CONSTRUCTION OF A NON-STANDARD INTEGRAL ON *AC* [0, 1]

THESIS

Presented to the Graduate Council of
Southwest Texas State University
in Partial Fulfillment of
the Requirements

For the Degree

Master of SCIENCE

by

Gerard Leonard Ornas, Jr., B.S.

San Marcos, Texas
August, 1998

ACKNOWLEDGEMENTS

I would like to thank Dr. Stanley G. Wayment for his guidance, advice, and help with this paper and throughout my studies here. He has made this paper possible and has shown me entirely more patience than I deserve.

I would like to also thank the Mathematics faculty here at SWTSU, especially Dr. Thomas Thickstun and Dr. Terence McCabe for all they have taught me and all the help they have given me.

I would like to thank my family for all their help, support, emotional, physical, and fiscal through ten years of college and graduate school, with more to come!

Finally, I would like to thank my grandmother, Margaret Bertha Zunker (18 June 1919 – 11 February 1998) for giving me a place to live for the past three years, and allowing me to get this far.

Thank You.

July 1998

TABLE OF CONTENTS

ABSTRACT

CONSTRUCTION OF A NONSTANDARD ITEGRAL ON *AC* [0, 1]

by

Gerard Leonard Ornas, Jr., BS
Southwest Texas State University
August 1998

Supervising Professor: Stanley G. Wayment

A review of the Riemann and Lebesgue integration leads into the development and proof of the Riesz Representation Theorem. After this a new function space and a new norm are presented. A dense subset is extracted and used to approximate the functions in the new space. These approximations are used to define a nonstandard integral and an accompanying integral representation theorem.

Chapter One is a review of the basic definitions needed throughout the remainder of the thesis. For the less common definitions an example function is provided. A very fundamental theorem, that the space C [*0, 1*] is complete is worked out in detail. The Riemann integral is defined, both as the supremum of Riemann sums over partitions, and as the limit of Riemann sums as the norm of the partition goes to zero. Finally, the Fundamental Theorem of Calculus is given.

Chapter Two traces the development of the Lebesgue integral, following the path of H. L. Royden's *Real Analysis*[1]. The chapter begins by showing the deficiencies of the Riemann integral. The first step in constructing the Lebesgue integral to avoid these difficulties is to define measure for sets. Next the Lebesgue integral is defined for a wider and wider range of functions. Starting with simple functions and moving up to the general

[1] H. L. Royden, *Real Analysis, 3rd Edition,* Macmillan, New York, 1988

case of Lebesgue integrable functions. The theorem that a function is the anti-derivative of another function if and only if it is absolutely continuous is stated. Finally, the chapter closes with an example of a very peculiar function, which illustrates the distinction between integration and antidifferentiation.

Chapter Three also follows the path in Royden. More care is given however to show and fill in the proofs. The chapter is mostly a sequence of lemmas, definitions, and propositions, leading to the very important Riesz Representation Theorem. The chapter closes with the statement and a detailed proof, following that of Royden.

Chapter Four is the actual development of the nonstandard, variation integral, developed by Edwards and Wayment[2]. The chapter starts by setting up the space of absolutely continuous functions on [*0, 1*]. Next the norm is defined and proved to be a norm. Next the set of polygonal functions is shown to be dense in the space of absolutely continuous functions. Using this subset to approximate the absolutely continuous functions, the computable variation integral is defined. Simultaneously, an integral representation theorem for the new integral, representing the integrals as bounded linear functionals is developed. Finally, a couple of examples are given to illustrate the use of the integral.

[2] J. R. Edwards and S. G. Wayment, *Representations for Transformations Continuous in the BV Norm,* Transactions of the American Mathematical Society, **154**, 1971 (251 – 265)

CHAPTER I

INTRODUCTORY MATERIAL

One quality of a college education is that it forces us to develop the tools with which we analyze the world around us. A sophomore English major would cringe in embarrassment if he was forced to read a book report he wrote analyzing a book like *Moby Dick* in middle school. The ideas he uses to look at the world of literature change very rapidly. Yet a doctoral candidate in biology would look with pride at her dissertation, even though it was full of math based on the integration technique she learned as a senior in high school calculus.

Why is it that our methods of viewing things change so dramatically in the liberal arts, yet almost not at all in mathematics. The answer, to the latter half of the question anyway, is the awesome power, ubiquity, and most importantly computability of the Riemann integral. For a tremendous range of uses, including most used in engineering and physics, the Riemann integral is a perfectly valid tool. However, it is not the only way to define an integral operator on a function space. Nor is it a one size fits all tool. As we will see in Chapter 2, it does not take a very difficult function to gum up the works of Riemann integration.

We will spend the rest of Chapter 1 laying out definitions and some important results we will need throughout the rest of the thesis. In Chapter 2 we will define the first of our alternative integrals, the Lebesgue integral. In Chapter 3 we will use this integral in an important theorem which lets us represent functionals on a space of functions as integrals. In the Chapter 4 we will develop the second of our alternate integrals, the variation integral. We will also develop another representation theorem analogous to the Riesz Representation Theorem in Chapter 3.

To start with however, we will review some definitions we will be working with throughout the rest of the thesis. To begin the definitions we will list some of the various forms of continuity of functions.

Definition: *A function f is said to be continuous at a point x, if given any $\varepsilon > 0$, there is a $\delta > 0$ so that if $y \in (x-\delta, x+\delta)$,* then we have that $f(y) \in (f(x)-\varepsilon, f(x)+\varepsilon)$.

A function that is continuous at every point in a set is said to be continuous on that set. This is the classical definition of continuity.

Definition: *A function f is said to be uniformly continuous on an interval [a,b] if for any $\varepsilon > 0$ there is a $\delta > 0$, so that for any $x \in [\, a, b\,]$, if $y \in (x-\delta, x+\delta)$ then $|f(x)-f(y)| < \varepsilon$.*

This definition differs from the above definition of continuity in that in the first definition, δ was a function of both ε and x. In uniform continuity, δ is a function only of ε. In other words, one δ will work for the entire set once ε is given. Of course a function which is uniformly continuous is continuous.

An example of a function which is not uniformly continuous on the reals is $f(x) = x^2$. A simple series of algebraic calculations will show that in order for $|f(x) - f(y)| < \varepsilon$ given $\varepsilon > 0, \delta > 0, x > 0$ and $y \in (x, x + \delta)$, then it must be the case that $x < \frac{\varepsilon - \delta^2}{2\delta}$. Since $\delta > 0$ we have that $x < \frac{\varepsilon}{2\delta}$. Rearranging this gives us (since $x > 0$) that $\delta < \frac{\varepsilon}{2x}$. Thus δ must depend very much on our x. Hence x^2 is not uniformly continuous.

