = |y_j t - x_j|$, and that $\operatorname{sgn}(y_j t * x_j) = -\operatorname{sgn}(y_i t * x_i)$ so that we cannot reduce both values by changing t. Further, if we look at which two dimensions vary the most from x, we see that these two must be i and j. Thus, if without loss of generality we say that $y_1 - x_1 \geq y_2 - x_2 \geq y_3 - x_3$, we find the minimum value at $|x_1 - ty_1| = |x_3 - ty_3|$.
Consider the line $Y = t(1, 1, 1)$. We see that for any points x and $Y(t)$, $d(x, Y(t)) = \max\{|t - x_1|, |t - x_2|, |t - x_3|\}$. Without loss of generality, as-

sume $x_1 \geq x_2 \geq x_3$. As we have seen, at a point minimizing t we must have that $|t - x_1| = |t - x_3|$. Solving for t gives us $t = \frac{x_1 + x_3}{2}$. Our condition on A implies that $x_1^2 + x_3^2 \leq 1$, therefore $x_1 + x_3 \leq \sqrt{2}$ and we see that for the line $Y = t(1, 1, 1)$, $d(A, Y) = \frac{\sqrt{2}}{2}$, and therefore $d_1(A, X) \leq \frac{\sqrt{2}}{2}$.

Let us now consider any line $Y = t(y_1, y_2, y_3)$. Without loss of generality, let $y_1 \geq y_2 \geq y_3$. Consider the point $x = (0, 0, 1) \in A$. We see that the set of distances is modelled by $|ty_1|, |ty_2|, |ty_3 - 1|$. Note that $|ty_1| \geq |ty_2| \geq |ty_3|$ by our choice of indexing. Thus, if we consider the closest point in Y to x, we see that $|ty_1| = |ty_3 - 1|$, otherwise some adjustment of t could lower the maximum value. Solving for t gives us $t = \frac{1}{y_1 + y_3}$, and thus $\rho(x, Y) \leq |\frac{y_1}{y_1 + y_3}|$. Recalling that $|y_1| > |y_3|$, This proves that for any line, there is a point at least $\frac{\sqrt{2}}{2}$ away, so $d_1(A, X) \geq \frac{\sqrt{2}}{2}$.

3 Weierstrass Approximation Theorem

3.1 Dirac sequences

First, let us define the convolution operator $*$ as

$$f * g(x) = \int_{-\infty}^{\infty} f(x - t)g(t)dt$$

We limit the functions to continuous functions on $\mathbb{R}$ that tend to 0 as $x \to \pm\infty$ so that all integrals we take converge.
Let us now define a Dirac sequence as a sequence of continuous functions $\{K_n\}$ that have the following properties:

1. $K_n \geq 0$ for all n.
2. $\int_{-\infty}^{\infty} K_n = 1$ for all n.
3. For any $\varepsilon, \delta > 0$ there exists N such that for all $n \geq N$, $\int_{|x| \geq \delta} K_n(x)dx < \varepsilon$.

Essentially, Dirac sequences have a constant infinite integral of 1, but as n becomes higher the distribution of the area under the curve becomes more

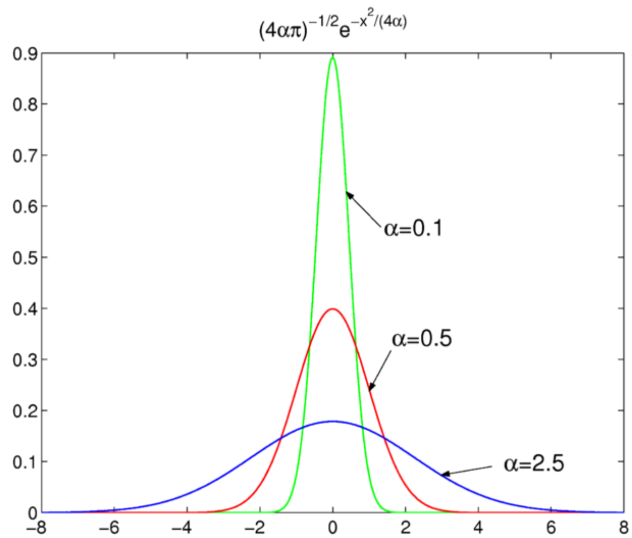

Figure 1: An example of a dirac sequence

closely centered around 0. We can now show that while the Dirac sequence does not converge to a function, its limit does serve as the identity function for convolution.

Theorem 3.1 (General approximation theorem)**.** *Let f be a bounded continuous function on $\mathbb{R}$. Then the sequence $K_n * f$ converges to f uniformly on any compact set. In other terms,*

$$\lim_{n\to\infty}(K_n * f)(x) = f(x)$$

and the convergence is uniform.

Lang [6] Provides a proof of this.

Proof. The proof relies only on the properties of the Dirac sequence, and the fact that convolution is commutative. Using the commutation and the definition of convolution, we see that

$$K_n(x) * f = f * K_n(x) = \int_{-\infty}^{\infty} K_n(t)f(t)dt$$

We can also rewrite $f(x)$ using the property 2, giving us

$$f(x) = f(x)\int_{-\infty}^{\infty} K_n(t)dt = \int_{-\infty}^{\infty} K_n(t)f(x)dt$$

If we now examine the difference, we see

$$f * K_n(x) - f(x) = \int_{-\infty}^{\infty} K_n(t)[f(x-t) - f(x)]dt$$

Because f is bounded, we know there exists a bound B such that $|f(x)| \leq B$ for all x. Recall that f is also uniform continuous, and thus for any $\varepsilon > 0$, there exists $\delta > 0$ such that $|f(x-\delta) - f(x)| < \varepsilon$. Extending this, we can conclude that if $|t| \leq \delta$,