A well known result of classical analysis is that a function which is continuous on a closed interval is uniformly continuous.

Definition: *A function f is said to be absolutely continuous on an interval [a, b] if given $\varepsilon > 0$, there is a $\delta > 0$, so that, given any finite collection $\{(x_i, y_i)\}_{i=1}^{n}$ such that $\sum_{i=1}^{n} (y_i - x_i) < \delta$, then it is the case that $\sum_{i=1}^{n} |f(y_i) - f(x_i)| < \varepsilon$.*

Notice, a function which is absolutely continuous would clearly be uniformly continuous. We can choose any finite collection of intervals whose lengths sum to less than δ. If we choose the collection to consist of only one interval, a very finite collection indeed, then we have the very definition of uniformly continuous.

We will now show two functions which are uniformly continuous, but not absolutely continuous. The first will not be absolutely continuous on the entire non-negative real line, but will be on any closed and bounded interval of the non-negative reals. The second example, which we will see in Chapter 2, will not be uniformly continuous on the closed interval [0, 1].

In order that a function be uniformly continuous, but not absolutely continuous, it must grow very much in steep jumps in order that a collection of intervals of finite length cannot contain the growth. But, it cannot take very large jumps, although they must be steep , in order that the function be uniformly continuous.

The function we will look at is

$$f(x)=\begin{cases}\displaystyle\sum_{i=0}^{n}\frac{1}{i+1} & x\in[n,(n+1-\frac{1}{2^n}]\\ \displaystyle\sum_{i=0}^{n}\frac{1}{i+1}+\left(x-(n+1-2^{-n})\right)\left(\frac{2^n}{n+2}\right) & x\in[n+(1-\frac{1}{2^n}),n+1]\end{cases}$$

where n is an integer greater than or equal to zero.

While the above description looks very messy a geometric interpretation will shed much light on the function. The function can be pictured as a sequence of steps, where the height of the n^{th} step is $\frac{1}{n+1}$. The length of each step is $1-\frac{1}{2^n}$. The height of all the stairs is the sum of the harmonic series, which is infinity. The sum of the length of the intervals over which the function's slope is greater than zero is $\sum_{n=0}^{\infty}\frac{1}{2^n}=2$. The increment between successive steps is going to zero. For any ε there is an N so that the height of the jump of the n^{th} step for all $n>N$ is less than ε. This makes it easy to show that the function is uniformly continuous. But the horizontal distance of the intervals over which the steps grow is also going to zero, and is doing so quickly enough for the total sum of their lengths to be finite. Since the harmonic series diverges, we can get a set of partial sums to exceed any given number. We can do this with an arbitrarily narrow finite collection of intervals, since the length of the intervals over which the function is growing are shrinking as $\frac{1}{2^n}$. Hence the function is not absolutely continuous.

Notice that it was the fact that the function grew to infinity in the limit that caused our function to not be absolutely continuous. If we restrict our function to any bounded interval, then the function is absolutely continuous on this interval.

We will state and prove a few results about absolutely continuous functions. These are that sum and product of absolutely continuous functions are absolutely continuous.

Proposition: *If f and g are absolutely continuous functions then their sum* $(f+g)$ *is also absolutely continuous.*

Proof: Let ε > 0. Since f and g are both absolutely continuous there exist δ_f and δ_g, so that if we are given any finite collection of intervals $\{(x_i, y_i)\}_{i=1}^n$, with the condition that

$\sum_{i=1}^n (y_i - x_i) < \delta_f$, then $\sum_{i=1}^n |f(y_i) - f(x_i)| < \frac{\varepsilon}{2}$, and that if

$\sum_{i=1}^n (y_i - x_i) < \delta_g$, then $\sum_{i=1}^n |g(y_i) - g(x_i)| < \frac{\varepsilon}{2}$. We choose δ to be the minimum of δ_f and δ_g.

Now consider a collection $\{(x_i, y_i)\}_{i=1}^n$, with $\sum_{i=1}^n (y_i - x_i) < \delta$. Since δ is smaller than either δ_f or δ_g, we have that

$$\sum_{i=1}^{n}\left|\left((f(y_i)+g(y_i)\right)-\left((f(y_i)+g(y_i)\right)\right|=$$
$$=\sum_{i=1}^{n}\left|\left((f(y_i)-f(x_i)\right)+\left((g(y_i)+g(x_i)\right)\right|$$
$$\leq\sum_{i=1}^{n}\left(\left|\left((f(y_i)-f(x_i)\right)\right|+\left|\left((g(y_i)+g(x_i)\right)\right|\right)$$
$$=\sum_{i=1}^{n}\left|f(y_i)-f(x_i)\right|+\sum_{i=1}^{n}\left|g(y_i)-g(x_i)\right|$$
$$<\frac{\varepsilon}{2}+\frac{\varepsilon}{2}=\varepsilon.$$

Thus $(f+g)$ is absolutely continuous.

QED.

Proposition: *If f and g are absolutely continuous on* [*a*, *b*], *fg is also absolutely continuous on* [*a*, *b*].

Proof: Notice that the proposition is trivial if either function is the zero function. Since f and g are absolutely continuous they are bounded on [*a*, *b*]. We will let two numbers nonzero numbers F and G be numbers so that $|f(x)|<F, |g(x)|<G,$ for all $x\in[a,b]$.

Now since f is absolutely continuous given a finite collection $\{(x_i,y_i)\}_{i=1}^{n}$, of sub-intervals of [*a*, *b*], there is a $\delta_f>0$, so that if

$\sum_{i=1}^{n}(y_i-x_i)<\delta_f$, then $\sum_{i=1}^{n}|f(y_i)-f(x_i)|<\frac{\varepsilon}{2G}$. Since g is absolutely continuous,

there is also a $\delta_g>0$, so that given $\{(x_i,y_i)\}_{i=1}^{n}$, such that

$\sum_{i=1}^{n}(y_i-x_i)<\delta_g$, then $\sum_{i=1}^{n}|g(y_i)-g(x_i)|<\frac{\varepsilon}{2F}$. We now choose δ to be the minimum

of δ_f and δ_g. We choose next $\{(x_i,y_i)\}_{i=1}^{n}$, so that $\sum_{i=1}^{n}(y_i-x_i)<\delta$. Now consider:

$$\sum_{i=1}^{n}\left|(f(y_i)g(y_i)) - (f(x_i)g(x_i))\right| =$$
$$\sum_{i=1}^{n}\left|(f(y_i)g(y_i)) - (f(x_i)g(y_i)) + (f(x_i)g(y_i)) - (f(x_i)g(x_i))\right|$$
$$\leq \sum_{i=1}^{n}\left(\left|(f(y_i)g(y_i)) - (f(x_i)g(y_i))\right| + \left|(f(x_i)g(y_i)) - (f(x_i)g(x_i))\right|\right)$$
$$\leq \sum_{i=1}^{n}\left|(f(y_i) - f(x_i))g(y_i)\right| + \sum_{i=1}^{n}\left|f(x_i)(g(y_i) - g(x_i))\right|$$
$$\leq \sum_{i=1}^{n}\left|(f(y_i) - f(x_i))\right|G + \sum_{i=1}^{n}\left|(g(y_i)) - (g(x_i))\right|F$$
$$= G\sum_{i=1}^{n}\left|f(y_i) - f(x_i)\right| + F\sum_{i=1}^{n}\left|g(y_i) - g(x_i)\right|$$
$$< G\frac{\varepsilon}{2G} + F\frac{\varepsilon}{2F} = \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus *fg* is absolutely continuous.