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

With these bounds, we can estimate the error of approximation $|(K_n * f)(x) - f(x)|$. Let us start by taking the integral of this, and splitting it into two pieces to yield

$$\int_{t \leq \delta} K_n(t)|f(x-t) - f(x)|dt + \int_{t \geq \delta} K_n(t)|f(x-t) - f(x)|dt$$

By the bound on f we can conclude $|f(x-t) - f(x)| \leq 2B$, and by property 3 we can say there exists N such that for all $n \geq N$, $\int_{|t| \geq \delta} K_n(x)dx < \varepsilon$. Combining these lets us bound the right-most term, and gives us

$$\int_{t \geq \delta} K_n(t)|f(x-t) - f(x)|dt \leq \varepsilon 2B \forall n > N$$

In the left-most term, recall that $|f(x-t) - f(x)| < \varepsilon$ and K_n is positive, giving us

$$\int_{t \leq \delta} K_n(t)|f(x-t) - f(x)|dt \leq \int_{-\infty}^{\infty} K_n(t)\varepsilon dt = \varepsilon$$

Combining these, we can conclude that

$$|(K_n * f)(x) - f(x)| \leq \varepsilon + \varepsilon 2B$$

This proves that the Dirac sequence approximates f uniformly. □

3.2 Weierstrass Approximation Theorem

Using this property of Dirac sequences, we can prove the Weierstrass Approximation Theorem in a way that gives us a useful construct of the actual approximating polynomial.

Theorem 3.2 (Weierstrass Approximation Theorem)**.** *Let f be a continuous function on a finite closed interval $[a, b]$. Then f can be uniformly approximated by polynomials on $[a, b]$ to within an arbitrarily low degree of error.*

Proof. Recall that, without loss of generality, the function can be translated and dilated to be on the interval $[0, 1]$. Next, let us define a linear function L such that $L(0) = f(0)$ and $L(1) = f(1)$. L is just a polynomial of degree 1, so any approximation we find of f can just as easily be used on $f - L$. This allows us to assume that $f(0) = f(1) = 0$ with no loss of generality. Let us now define the **Landau sequence** as

$$K_n(x) = \begin{cases} \frac{1}{c_n}(1 - x^2)^n & : -1 \leq x \leq 1 \\ 0 & : |x| \geq 1 \end{cases}$$

We want to ensure that this sequence serves as a Dirac sequence. We can see that it satisfies property 1, as long as c_n is positive. As c_n is a constant of our choosing, let us choose it such that the second property is also satisfied. Namely, let $c_n = \int_{-1}^{1}(1 - x^2)^n$. We can see from this that

$$\int_{-1}^{1} K_n(x) = \frac{1}{\int_{-1}^{1}(1 - x^2)^n} \int_{-1}^{1}(1 - x^2)^n = 1$$

satisfying the second property.
This sequence also satisfies the third property of Dirac sequences. Consider any $\varepsilon > 0$ and $\delta > 0$. We see that

$$\int_{|x| \geq \delta} K_n(x)dx = \int_{-\infty}^{\infty} K_n(x)dx - \int_{-\delta}^{\delta} K_n(x)dx$$

Note that the infinite integral is just 1, as outside of $[-1, 1]$ the value is 0, and the integral of $[-1, 1]$ is 1 by the second property of the Diroc sequence. Thus,

$$\int_{|x| \geq \delta} K_n(x)dx = 1 - \frac{1}{c_n} \int_{-\delta}^{\delta}(1 - x^2)^n dx$$

If $\delta \geq 1$, then the second term in the equation becomes 1, and the integral is thus 0. Note that, for $\delta < 1$, $c_n > \int_{-\delta}^{\delta} K_n(x)dx$ because the function is strictly positive in this range. Also, as $n \to \infty$, $\frac{1}{c_n}\int_{-\delta}^{\delta} K_n(x)dx \to 1$. The whole integral thus converges to $1 - 1 = 0$, and so for some n_0 high enough, must pass below ε. This satisfies the third property, and we have shown that $\{K_n\}$ is a Dirac sequence.
Using the general approximation theorem, we now know that $K_n * f$ converges uniformly to f on $[0, 1]$. Thus, if $K_n * f$ is a polynomial, we have demonstrated a convergent polynomial for f and satisfied the theorem.
Recall that

$$K_n * f = \int_{-1}^{1} K_n(x - t)f(t)dt$$

$K_n(x - t)$ is a polynomial, and can be expanded to $\frac{1}{c_n}(1 - (x - t)^2)^n = \frac{1}{c_n}(1 - x^2 + 2xt - t^2)^n$. Now using a binomial expansion in x and t we can get $\sum_{i+j=n} a_{ij}x^i t^j$ where $a_{ij} \in \mathbb{R}$. Thus, the convolution can be written as

$$K_n * f = \sum_{i+j=n} a_{ij}x^i \int_{-1}^{1} t^j f(t)dt = \sum_i \sum_j a_{ij} \int_{-1}^{1} t^j f(t)dt x^i$$

We have shown that $K_n * f$ converges to f, and that $K_n * f$ is a polynomial. This proves the theorem, giving us the existence of a uniform convergent polynomial for any f. Importantly, this proof also gives us the explicit means of constructing such a polynomial. □

3.3 Chebyshev Equioscillation Theorem

We have now shown that there always exists a convergent series of polynomials for any continuous function. We have even given one way of constructing such a polynomial using Dirac sequences. However, what is of equal interest is the task of approximating constrained polynomials, such as those of degree at most n. We have already shown two examples to doing this for specific types of functions: smooth functions by using the Jackson Inequality Theorems, and functions of the type $f(x) = \frac{1}{x - a}$.