QED.

Observation: *Constants functions are absolutely continuous.*

Proof: Let $f(x) = k$, for some number *k*. Let ε > 0, δ > 0. Consider $\{(x_i, y_i)\}_{i=1}^{n}$, so that

$\sum_{i=1}^{n}(y_i - x_i) < \delta.$

$$\sum_{i=1}^{n}\left|f(y_i) - f(x_i)\right| = \sum_{i=1}^{n}|k - k| = 0 < \varepsilon.$$

Thus f is absolutely continuous.

QED.

The next couple of definitions will deal with the notion of sequences of functions converging to a function.

Definition: *A sequence of functions, $\langle f_n \rangle$, defined on an interval [a, b] is said to converge (pointwise) to a function f (called the pointwise limit of the sequence) if, given any ε > 0, there is a natural number, N, so that if n > N, then $|f_n(x) - f(x)| < \varepsilon$.*

Notice that the N in the above definition is a function of both x and ε, in much the same way as δ was in the definition of continuous function. This analogy will be carried further with the next definition.

First, we will give an example of a sequence of functions which converge pointwise to the function $f(x)=0$. The sequence of functions is given by $f_n(x)=\frac{1}{n}$. That this sequence converges can be seen by observing that given ε > 0, there is an N, given by the Axiom of Archimedes, so that $\frac{1}{N}<\varepsilon$. This will be the N for the sequence.

Definition: *A sequence of functions, $\langle f_n \rangle$, defined on an interval [a, b], is said to converge uniformly to a function, f, if given $\varepsilon > 0$, there is a natural number N, so that if $n > N$, then $|f_n(x)-f(x)|<\varepsilon$.*

Notice that, as we foreshadowed in the previous definition, the difference in our two forms of convergence of sequences of functions is that in uniform convergence, the N is totally dependent on ε, and has no dependence whatsoever on x. An example of a function which is uniformly convergent is the example, $f_n(x)=\frac{1}{n}$, above.

We next give a sequence of functions which converges, but does not do so uniformly. The sequence of functions we describe will converge pointwise, as did the previous sequence to the zero function on the interval [0, 1]. Let the sequence $\langle g_n \rangle$ be defined as follows on the interval [0, 1]:

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

This is a sequence of Gaussian curves whose peak value of 1 occurs at $x = 1 - \frac{1}{n}$.

The n^n term in the exponential makes the curves become narrow as than they slide to the right. As a result, any point p less than 1 will be passed by the peak of the curve for some n and the function values at p will then very rapidly go to zero.

To show this consider that given an $x \in [0,1)$, there is an n so that $x < 1 - \frac{1}{n}$. Now consider the exponent in the sequence of real numbers, $\langle g_k(x) \rangle$ for all $k > n$. The k^k term grows very, very quickly. The $\left(1 - \frac{1}{k}\right) - x$ term also grows. Thus the exponent goes to -∞, and the sequence $\langle g_k(x) \rangle$ goes to zero.

As the curves become narrower, the function goes to zero at x=1, as well. To show this consider the sequence $\langle g_k(1) \rangle$. This sequence is equal to $e^{-\left(k^k\left[\left(1-\frac{1}{k}\right)-1\right]\right)^2}$. This simplifies to $e^{-\left(k^{k-1}\right)^2}$, which also converges to zero.

Thus, the sequence converges pointwise to the zero function. However, for any value of n, there is a point, namely $1 - \frac{1}{n}$, for which the function value is one. This means if we take $0 < \varepsilon < 1$ for any n there is an x for which $|g_n(x) - g(x)| > \varepsilon$. Hence the sequence of functions does not converge uniformly.

We will now define a property of functions which will be very useful to us in the future.

We will consider a function f defined on an interval $[a, b]$. Next define a set of points $\sigma = \{x_0 = a < x_1 < x_2 < \cdots < x_{n-1} < x_n = b\}$, which we will call a partition of $[a, b]$. We will define the positive variation of the function under the partition to be

$$p = \sum_{i=1}^{n} [f(x_i) - f(x_{i-1})]^+$$

where given a function g,

$$g^+(x) = \begin{cases} g(x) & g(x) > 0 \\ 0 & g(x) \le 0 \end{cases}.$$

We will define the negative variation to be

$$n = \sum_{i=1}^{n} [f(x_i) - f(x_{i-1})]^-,$$

where g^- is defined, analogously to g^+ by :

$$g^-(x) = \begin{cases} -g(x) & g(x) \le 0 \\ 0 & g(x) > 0 \end{cases}.$$

We define the total variation of f over our partition to be $t = p + n$.

We define the positive, negative, and total variations of f on the interval $[a, b]$:

$$\begin{aligned} P &= \sup p \\ N &= \sup n \\ T &= \sup t \end{aligned}$$

where the supremum is taken over all possible partitions of the interval.

A function, defined on $[a, b]$ is said to be of bounded variation if $T(f) < \infty$.

Notice that a function of bounded variation need not be continuous. For example, consider the function $g(x) = 0$, where $x \neq 0.5$ and $g(x) = 1$, when $x = 0.5$. It is not difficult to show that $T(g) = 2$ on $[0, 1]$.

Functions which are not of bounded variation take a little effort to construct, but not too much. Consider the function, defined on $[\ 0, \frac{1}{\pi}\ ]$

$$f(x)=\begin{cases}\sin\frac{1}{x} & x\neq 0\\ 0 & x=0\end{cases}.$$

If we look at a sequence of partitions each consisting of the first k numbers of the form $\left[\frac{1}{\left(n+\frac{1}{2}\right)\pi}\right]$, along with zero and $\frac{1}{\pi}$. These partitions will give an sequence of total variations which diverges. Hence this function is not of bounded variation. A more subtle example, based upon this idea would be the function

$$g(x)=\begin{cases}x\sin\frac{1}{x} & x\neq 0\\ 0 & x=0\end{cases}.$$

The analysis of this example is mostly the same as the previous one, except that the divergence of the total variations of the partitions is a result of the divergence of the harmonic series.