When examining special cases such as this, it can be very difficult to establish just when a polynomial is exactly the best polynomial of that degree. However, we do have a powerful tool for this in the form of Chebyshev's Equioscillation Theorem, which have already used in a proof above. This theorem essentially builds the fact that an n-degree polynomial can only oscillate $n+2$ times to show that if the maximum distance from a function to a polynomial is achieved exactly this many times, then the polynomial is the best approximation. The following proof is an original expansion of the de La Valle Poussin Theorem, which only provides a lower bound on the approximation.

Theorem 3.3 (Chebyshev Equioscillation Theorem)**.** *Let f be a continuous function on $[a,b] \in \mathbb{R}$, and p a polynomial of degree n. Among all polynomials of degree $\leq n$, p is the best approximation of f (meaning p minimizes $||f-p||$) if and only if there are $n+2$ points such that $a \leq x_0 < x_1 < x_2 \dots x_{n+1} \leq b$ such that $f(x_i) - p(x_i) = \sigma(-1)^i ||f-g||$ where $\sigma = \pm 1$.*

Proof. Let us first define an alternating set, where we note that the set used in the lemma above was an alternating set of length 2. An alternating set of function $f \in C[a,b]$ and polynomial p of degree n is $X = \{x_0, x_1 \dots x_{n-1}\}$ such that $a \leq x_0 < x_1 < \dots < x_{n-1} \leq b$ and $x_i = (-1)^i e_i$ where e_i is strictly positive or negative for all i. Then length of an alternating set is the value n. We call this set uniform if $e_i = E$ for all i.

What we will show is that if there is a uniform alternating set of length $n+2$, and if there exists a better approximating polynomial of equal degree, then the difference of these two polynomials will have $n+2$ roots, which is not possible for degree n polynomials. We show that $d_n(f) \geq ||f-p||$ when a uniform alternating set X of length $n+2$ exists, which of course proves that p is the polynomial of best approximation of f.
Let us define $E = ||f-p||$. Suppose there exists a polynomial of best approximation q, which implies that $||f-q|| <= E$. We see that, for every $x_i \in X$

$$|f(x_i) - q(x_i)| < E = |f(x_i) - p(x_i)|$$

Let us consider the polynomial $p-q$, and show that X is an alternating series on this as well. First, note that $p-q = p-f-(q-f)$. Let us first consider all points x_i such that i is odd, and assume that $\sigma = 1$, meaning $(p-f)(x_i) > 0$. We see that $(p-q)(x_i) = (p-f)(x_i) - (q-f)(x_i) = E - e_i$ where $|e_i| < E$.

We can limit the range of this point to get $(p-q)(x_{2i+1}) \in [0, 2E]$. If we do the same examination for the even i's, we see that $(p-q)(x_{2i}) \in [-2E, 0]$. Note that if $\sigma = -1$ we can switch the even and odd indices to arrive at the same conclusion.
Noting that there are $n+2$ such points, we see that X is an alternating set on $p-q$, and thus $p-q$ changes sign at least $n+2$ times. However, $p-q$ is of degree at most n, and thus $p-q=0$ and we see that p is the best approximation.

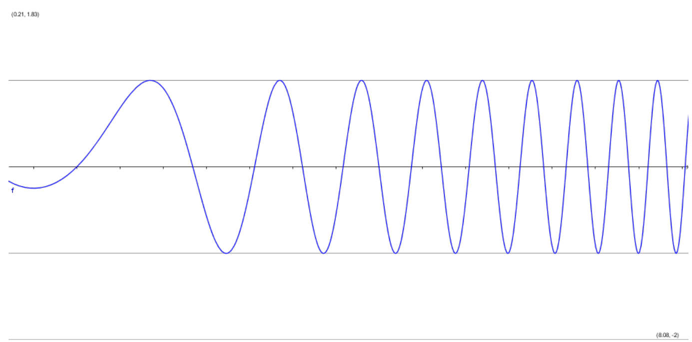

Figure 2: What (f - p)(x) must look like

□

We can also begin to see the other direction, that in fact all best approximations share some properties, by examining the weaker case in this proof.

Theorem 3.4. *Let f be a continuous function on $[a, b]$ and suppose that p is the best approximating polynomial of degree n. There exists points $a \leq x_1 < x_2 \leq b$ such that $f(x_1) - p(x_1) = E||f-p||$ and $f(x_2) - p(x_2) = -E||f-p||$ where $E \in \{\pm 1\}$.*

Proof. We begin by showing that the minimum and maximum must be equal, or a better approximation would exist. Let x_0 be the minimum and x_1 the maximum of $f(x) - p(x)$. Because the functions are defined on a closed set, we know the min and max exist on this set. Let $m_0 = f(x_0) - p(x_0)$ and $m_1 = f(x_1) - p(x_1)$. We see that if these are not opposite in sign and of equal magnitude, we could add a constant to p to get a better approximation.

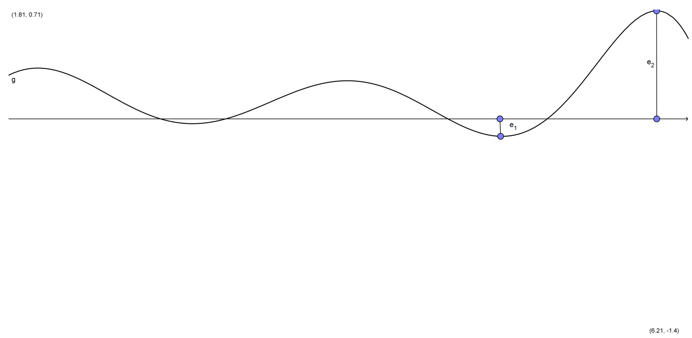

Figure 3: $f-p$ of a non-optimal approximation

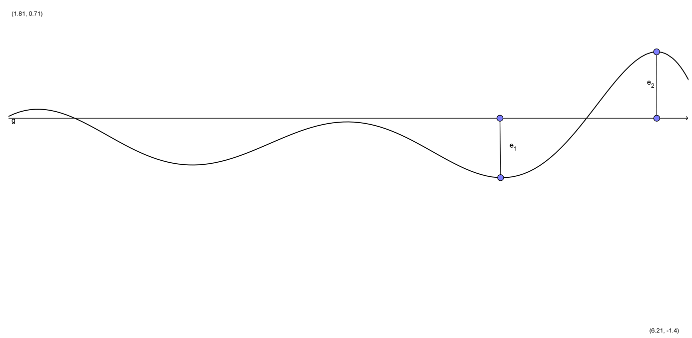

Figure 4: A better approximation

Specifically, we could construct a polynomial $q(x) = p(x) + \frac{m_0 + m_1}{2}$, and notice that because