Now that we have defined bounded variation, we can state a theorem which will be very important later on, in Chapter 2.

Theorem: *A function f is of bounded variation on* $[\ a,\ b\ ]$ *if and only if f can be written as the difference of two monotone real-valued functions on* $[\ a,\ b\ ]$.[1]

Another observation we can make right now is:

[1] H.L. Royden, *Real Analysis, 3rd Edition,* MacMillan, New York, 1988, pg. 103

Observation(5.11): *An absolutely continuous function f on a closed interval is of bounded variation.*[2]

Proof: Let ε > 0. We have a δ > 0 for which *f* on any collection of sub-intervals of length less than δ will grow less than ε. If we have any partition of the interval, we can add points to this partition if necessary to get a collection of intervals so that we can group them into at most $1+\frac{(b-a)}{\delta}$ groups, so that each group has a total length of less than δ. The total change in these functions on any group is less than ε. Thus the total change in the function is less than $\varepsilon\left(1+\frac{(b-a)}{\delta}\right)$. Thus we have a number which bounds the total variation of the function over any partition. This number bounds the supremum of the variation over all partitions. So we have a number, which is a bound for the variation of the function and the function is of bounded variation.

QED.

We will now turn our attention to some definitions which we will need in our work in Chapters 3 and 4 concerning functional analysis.

Definition: *A set X is said to be a real linear vector space if:*

1. *X is a group under an operation* $\oplus$
2. *Given* $a \in \Re$ *(the field of real numbers) and* $x \in X$, $ax \in X$.
3. *Given* $a,b \in \Re$ *and* $x,y \in X$, *then both* $(a \oplus b)x = ax \oplus bx$, *and* $a(f \oplus g) = af \oplus gf$.

---

[2]Ibid., pg. 108

One obvious example of a linear space is to let X the real numbers. Another example is to let X be the set of continuous functions on the interval [$0, 1$], heretofore denoted $C[0, 1]$. To demonstrate this we observe that the sum of two continuous functions is itself continuous. Also if we take a continuous function and multiply it by a constant we again get a continuous function. Also the real numbers distribute over the space of continuous functions. Further, since we have closure under multiplication, $C[0, 1]$ also have another structure. They form an algebra.

Let $AC[0, 1]$ be the set of absolutely continuous functions on the unit interval. We recall that products of absolutely continuous functions on a closed interval are absolutely continuous. Further constant functions are absolutely continuous. Hence, scalar products of absolutely continuous functions are absolutely continuous. Given this and the fact that the sum of absolutely continuous functions are absolutely continuous, and we can observe that the absolutely continuous functions on the interval [$0, 1$], form a linear vector space, and also an algebra.

Definition: *A linear space X is said to be a normed space if there is a function, denoted $\|f\|$ which maps elements of X to the non-negative reals, subject to the following constraints: 1. $\|\alpha f\| = |\alpha|\|f\|$ for any constant α, and any $f \in X$, 2. $\|f\| = 0$ if and only if f = 0, and 3.. $\|f + g\| \leq \|f\| + \|g\|$,.*

This third criterion is called either the Minkowski inequality or the triangle inequality. It is one of the most commonly used tools in analysis.

One norm which comes up quite frequently when the linear space in question is $C[0, 1]$ is the sup norm. We define the sup norm as follows:

$$\|f\|_\infty = \sup_{0 \le x \le 1} |f(x)|$$

It is not very difficult to show that this does in fact describe a norm. The first property follows from the facts that (1.) The absolute value of a product is the product of an the absolute values and (2.) The supremum of a constant over any non-empty set is simply the constant itself. The second property follows from the fact that no number has an absolute value less than zero. Hence if the norm of a function is zero, then the greatest the absolute value of the function could be anywhere in the unit interval would be zero. Thus the function must be identically zero. Finally, the hardest part in showing that the sup norm is in fact a norm is the triangle inequality. This is not surprising as it will almost always be the hardest part in showing any purported norm to be an actual norm. Let us consider the norm of the sum of two functions, f and g. We see that

$$|f(x) + g(x)| \le |f(x)| + |g(x)|$$

By the classical triangle inequality. Further:

$$|f(x)| + |g(x)| \le |f(x)| + \sup_x |g(x)|$$
$$\le \sup_x |f(x)| + \sup_x |g(x)|$$

Hence,

$$\|f + g\|_\infty = \sup_x |f(x) + g(x)|$$
$$\le \sup_x |f(x)| + \sup_x |g(x)| = \|f\|_\infty + \|g\|_\infty.$$

Hence the sup norm is in fact a norm on *C[0, 1].*

The last quality of a linear normed space we will consider is whether or not the space is complete. A space is complete if every Cauchy sequence in the space converges to an element of the space. This necessitates our defining the term Cauchy.

Definition: *A sequence, $\langle f_n \rangle$, in a linear normed space, X, is said to be Cauchy if given an $\varepsilon > 0$, there is a natural number N, so that for any n, $m > N$, $\|f_n - f_m\| < \varepsilon$..*

The real numbers, using the standard euclidean metric form a normed linear space. An example of a Cauchy sequence in the reals is $\langle x_n \rangle = \left\langle \frac{1}{n} \right\rangle$. For any ε, we know that there is a positive integer N for which $\frac{1}{N} < \frac{\varepsilon}{2}$. If we choose n, $m > N$. Then

$$\|x_n - x_m\| = \left| \frac{1}{n} - \frac{1}{m} \right|$$

But since n and m are real numbers, as are their reciprocals, we can apply the triangle inequality to get

$$\left| \frac{1}{n} - \frac{1}{m} \right| \leq \left| \frac{1}{n} \right| + \left| \frac{1}{m} \right| < \left(\frac{\varepsilon}{2} + \frac{\varepsilon}{2} \right) = \varepsilon.$$

Thus we have that our sequence is Cauchy. In fact we can use a similar argument to show that any convergent sequence is Cauchy.

We can demonstrate a point here with an example which is not Cauchy. This will demonstrate that just because the distance between successive terms goes to zero, it is not necessarily the case that the sequence is Cauchy.

The sequence we will use is the partial sums of the harmonic series, that is

$$x_n = \sum_{i=1}^{n} \frac{1}{n}.$$

It is a very well known result that the sequence $\langle x_n \rangle$ diverges to infinity. Yet as n gets very large, the distance between consecutive terms goes to zero. We will now show that even though this is the case, the sequence is not Cauchy.

We will start our proof by letting ε be a number greater than zero. Next we will choose n to be an arbitrary number. The sum $\sum_{i=1}^{n}\frac{1}{n}$ is a number we will call N. Now consider the number $(N + \varepsilon)$. Since the sums of the partial sequences go to infinity, there must be an m so that $\sum_{i=1}^{m}\frac{1}{i} > (N+\varepsilon)$. For this m, we will consider

$$\left\|x_n - x_m\right\| = \sum_{i=1}^{m}\frac{1}{i} - \sum_{i=1}^{n}\frac{1}{i} \geq (N+\varepsilon) - N = \varepsilon.$$

Thus given an ε and *any* n whatsoever, there is an m greater than n so that the norm of the distance between the two points is greater than ε. Thus the sequence is not Cauchy. As the previous example demonstrated that any convergent sequence is Cauchy, this example demonstrates that any if $\|f_n\| \to \infty$, the sequence is not Cauchy. In fact, any sequence which diverges to ±∞ cannot be Cauchy.

The only question this leaves regarding when a sequence is Cauchy is "Is every Cauchy sequence convergent?" The answer to this is in general no. For instance, the Cauchy sequence of rational numbers $\left\langle \sum_{i=0}^{n}\frac{1}{n!} \right\rangle$, converges to the number e, as is known from calculus. However, e is not a rational number. So we have a set of numbers, namely the rationals, in which not every Cauchy sequence converges in the set. There are sets in which every Cauchy sequence does converge. It is an axiom of the real numbers that every Cauchy sequence of real numbers converge to a real number. Similarly every Cauchy sequence of complex numbers converge to a complex number. This property of both the reals and complex numbers is called completeness.

Definition: *A normed linear space , X, is complete if every Cauchy sequence in X converges in X.*

We will now show another example, namely, that the set C [*0, 1*], the continuous functions on the interval [*0, 1*] under the sup norm is complete.

Theorem: *Every Cauchy sequence of continuous functions on* [*0, 1*] *under the sup norm converges to a continuous function.*

Proof: We will prove this theorem in three steps. First we will define what we want our limit function to be. Second we will show that this candidate function is in fact continuous. Finally we will show that this function is the limit of the sequence under the norm.

In Chapter 4 we will be proving a similar theorem with a different set of functions and a different norm, but the proof will follow this same set of steps.

We will start with a Cauchy sequence of functions and observe that given any $\varepsilon > 0$ and any x in [*0, 1*] , there is an $N > 0$ so that if n and $m > 0$, then $|f_n(x) - f_m(x)| < \varepsilon$. Thus for each point x the sequence $\langle f_n(x) \rangle$ is a Cauchy sequence of real numbers. As we stated above, the real numbers are complete, so for each x, the sequence $\langle f_n(x) \rangle$ has a limit. We now define our candidate for our limit function, $f(x) = \lim_{n \to \infty} f_n(x)$.

The next step is to show that this function is continuous at x. That is, given $\varepsilon > 0$ and $x \in [0,1]$, we must find a $\delta > 0$, so that if $y \in (x - \delta, x + \delta)$, then $|f(x) - f(y)| < \varepsilon$.

Since $\langle f_n(x) \rangle$ converges to $f(x)$ there is an N' so that if $n > N'$, then $|f_n(x) - f(x)| < \frac{\varepsilon}{4}$. Since $\langle f_n \rangle$ is Cauchy there is an N'' so that if $n, m > N''$ then

$|f_n(x) - f_m(x)| < \frac{\varepsilon}{4}$ for all $x \in [0,1]$. We now fix N=N'+N''. Since f_N is continuous,

there is a δ so that if $|x - y| < \delta$, then $|f_N(x) - f_N(y)| < \frac{\varepsilon}{4}$. This is the δ we will use for

the definition of continuity. We next let $y \in (x - \delta, x + \delta)$. As we observed above

$\langle f_n(y) \rangle$ converges, so there is an M so that if $k > M$, then $|f_k(y) - f(y)| < \frac{\varepsilon}{4}$. Let

m=M+N. Now consider:

$$
\begin{aligned}
&|f(x) - f(y)| = \\
&= |f(x) - f_N(x) + f_N(x) - f_N(y) + f_N(y) - f_m(y) + f_m(y) - f(y)| \\
&\le |f(x) - f_N(x)| + |f_N(x) - f_N(y)| + |f_N(y) - f_m(y)| + |f_m(y) - f(y)| \\
&< \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \varepsilon.
\end{aligned}
$$

Thus f is continuous.

Finally, we must show that given ε > 0, there is an N so that for any $n > N$,

$|f(x) - f_n(x)| < \varepsilon$, for any $x \in [0,1]$. Choose N as follows:

Since $\langle f_n \rangle$ is Cauchy, there is an N so that $|f_n(x) - f_m(x)| < \frac{\varepsilon}{2}$, for all $m,n > N$

and all $x \in [0,1]$. Then for $n,m > N$ and $x \in [0,1]$ we have

$$|f_n(x) - f(x)| \le |f_n(x) - f_m(x)| + |f_m(x) - f(x)|$$

for any choice of m. Notice that the left hand side of the inequality is independent of m.

However, the first term in the right hand side is less than $\frac{\varepsilon}{2}$ when $m > N$. Since the

sequence of real numbers $\langle f_n(x) \rangle$ converges to $f(x)$, there is an M, which depends on x, so that if $m > M$ then $|f_n(x) - f(x)| < \frac{\varepsilon}{2}$. If $m > N+M$ both terms on the right are less than $\frac{\varepsilon}{2}$. Thus if n is greater than N and $x \in [0,1]$ we have $|f_n(x) - f(x)| < \varepsilon$.

We now let y be in the unit interval, and $n > N$, and $m = N + M$. Now consider

$$\begin{aligned} |f_N(y) - f(y)| &\le |f_N(y) - f_m(y)| + |f_m(y) - f(y)| \\ &\le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

Thus our arbitrary Cauchy sequence of continuous functions converges in the sup norm to a continuous function.

QED.

We will spend most of the rest of this chapter defining the Riemann integral. This is the integral for which we will spend most of the remainder of the thesis finding an alternative. This will mostly be a statement of the definition and some theorems and results. For a basic development, many freshman calculus texts will suffice. For a careful and rigorous development of the Riemann integral a very good source is Rudin's *Principles of Classical Analysis.*[3]

Given an interval, [a, b], we have defined a partition on this interval to be a set $\sigma = \{a = x_0 < x_1 < x_x < \cdots < x_{n-1} < x_n = b\}$. Given a function f defined on this interval, we define the upper Riemann sum of this function on this partition to be

$$U(P,f) = \sum_{i=1}^{n} (\sup f(x))(x_i - x_{i-1}) .$$

[3] W. Rudin, *Principles of Mathematical Analysis, 3rd Edition,* McGraw-Hill, New York, 1976

Geometrically, we are finding the area of a rectangle, whose base is the interval between two points in our partition. The height of this rectangle is the supremum of the function on the interval in our partition. We are then summing the area of all the rectangles defined by our partition. This sum is an overestimate of the area under the curve f on the interval [a, b].