$$m_0 \leq f(x) - p(x) \leq m_1$$

then we can say

$$m_0 - \frac{m_0 + m_1}{2} \leq f(x) - q(x) \leq m_1 - \frac{m_0 + m_1}{2}$$

Simplifying gives us

$$\frac{m_0 - m_1}{2} \leq f(x) - q(x) \leq \frac{m_1 - m_0}{2}$$

Thus, we see that $||f-q|| = |\frac{m_0 - m_1}{2}|$, while $||f-p|| = \max\{|m_0|, |m_1|\}$. Thus, either $m_0 = -m_1$ so that $||f-q|| = |m_0| = \max\{|m_0|, |m_1|\} = ||f-p||$ or q is a better approximation than p. $\square$

4 Lethargy

4.1 Bernstein Lethargy Theorem

We have seen now that for any function in $C[-1,1]$ we can find an approximating polynomial to any arbitrarily small degree of error. However, this says nothing of the speed of convergence. This is a vital question if we want to save any computational time by using an approximation, because if a polynomial of sufficiently low degree of error happens to have an arbitrarily large degree, there is little practical use to be found in finding it. What we first show is that there are in fact functions that can cause polynomials to converge arbitrarily slowly to 0. Bernstein originally established this on $C[-1,1]$ and Timan later generalized it to all Banach spaces. Borodin[2] gives a proof of this.

Theorem 4.1 (Lethargy). *Given a Banach space X and a series of nested finite-dimensional subspaces $Y_1 \subset Y_2 \subset \ldots$, if $\{d_1, d_2 \ldots\}$ is a monotone decreasing sequence converging to 0, then there exists a point $x \in X$ such that $\rho(x, Y_k) = d_k$ for all $k > 0$.*

Proof. We first examine the case of a finite series of nested subspaces and values, let us call these $Y_1 \subset Y_2 \ldots Y_n \subset X$ and $d_1 > d_2 > \cdots > d_n \geq 0$. We can create a series of points that follows the nesting of our sets by choosing $x_k \in Y_{k+1}$
Y for $k = 1, 2 \ldots n-1$. We create an initial scalar λ_n such that

$$\lambda_n = \frac{d_n}{\rho(x_n, Y_n)}$$

Note that this gives us $\rho(\lambda_n x_n, Y_n) = d_n$.

We now need a value ε such that $d_k + \varepsilon \leq d_{n-1}$. We can find this by choosing $\varepsilon = \min\{d_1 - d_2, d_2 - d_3 \ldots d_{n-1} - d_n\}$. In the subspace Y_n, we know that some point p is a limit point of Y_n such that $||\lambda_n x_n - p|| = \rho(\lambda_n x_n, Y_n) = d_n$ because this is the definition of ρ. Thus, we can say that there is some element $y_n \in Y_n$ within ε of p, and conclude that $||\lambda_k x_k - y_k|| \leq d_k + \varepsilon$.

Let us define a function $f(t) = \rho(\lambda_n x_n - y_n + tx_{n-1}, Y_n)$. First, we see that this function is continuous for all $t \geq 0$, because if we consider $f(t+\delta) = \rho(\lambda_n x_n - y_n + tx_{n-1} + \delta x_{n-1}, Y_n) \leq \rho(\lambda_n x_n - y_n + tx_{n-1}, Y_n) + \delta||x_n||$, we can choose an arbitrarily low δ to be within any disk. We also note that

$$f(0) = \rho(\lambda_n x_n - y_n, Y_n) \leq ||\lambda x_n - y_n|| \leq d_n + \varepsilon \leq d_{n-1}$$

and

$$\lim_{t \to \infty} f(t) = \infty$$

By the Intermediate Value Theorem, we see that there is some value λ_{n-1} such that $f(\lambda_{n-1}) = d_{n-1}$, and therefore $\rho(\lambda_n x_n - y_n + \lambda_{n-1} x_{n-1}, Y_n) = d_{n-1}$. In Y_{n-1} we can also find an element y_{n-1} such that

$$||\lambda_n x_n - y_n + \lambda_{n-1} x_{n-1} - y_{n-1}|| \leq d_{n-1} + \varepsilon$$

using the same argument that was used to find y_n. From this, we conclude that

$$\rho(\lambda_n x_n - y_n + \lambda_{n-1} x_{n-1} - y_{n-1}, Y_n) = \rho(\lambda_n x_n, Y_n) = d_n$$

and

$$\rho(\lambda_n x_n - y_n + \lambda_{n-1} x_{n-1} - y_{n-1}, Y_{n-1} = \rho(\lambda_n x_n - y_n + \lambda_{n-1} x_{n-1}, Y_{n-1}) = d_{n-1}$$

We can repeat this procedure for all our x_i and Y_i until we arrive at an element $x = \lambda_n x_n - y_n + \lambda_{n-1} x_{n-1} - y_{n-1} + \cdots + \lambda_2 x_2 - y_2 + \lambda_1 - y_1$ such that

1. $\rho(x, Y_k) = d_k$ for all $k = 1, 2 \ldots n$
2. $||x|| \leq d_1 + \varepsilon$
3. $x - \lambda_n x_n \in Y_n$

With this element known to exist for every finite set of subspaces and real numbers, we can expand to the countably infinite case. We construct a sequence of elements $x_m \in Y_{m+1}$ such that x_m is the element proven to exist above, and $\rho(x_m, Y_k) = d_k$ for all $k < m$. Noting that the given real values converge to 0, there must be some point $d_N < 1$, which means that $\varepsilon_m < 1$ for all finite cases where $m > n$. Thus we also get $||x_m|| \leq d_m + 1$ for all $m > N$.

Let y_{mk} be the element in Y_k that is closest to x_m. If we hold k constant, we get the sequence $\{y_{km}\}$ for all $m \geq k$. By definition, we get that

$$||y_{km}|| \leq ||x_m|| + ||x_m - y_{km}|| \leq d_1 + 1 + d_k \leq 2d_1 + 1$$

The magnitude of each y_{km} is thus bounded, and resides within a finite subspace Y_k, meaning that the closure of the sequence is bounded and thus compact. Because the sequence we have is infinite, it must have a convergent subsequence, which must be Cauchy. Let us call the indices of this sequence A such that $\{y_{km}\}_{m\in A}$ is the Cauchy subsequence. Recall that this is general for all k. We see that if we make a matching sequence of $\{x_m\}_{m\in A}$ then this sequence is also Cauchy. Examine the difference for some $l > m$

$$||x_m - x_l|| \leq ||x_m - y_{km}|| + ||x_l - y_{kl}|| + ||y_{kl} - y_{km}|| = 2d_k + ||y_{kl} - y_{km}||$$

We see $d_k \to 0$ as k grows, and $||y_{kl} - y_{km}|| \to 0$ because that sequence is Cauchy. Thus we know that this sequence converges to a limit, and the limit must meet the distance requirements of all the previous elements. This gives us an element x such that $\rho(x, Y_k) = d_k$ for all k.

□

This is a powerful theorem, and essentially proves that for some functions, the rate of convergence of the best approximation can be arbitrarily slow, if we choose an decreasing sequence that converges to 0 extremely slowly. Thus, for any given function or type of function, we are never guaranteed the existence of a bound on $\rho(f, P_n)$, and thus must evaluate them on a case-by-case basis. Bernstein's Theory has been improved by Borodin [2] to include infinite-dimensional subspaces as well. To see this, we must first prove a lemma.

Lemma 4.2. *Let $d_1 \geq d_2 \geq \cdots \geq d_n \geq 0$ be a finite decreasing sequence and $Y_1 \subset Y_2 \subset \cdots \subset Y_n$ be a system of strictly nested subspaces of Banach space X that meets the following property: There exists a series of elements q_n such that $q_n \in X\ Y_n, q_k \in Y_{k+1}\ Y_k, k = 1, \ldots, n-1$, such that any element q in the linear span $\langle q_k, q_{k+1}, \ldots q_n\rangle$ satisfies the inequality*

$$\frac{\rho(q, Y_k)}{\rho(q, Y_{k-1})} \geq \frac{d_k}{d_{k-1}}, k = 2, 3, \ldots, n$$

There is some element x in the cone $\langle q_1, q_2, \dots q_n\rangle^+$ (the linear combinations of $q_1 \dots q_n$ with positive nonnegative) which satisfies $\rho(x, Y_k) = d_k$, $k = 1, \dots n$.

We see this is a very similar claim to the beginning of Theorem 4.1, except that the condition of finite-dimensional subspaces has been replaced with the above inequality.

Proof. We will prove this by induction on n. When $n = 1$, we see that the desired element is trivially

$$x = \frac{d_1}{\rho(q_1, Y_1)} q_1$$

because $\rho(x, Y_1) = \dfrac{d_1}{\rho(q_1, Y_1)} \rho(q_1, Y_1) = d_1$.
When $n = 2$, let us consider a similar construction to the above case, but vary it linearly by q_1. Namely, define

$$x(\lambda) = \frac{d_2}{\rho(q_2, Y_2)} q_2 + \lambda q_1$$

Because $q_1 \in Y_2$ and subspaces are closed under scalar multiplication, the q_1 term drops out when considering rho, and we get

$$\rho(x(\lambda), Y_2) = \rho(\frac{d_2}{\rho(q_2, Y_2)} q_2, Y_2) = d_2$$

To find the correct value for lamba, we note that

$$\rho(x(0), Y_1) = \rho(\frac{d_2}{\rho(q_2, Y_2)} q_2, Y_1) = \frac{d_2}{\rho(q_2, Y_2)} \rho(q_2, Y_1) \leq d_1$$

because $\dfrac{\rho(q_2, Y_2)}{\rho(q_2, Y_1)} \geq \dfrac{d_2}{d_1}$ by our assumptions. Also, $\rho(x(\lambda, Y_1)) \to \infty$ as $\lambda \to \infty$. We recall that $rho(x(\lambda), Y_1)$ is a continuous function with regards to x, and so there must be some λ_1 value that satisfies $\rho(x(\lambda_0) Y_1) = d_1$.