We also define the lower Riemann sum, $L(P, f)$, of the function similarly, except that the height of the rectangle will be the infimum of the function on the interval.

We will now define the upper Riemann integral, $U\int_a^b f(x)dx = \inf U(P,f)$, where the infimum is taken over all partitions of [a, b]. We will similarly define the lower Riemann integral to be $L\int_a^b f(x)dx = \sup L(P,f)$. Again, the supremum is taken over all partitions of [a, b]. If the upper and lower Riemann integrals are equal to one another, we say the Riemann integral of f on the interval [a, b] exists and is equal to the lower (and of course the upper) Riemann integral.

We can equivalently define the Riemann integral by taking the right hand value (or left hand, or center) of the function on each interval. As we let the norm of the partition, that is the length of the longest interval in our partition, go to zero we get the same integral as above.

While the first definition has theoretical and pedagogical advantages in proofs, the second definition has a very large advantage of its own. We can take any partition and find the right hand endpoint and evaluate the function there. Since we know that this is very close to the integral, for a fine enough partition, evaluating the Riemann integral, within any chosen accuracy becomes an arithmetic process. That is, we can actually compute the Riemann integral, even when the Fundamental Theorem of Calculus does not apply. This is not a property held by the Lebesgue integral we will look at in Chapter 3. It is an important property of the integral we will develop in Chapter 4.

The two single most important developments of mathematics in the seventeenth century (and arguably since) are the Riemann integral (although it was not called such, or defined in the above fashion at the time) and the following theorem.

Theorem: (The Fundamental Theorem of Calculus) *If f is Riemann integrable on the interval [a, b] and has an anti-derivative F, then*

$$\int_a^b f(x)dx = F(b) - F(a).$$

If we define the indefinite integral of f(x) to be $F(x) = \int_a^x f(t)dt$, *then at points where f is continuous,* $F'(x) = f(x)$.

This theorem tells us that integration and differentiation are inverse procedures and also tells us how to evaluate a Riemann integral. If a Riemann integrable function has an anti-derivative, then the integral over the interval [a, b] is just the anti-derivative of the function evaluated at the endpoints a and b. One of the major questions this leaves unanswered is "How do we know whether a function has an anti-derivative or not?" This is one of the major questions we will answer in the next chapter. Another question we will face is "Can a function have an anti-derivative and yet fail to be integrable?"

CHAPTER II

THE LEBESGUE INTEGRAL, AN OVERVIEW

In the previous chapter, we made several definitions and defined Riemann Integration. In this chapter, we will point out the shortcomings of the Riemann integral and review the development, as presented in Royden's *Real Analysis*[1] of the first of our two alternate integrals, the Lebesgue integral. We will find its basic properties, and state several results concerning it. We will finish with the discussion of some interesting, but pathological functions.

In Riemann integration, the "area under the curve" is approximated by slicing the graph along the x-axis and approximating the area of these slices with rectangles whose height is some value the function takes on in the slice. The integral is then defined to be the limit (if it exists) of the approximations as the width of the widest slice goes to zero. Using this definition is very useful and leads to the fundamental theorem of calculus, which gives us a way of actually computing integrals.

At this point we might ask, "Why did we slice up the x-axis? Couldn't we have sliced up the y-axis just as easily?" The answer that comes to mind first is that because a function can only have one value for a given x, we know that our slices will be slices. If we were to cut up the y-axis we might get many different pieces, which would be harder

---

[1] Royden, *Real Analysis, 3rd Edition,* Macmillan, New York, 1988

to add up. On the other hand, we are taking the same area, and just cutting it differently. A large pizza is still large whether we cut it into wedges, or strips. Both these points are valid. The integral we get by cutting the y-axis will not in general be easily computable, except by the fact that it will turn out to be the same as the standard Riemann integral, wherever the Riemann integral exists. There will however be some advantage to developing this integral, in that we can integrate more functions with this integral than we could with the Riemann integral. Further, we will be able to tighten up the constraints on the Fundamental Theorem of Calculus during the development of the new Lebesgue integral.

To start with, we will show that there are bound functions on finite domains which are not Riemann integrable. While as we saw in the first chapter, it can take a very strange function indeed to get around some of the classifications we, it will not take much to describe our pathological case here. Define a function f (x), the Dirichlet or so-called "salt and pepper" function on the interval [0,1] as follows:

$$f(x)=\begin{cases} 0 & x \in \text{irrationals} \\ 1 & x \in \text{rationals} \end{cases}$$

Notice that, since both the rational and irrational numbers are dense in [0,1], in any slice we make in the domain of f, there will be an x where $f(x)=1$ and a y where $f(y)=0$. Thus the upper Riemann sum will always be 1 and the lower Riemann sum will always be 0. Thus the integral, which is the limit of the Riemann sum cannot exist.

If we could find a way to measure the "size" of the rationals and the irrationals, then integrating this function would be trivial. We would multiply 0 by the "size" of the irrationals, multiply 1 by the "size" of the rationals, and add the two products. So the issue with our new integral is, what do we mean by the size of a set?

This evaluation of the size is what we will call the "measure" of the set. Before we can define the measure of a set however, we must first make a pair of definitions. First we define an algebra of sets to be a collection of sets, S, so that if the two sets, A and B are both in S, then their union, intersection, and complements are also in S. By induction, any finite union of sets from S is also in S. We will also define a special class of algebras, the σ-algebra. An algebra of sets $\mathcal{B}$ is a σ-algebra if every countable union of sets in $\mathcal{B}$ is also in $\mathcal{B}$. All the other properties of algebras hold, of course, because if $\mathcal{B}$ is a σ-algebra it is by definition, an algebra.

With these definitions out of the way, we can now proceed to define a measure. A measure is defined as a function with the following properties: First it maps a σ-algebra of sets to the non-negative extended real numbers (i.e. the real numbers along with $\pm\infty$). It also has a property called countable additivity, that is, the measure of a countable union of sets is equal to the sum of the measures of the individual sets in the unions when the sets are disjoint. It would also be nice if the σ-algebra the function was defined on was as large as possible.