Let us now consider when $n \geq 3$, and assume the theorem has been proven for $n - 1$. Let us take any nonzero element $u \in \langle q_{n-1}, q_n\rangle^+$ and any element $q \in \langle q_k, \dots, q_{n-2}, u\rangle$. By our assumptions, we have that

$$\frac{\rho(q, Y_k)}{\rho(q, Y_{k-1})} \geq \frac{d_k}{d_{k-1}}$$

$$\frac{\rho(q, Y_{n-1})}{\rho(q, Y_{n-2})} \geq \frac{d_{n-1}}{d_{n-2}}$$

By our inductive hypothesis, there is some element $x(u) \in \langle q_1, \ldots, q_{n-2}, u\rangle^+$ such that $\rho(x(u), Y_k) = d_k, k = 1 \ldots, n-1$. Let us define a scalar $\lambda(u)$ such that $x(u) = \lambda(u)u$. We have that $\rho(x(u), Y_{n-1}) = d_{n-1} = \lambda(u)\rho(u, Y_{n-1})$, so therefore $\lambda(u) = \dfrac{d_{n-1}}{\rho(u, Y_{n-1})}$. Clearly, $\lambda(u)$ is continuous on u. Thus, $\rho(x(u), Y_n) = \rho(\lambda(u)u, Y_n)$ and is also continuous on u. We see that $\rho(x(q_{n-1}, Y_n)) = 0$ and that

$$\rho(x(q_n), Y_n) = \rho(\lambda(q_n)q_n, Y_n) = \frac{d_{n-1}}{\rho(q_n, Y_{n-1})}\rho(q_n, Y_n) \geq d_n$$

Therefore there must be some u_n such that $x = x(u_n)$ satisfies $\rho(x, Y_n) = d_n$.
□

We have now shown the case for a finite number of infinite-dimensional subspaces. Similar to Bernstein's proof, we can now extend this to the infinite case, by examining the infinite sequence and finding a Cauchy subsequence.

Theorem 4.3. *Let $d_1 \geq d_2 \geq \cdots \geq 0$ be a decreasing sequence converging to 0 and $Y_1 \subset Y_2 \subset \cdots \subset X$ be a system of strictly nested subspaces of Banach space X that meets the following property: There exists a series of elements q_n such that $q_n \in Y_{n+1}\ Y_n$ such that for all $k \in I\!N$ any element q in the linear span $\langle q_k, q_{k+1}, \ldots q_n\rangle$ satisfies the inequality*

$$||q|| \leq \frac{d_k - 1}{d_k}\rho(q, Y_k)$$

There is some element x in the closure $\overline{\langle q_1, q_2, \ldots q_n\rangle}$ which satisfies $\rho(x, Y_n) = d_n$ for all $n \in N$.

Proof. First, let's consider the ratio between rho values. We see that $d_k leq d_{k-1}$ and therefore $||q|| \geq \rho(q, Y_k)$. This lets us conclude that

$$\frac{\rho(q, Y_k)}{\rho(q, Y_{k-1})} \geq \frac{\rho(q, Y_k)}{||q||} \geq \frac{d_k}{d_{k-1}}$$

We have satisfied the conditions of 4.2, so there exist elements $x_n \in \langle q_1, \ldots, q_n\rangle$ such that $\rho(x, Y_k) = d_k$ for $k = 1, \ldots, n$.
We find that the sequence x_n contains a convergent subsequence. Split x_n

into two groups of seperate factors at some point k, namely as $x_n = y_k^n + v_k^n$ where $k < n$, $y_k^n \in \langle q_1, \ldots, q_{k-1} \rangle$ and $v_k^n \in \langle q_k, q_k + 1, \ldots \rangle$. Using our given assumptions, we see that

$$||v_k^n|| \leq \frac{d_{k-1}}{d_k} \rho(v_k^n, Y_k) = \frac{d_{k-1}}{d_k} \rho(x_n, Y_k) = d_{k-1}$$

$$||y_k^n|| \leq ||v_k^n|| + ||x_n|| \leq \frac{d_0}{d_1} \rho(x_n, Y_1) + d_{k-1} = d_0 + d_{k-1}$$

We can now diagonalize to a convergent subsequence. First, we notice that for a fixed k, $\langle y_k^n \rangle$ is bounded, and also lies within a finite-dimensional subspace, meaning it must converge to some y_k.Thus, we look at a sequence of these y_n and notice that for any m, n and $k < \min\{m, n\}$ we get that

$$\begin{aligned} ||x_n - x_m|| &\leq ||y_k^n - y_k^m|| + ||v_k^n|| + ||v_k^m|| \\ &\leq ||y_k^n - y_k^m|| + 2d_{k-1} \\ &\leq y_k + 2sk - 1 \to 0 \end{aligned}$$

Thus, we see that $\{x_n\}$ has a Cauchy subsequence, which converges to the desired element x. $\square$

4.2 Lethargy and Smoothness

Here we provide an example of applications of the Bernstein Lethargy Theorem, which is used to show that Jackson's bounds on smooth functions only apply to globally smooth ones, and that if we even relax a single endpoint from the smoothness condition we can find arbitrarily slowly convergence. First, we define some terms.

Definition 4.4 (Approximation Scheme)**.** Let X be a quasi-Banach space (The Triangle Inequality is replaced with $||x+y|| \leq K(||x|| + ||y||)$ for some $K > 1$.), and let $\{0\} = A_0 \subset A_1 \subset \cdots \subset A_n \subset \cdots \subset X$ be an infinite chain of strictly inclusive subsets. We say that $\{A_n\}$ is an **approximation scheme** of X if the following hold:

(i) There exists a map $K : \mathbb{N} \to \mathbb{N}$ such that $K(n) \geq n$ and $A_n + A_n \subseteq A_{K(n)}$ for all $n \in \mathbb{N}$.

(ii) For every scalar λ, $\lambda A_n \subset A_n$.

(iii) $\cup A_n$ is a dense subset in X.

A common example of an approximation scheme is the sequence of sets of polynomials of degree n in the space of continuous functions. Also of note, if the sets A_n are linear subspaces, then $K(n) = n$ and we call this a *Linear Approximation Scheme.*

Definition 4.5 (de La Vallee Poisson Theorem)**.** This is a relaxation of the Chebyshev Equioscillation Theorem. Let P_n be the polynomials of degree n, f be a continuous function and $\{X_n\}$ be an alternating series on function f of size $n+2$. If $\min_{x \in X_n} |f(x)| = \varepsilon$ then $\rho(f, A_n) \geq \varepsilon$.

We see that this is simply a relaxed version of the Equioscillation Theorem, where we do not require the alternating set to be uniform, and thus achieve only a lower bound on the degree of error. Proof 3.3 can be used to prove this as well, by simply changing the conditions of the alternating set. We now consider cases when other approximation schemes share this property, not necessarily just polynomials, to get the de La Vallee Poussin Condition.

Definition 4.6 (de La Vallee Poussin Condition)**.** Let $\{A_n\}$ be an approximation scheme in $C[a,b]$ such that $A_0 \subset A_1 \subset \cdots \subset A_n \subset \cdots \subset C[a,b]$. We say that $\{A_n\}$ satisfies the de La Vallee Poisin condition if for every A_n there exists a number $m_n \in N$ such that for any function $g \in A_n$, an alternating series X on $|f-g|$ of size m_n guarantees $\rho(f, A_n) \geq \varepsilon_n$.

We see that for polynomials, $m_n = n+2$ where n is the degree of the subspace. We can in fact find many approximation schemes that meet this condition, especially if we can limit the number of zeros for any given dimension.

Definition 4.7. The vector space $A \subset C[a,b]$ is named Haar on $[a,b]$ if dim $A = n$ and the only element from A which has more than $n-1$ zeros is the null function.

We see that all polynomial nested sets satisfy the de La Vallee Poussin Condition, but it in fact extends to all Haar sets of linear subspaces of increasing dimension.
We also include a theorem by Carleman [4], a proof of which can be found in [5] Chapter 1 Theorem 4.3. This is essentially a version of the Weierstrass Theorem adapted to entire functions instead of polynomials.

Theorem 4.8 (Carleman's Theorem)**.** *[4] For each $f \in C(I\!R)$ and each $g \in C(I\!R), g(x) > 0, x \in I\!R$, there exists an entire function e such that*

$$|f(x) - e(x)| < g(x), x \in I\!R$$

We are now ready to prove Almira's[1] theorem, which states that for any decreasing positive sequence of numbers, we can find a continuous function that is analytic on all but one endpoint of the interval, and whose best approximation in any sequence of subspaces is still bounded by the sequence of numbers. Essentially, this shows that for functions where a single point is not smooth, the smoothness of the rest of the function cannot put a bound on the error of approximation.

Theorem 4.9 (Almira's Theorem)**.** *[1] Let $0 < \alpha < \beta$ and $[a, b] = [0, 1]$ or $[\alpha, \beta]$. Let $\{A_n\}$ be an infinitely nested sequence of subspaces in $C[a, b]$ that meets the de La Vallee Poussin condition with the set $\{M_n\} = \{m_0, m_1 \dots\}$ such that m_n is the size needed to apply the de La Vallee Poussin Theorem to A_n. Let $\{\varepsilon_n\}$ be a non-increasing sequence in $I\!R^+$ that converges to 0. Then there exists a function $f \in C[a, b]$ such that f is real and analytic on $(a, b]$ and $\rho(f, A_n) \geq \varepsilon_n$ for all n.*

Proof. We will first prove the case for $[a, b] = [0, 1]$. If $\{\varepsilon_n\}$ is constant the result is trivial, so we can assume $\varepsilon_n > 0$. Let us define a polygonal line $p : \mathbb{R} \to \mathbb{R}$ with vertices $\{n, p(n)\}$ for $n \in Z$. Thus, this line has vertex on every integer of the x-axis, and the y value is dependent on x. Let us define p as

(i) $p(t)$ is even, and $p(0) = p(1) = \cdots = p(m_0 - 1) = 3\varepsilon_0$

(ii) $p(m_0+m_1+\cdots+m_n+k) = 3\varepsilon_{n+1}$ for all $n \in \mathbb{N}$ and $k = 0, 1, \dots, m_{n+1}-1$

What this gives us is a polygonal path that is constant at $3\varepsilon_0$ at all points before the first alternating point m_0, then decreases by steps, so between m_0 and m_1 it equals ε_1, between m_1 and m_0+m_1 it equals ε_2, etc. If we consider dividing this line by 3, it remains continuous and positive, so by 4.8 there exists an entire function $e(t)$ such that

$$|p(t) - e(t)| \leq \frac{p(t)}{3}, t \in \mathbb{R}$$

If we examine the range of values e can take, we see that $\frac{1}{3}p(t) \leq e(t) \leq \frac{5}{3}p(t)$ and thus $e([0, m_0 - 1]) \subseteq (\varepsilon_0, 5\varepsilon_0)$, and in the general case $e([m_0 + m_1 + \cdots +$

$m_n, m_0 + m_1 + \cdots + m_{n+1} - 1]) \subseteq (\varepsilon_{n+1}, 5\varepsilon_{n+1})$.
We are now ready to construct a function f. Let $f(0) = 0$, and for $t \in (0, 1]$, $f(t) = e(\frac{1}{t}) \cos \frac{2\pi}{t}$. We see that $\lim_{t\to 0} e(\frac{1}{t}) = \lim_{n\to\infty} \varepsilon_n = 0$, thus f is continuous on $[0, 1]$ and clearly analytic on $(0, 1]$. Let us consider $\rho(f, A_n)$ for any n. A_n satisfies the de La Vallee Poussin Theorem for alternating series of size m_n, so we see that we can construct such a series in the range, such as

$$\frac{1}{m_0 + m_1 + \cdots + m_{n-1}} \leq t_0 < t_1 \cdots < t_{m_n} \leq \frac{1}{m_0 + m_1 + \cdots + m_n - 1}$$