The measure we will use is the Lebesgue measure where the measure of an interval was the length. The Lebesgue measure is based on the outer measure of a set. The outer measure is defined for all subsets of the real numbers. We denote the outer measure of a set A, by m*(A). Consider a collection of intervals $\mathrm{A}' = \{(x_i, y_i)\}$ which cover A. If we denote the sum of the lengths of the intervals in A' by α, m*(A) is the infimum over all collections of intervals covering A of α. Since any subset of the reals can be covered by the real line, we know that the set of all collections of intervals covering A is non-empty. So the set of the α's is a non-empty subset of the reals, bounded below by zero, hence it has an infimum. This infimum is the outer measure. We call a set E measurable if given any test set A the following equation holds:

$$m^*A = m^*(A \cap E) + m^*(A \cap \widetilde{E})$$

Where $\widetilde{E}$ is the complement of E. By invoking the Axiom of Choice we can construct a set which does not meet this criterion, but without bringing in this big gun, most any set encountered will be measurable. If a set is measurable, we define its Lebesgue measure to be its outer measure.

A couple of classical results are that the (Lebesgue) measurable sets form a σ-algebra and that open intervals are in this σ-algebra. We call the smallest σ-algebra containing the open sets the Borel sets. Thus the Lebesgue measure is defined at least on the Borel sets. The second result is that the measure of an interval is its length. This result follows almost directly from the definition of outer measure.

We can use the definition of Lebesgue measure to find the measure of the rational numbers. Each rational number x can be fit into an interval $(x - \frac{1}{n}, x + \frac{1}{n})$, for any n. So the radius of the interval containing x is $2/n$. The infimum of the set of all such numbers is zero. Thus the measure of a point is zero. From here, we observe that there are countably many rationals. Thus

$$mQ = \sum_{i=1}^{\infty} m(q_i) \leq \sum_{i=1}^{\infty} 0 = 0.$$

thus, $mQ = 0$, where Q denotes the rationals.

This also proves that any countable set is of measure zero.

The measure of the interval $[0,1]$ is one. The irrationals are the complement of the rationals, and are thus disjoint from the rationals. The sum of the measure of the union of the rationals and irrationals is the measure of the interval. Hence the measure of the irrationals is one. This gives us what we need to integrate the Dirichlet function we started with. The measure of the irrationals is one and the function evaluates to zero on the irrationals, so the irrationals contribute a value of zero to the integral. The function evaluates to one on the rationals, but since the measure of the rationals is zero, they also contribute zero to the integral of the function. While we can evaluate this integral quite nicely using the Lebesgue integral (which we still have not formally developed yet), this will be the exception rather than the rule.

We have defined the notion of a measurable set earlier, now we are going to define what is meant by a measurable function, which we will need in defining our integral. A function f is measurable if given for each real α, $\{x: f(x) \le \alpha\}$ is measurable.

We have an algebra of measurable functions. It can be shown that linear combinations, and products of measurable functions are measurable. Further, the suprema, infima, limits, limits superior, and limits inferior of sequences of measurable functions are all measurable. This gives us a very wide latitude of functions with which to deal.

One more thing remains to be done before we can actually define the Lebesgue integral. We will state a pair of theorems which will allow us to approximate measurable functions with more well behaved functions and measurable sets with collections of intervals.

Theorem: *Let f be a measurable function defined on an interval* [a, b], *and assume that f takes on the values $\pm\infty$ only on a set of measure zero. The given ε>0, we can find a step function g and a continuous function h such that*

$$|f(x)-g(x)|<\varepsilon \text{ and } |f(x)-h(x)|<\varepsilon$$

except on a set of measure less than ε. Further, if $\exists$ m, M so that m $\leq$f(x) $\leq$M, for all x, then we may choose the functions g and h so that m $\leq$g(x) $\leq$M and m $\leq$h(x) $\leq$M.[2]

We will now, at long last, actually define the Lebesgue, integral and state and prove some of the theorems along the way. We will define the integral in stages, expanding the range of functions with which we work.

The first step we will take will be in defining an integral of a simple function. A simple function is a function which takes on only finitely many values over its domain. The "salt-and-pepper" function we defined earlier is an example of a simple function. It takes on only two values. Another example would be the "greatest integer" function on the interval [17, 42]. Since a simple function, $\varphi(x)$, takes on only finitely many values we can label these values a_i, where i ranges from 0 to some natural number n. We will

[2] Royden, 69 -- 70

define the sets E_i, to be the set of all x's where $\varphi(x) = a_i$. We will also define the characteristic function, χ_{E_i}, to be:

$$\chi_{E_i}(x) = \begin{cases} 1 & x \in E_i \\ 0 & x \notin E_i \end{cases}$$

We can thus write our function, φ, as the sum of each a_i times the characteristic function of E_i. We can do this because any given point will only fall into one E_i. Hence, for any given x at most one term in our sum will be non-zero as we sum over all i, i.e. $\varphi(x) = \sum_i a_i \chi_{E_i}$. This gives us the "canonical" representation of φ,

With our function thus defined we now define the integral to be exactly what would be intuitive: namely the sum of the values of the simple function times the measure of the set having that value. In rough geometric terms we are taking the height of the function and the "width" of that portion of the domain taking on that value. We must of course realize that the set taking on this value may be broken into infinitely many pieces, each of which may be "infinitesimally narrow." This is of course why we defined the Lebesgue measure in the first place. Symbolically, for a given simple function over a domain E:

$$\int_E \varphi(x)dx = \sum_i a_i mE_i \quad .$$

One very important, but not very difficult to demonstrate property of our integral is that it is a linear operator, that is:

$$a\int f + b\int g = \int (af + bg).$$

We can now use this definition to define an integral on a much wider class of functions. If we have a bounded function we can define functions $\varphi(x), \psi(x)$, which are both simple functions with :

$$\varphi(x) \leq f(x),$$
$$\psi(x) \geq f(x),$$

for all x in E. Clearly, for all such functions,

$$\int_E \varphi \leq \int_E \psi$$

since φ is always less than f, which is in turn always less than ψ. Given this we can state:

$$\sup_{\varphi \leq f} \int_E \varphi \quad \leq \quad \inf_{\psi \geq f} \int_E \psi .$$

It can be shown that if f is a bounded measurable function on an interval, then the above numbers will be equal. Hence, we can define the Lebesgue integral for a bounded measurable function:

$$\int_E f \quad = \quad \sup_{\varphi < f} \int_E \varphi.$$

Some results follow almost immediately from this definitions. The first of these is that the integral is still a linear operator on its expanded domain. Another is the Bounded Convergence Theorem. This is the first in a sequence of convergence theorems which give us different criteria under which the limit of a sequence of integrals is the integral of the limit of the integrands.