such that $\operatorname{sgn}(t_i t_{i+1}) = -1$. We see that $\varepsilon_n \leq e(\frac{1}{t}) \leq 5\varepsilon_n$ in this range, and that $\cos 2\pi t$ goes from $\cos(2\pi(m_0 + m_1 + \cdots + m_{n-1})$ to $\cos(2\pi(m_0 + m_1 + \cdots + m_n - 1)$. Thus, it cycles m_n times, so we can find an alternating series of size m_n. By construction, $e(t) \geq \varepsilon_n$ within this range, so applying de La Vallee Poussin theorem we have shown that $\rho(f, A_n) \geq \varepsilon_n$. This applies for all n and so we have ended the proof for $[a, b] = [0, 1]$. To apply this to an arbitrary $[a, b]$, we can construct a new function $g(t) = f(\frac{t-a}{b-a})$ This is a continuous map, so g is still analytic on $(a, b]$ and continuous on $[a, b]$ and we can see that the above arguments imply that $\rho(g, A_n) \geq \varepsilon_n$

□

4.3 Finite Point Approximations

We have seen that for some functions, the degree of approximation may reduce arbitrarily slowly as we increase the degrees of the polynomial. In these cases it is often impractical to actually find such polynomials of high enough degree to meet a required bound on error, in fact such polynomials could be of arbitrarily high degree.
However, we may still wish to get some form of approximation on these functions. An inconvenient function is often an essential one, and so we turn to ways to approximate the approximation. We do this by finding a finite set of points within the interval to approximate instead. In order for this to be a valid way to estimate approximations, we must show that we can find a best approximation on a finite set, and that these approximations will approach the real one as the number of points approaches infinity.
To begin, we refer to a set of points X_m as m distinct points such that

$x_1 < x_2 < \dots x_m$ and all are within the interval we wish to approximate, $[a, b]$. Consider approximating these points with a polynomial of degree n. We note that if $m \leq n+1$, then a polynomial can be made to match each point exactly, with no degree of error. Thus, we will only consider the situations when $m > n+1$. We note that this relates to 3.3, and this is no coincidence. Using the same proof that we did for Chebyshev's Theorem, we find that a best approximation p_n of X_m also must oscillate $n+2$ times at the degree of approximation.

Theorem 4.10. *$p_n(X_m)$ is the best approximation on X_m to f if there exists an alternating set of $n+2$ points in X_m such that $f(x_i) - p_n(x_i) = -(f(x_j) - p_n(x_j)$ for all even i and odd j.*

A clear corollary of this theorem is that the best approximation of a finite set of points is unique as well. If two polynomials of degree n match values at $n+2$ places, they must be the same polynomial.
We next show that in fact, $n+2$ points is sufficient to find a best approximation. We can reduce the best approximation on $[a, b]$ or on X_m to finding it for some X_{n+2}.

Theorem 4.11. *If $p_n \in P_n$ is the best approximation to f on $[a, b]$, there exists X^*_{n+2} such that*

$$\rho(f, [a, b]) = \rho(f, X^*_{n+2})$$

Additionally, for all $X_{n+2} \subset [a, b]$,

$$\rho(f, X_{n+2}) \leq \rho(f, [a, b])$$

Proof. The first part of the theorem is simply a result of Chebyshev's Theorem. If X^*_{n+2} is an alternating set for $f - p^*_n(x)$, then by definition $||f - p^*_n(x)|| = ||p_n(X_{n+2})^* - f)||$.
If we suppose that the best approximation on the points, $p_n(X_{n+2})^*$ is not equal to the best approximation p^*_n, then by the uniqueness of p^*_n we get

$$\rho(f, X_{n+2}) \leq \rho(f, [a, b])$$

and note that the equality is only possible when $p^*_n = p_n(X_{n+2})^*$. □

This theorem tells us that finding the alternating set of points that gives us the highest degree of error will give us the approximating polynomial. While this is not terribly useful on $[a, b]$, as there are infinitely many sets of

$n+2$ points, we can also apply it to a finite set of points X_m, and there are now only finitely many combinations of points to test. If we know $\rho(f, X_m)$ ahead of time, then we can find the optimal set of points to approximate on when we find a set where $\rho(f, X_{n+2}) = \rho(f, X_m)$.

References

[1] Almira, J. M., *A Lethargy Result for Real Analytic Functions* Applied Mathematics Letters 25.6 (2012): 1016-018. Web.

[2] Borodin, P. A., *On the existence of an element with given deviations from an expanding system of subspaces*, Mathematical Notes, Vol. 80, No. 5, 621-630, (2006). (Translated from Mathematicheskie Zametki).

[3] Butzer P. and and H. Berens,*Semi-groups of Operators and Approximation*, (1967).

[4] Carleman T., Sur un thorme de Weierstrass, Arkiv. Mat. Astron. Fys. 20 (1927) 15.

[5] Cheney, E. W., *Introduction to Approximation Theory*, (1966).

[6] Lang, S., *Math Talks for Undergraduates*, pp. 52, (1999).

[7] Lorentz, G., *Constructive Approximation: Advanced Problems*, pp. 2 (1996).

[8] Mayans, R., "The Chebyshev Equioscillation Theorem," Convergence (February 2007). Web.

[9] Oja, E., *On Bounded Approximation Properties of Banach Spaces* Banach Center Publ. Banach Center Publications 91 (2010): 219-31. Web.

[10] Pinkus, A., *N-widths in Approximation Theory.* Berlin: Springer-Verlag, 1985. Print.

[11] Rivlin, T. J., *An Introduction to the Approximation of Functions.* Waltham, MA: Blaisdell Pub., 1969. Print.

[12] Timan, A. F., *The Theory of Approximating Functions of Real Variables.* Oxford: Macmillan, 1973. Print.