Theorem: *Let $\langle f_n \rangle$ be a sequence of measurable functions defined on a set E of finite measure, and suppose that there is a real number M such that $|f_n(x)| \leq M$ for all n and all x. If $f(x) = \lim f_n(x)$ for each x in E, then*

$$\int_E f \quad = \quad \lim \int_E f_n.$$ [3]

One result concerning Riemann integrals is that if a bounded function has a Riemann integral, it will agree with the Lebesgue integral. It is possible however, if a function has an improper Riemann integral, that the function may not have a Lebesgue integral. Another result of note, which can be shown at this point is that the Riemann integral of a function exists if and only if the set of discontinuities of the function is of measure zero. Since the "salt and pepper" function above is discontinuous at every point in the unit interval, its Riemann integral does not exist based on this result. (Although we had already shown this from the definition of Riemann integral directly.)

It should be noted, mostly as a reminder, that while we can evaluate the Riemann integral of a function with possibly an uncountable number of discontinuities (as we shall see below), if even one discontinuity exists, we cannot use the fundamental theorem to evaluate the integral.

We will now demonstrate two examples of the above result. In the first example we will show a function with uncountably many discontinuities which has a Riemann integral. In the second, we will demonstrate a function with "only" countably many

---

[3] Ibid. 84

discontinuities, but discontinuities which are dense in the domain, which also has a Riemann integral.

First we will consider the following function defined on the interval [0 , 1].

$$g(x)=\begin{cases} 1 & x \in Cantor\ set \\ 2 & x \notin Cantor\ set \end{cases}$$

Since we will consider first the points in the Cantor set, let x be a point in the Cantor set. Let us also take two numbers $1 > \varepsilon > 0, \delta > 0$. Since the Cantor set is totally disconnected, there is a $y \in (x-\delta, x+\delta), y$ not in the Cantor set. Since y is not in the Cantor set:

$$|g(y) - g(x)| = |2 - 1| = 1 > \varepsilon.$$

Hence, g is not continuous on the Cantor set. To show that g is continuous on the complement of the Cantor function, we will let x be a point in the unit interval and not in the Cantor set.

We will next need to consider how the Cantor set is constructed in the first place. We start with the unit interval. The first step in the construction is to define the intervals which define the complement of the Cantor set. What remains after all the middle thirds are removed from the unit interval is the Cantor set.

We will do this in steps. The first set of intervals C_1 in the complement of the Cantor consists only of the open middle third or the unit interval, $(1/3, 2/3)$. The next set of intervals are the middle thirds of the intervals remaining when we remove the first interval. These are $(1/9, 2/9)$ and $(7/9, 8/9)$. Notice that the intervals $(3/9, 4/9)$ and $(5/9, 6/9)$ are sub intervals of $(\frac{1}{3},\frac{2}{3})$. This suggests a pattern we will use to define the complement of the Cantor set:

$$C_k = \left\{ \left(\frac{2n-1}{3^n}, \frac{2n}{3^n} \right) \middle| n \text{ is a natural number such that: } 1 \le 2n-1 < 3^k \right\}$$

The Cantor set is now defined to be:

$$C = [0,1] - \left(\bigcup_{k=1}^{\infty} C_k \right).$$

We note that at each step of the construction, the only sets thrown out of the unit interval are the middle thirds of the intervals left over from the previous step, since all other intervals in C_i we subintervals of C_k for some $k<i$. We will consider an equivalent definition a little later. Thus $(C_k - C_{k-1})$ is a collection of 2^{k-1} intervals of length $\frac{1}{3^k}$.

Returning to our function, we have that the point x is in an open interval in the complement of the Cantor set. Hence, by definition of open interval there is a $\delta > 0$ so that the interval $(x-\delta, x+\delta)$, is completely in the complement of the Cantor set. If we let $\varepsilon > 0$, then for any point, y, in our interval,

$$|g(x) - g(y)| = |2-2| = 0 < \varepsilon.$$

Thus our function is continuous on the complement of the Cantor set.

If we now apply additivity of the Lebesgue measure to the complement of the Cantor set, we find that its measure is:

$$\frac{1}{3} + \frac{2}{9} + \cdots + \frac{2^{n-1}}{3^n} + \cdots = \frac{1}{3}\sum_{n=0}^{\infty}\left(\frac{2}{3}\right)^n = \frac{1}{3}\left(\frac{1}{1-\frac{2}{3}}\right) = 1.$$

Thus the measure of the complement of the Cantor set is one. The measure of the Cantor set and its complement must add up to the measure of the unit interval. The measure of the unit interval is one, thus the measure of the Cantor set must be zero. Hence, the function has discontinuities of measure zero. Thus this function has a Riemann integral.

$$\int_0^1 g(x)dx = 1 \cdot m(\text{Cantor Set}) + 2 \cdot m(\text{Complement}) =$$
$$1 \cdot 0 + 2 \cdot 1 = 2$$

One interesting observation can be made at this point. While in the above example the function was constant on the Cantor set, we could have done anything we wanted with the values of the function on the Cantor set and it would all have occurred on a set of measure zero. As a result all the possible acrobatics we could have put the function through would not have affected either the Riemann integrability of g, nor the integral of the function.

We will now take a very interesting detour. Notice above, that the measure of the Cantor set was zero because we deleted the middle third of the unit interval. When we added the lengths (i.e. measures) of the intervals in the complement of the Cantor set, we got a geometric series which summed to one. If instead of removing the middle third we removed a smaller interval, the series can sum to a number smaller than one. For instance we will remove inductively at the k^{th} step an interval of length $\frac{1}{4^k}$ from the center of the intervals left over from the $(k\text{-}1)^{\text{st}}$ step, starting with a unit interval, in the same way we did with the standard Cantor set. We will thus remove 2^{k-1} intervals of length $\frac{1}{4^k}$, at the k^th^ step. Thus the measure of our generalized or "fat" Cantor set is:

$$1-\left(\frac{1}{4}+\frac{2}{16}+\frac{4}{64}+\cdots\right)=1-\left(\frac{1}{4}+\frac{1}{8}+\frac{1}{16}\right)=$$
$$=1-\frac{1}{2}\sum_{i=1}^{\infty}\left(\frac{1}{2}\right)^{n}=1-\frac{1}{2}(1)=\frac{1}{2}.$$

This would give us a set with most of the properties of the Cantor set, but it would have measure of one half. By using smaller intervals as our base (a sixteenth instead of a third or fourth, for instance), we can make our fat Cantor set have measure arbitrarily close to one.

Notice that our function *g* above, if we re-define it on our fat Cantor set has a set of discontinuities of measure one half. This means that the Riemann integral of this function does not exist. It is however a trivial matter for the Lebesgue integral. The value of the function on the fat Cantor set is one, and the measure is one half. The value of the function on the complement of the fat Cantor set is two and its measure is one half. Thus the Lebesgue integral of the function on the unit interval is three halves.

Our next example will be the "smog" function. The name of the function is an analogy to the way smog just seems to settle over a city, becoming more dense as the altitude becomes lower. This function, which we will refer to as *h*, in keeping with the alphabetic progression we have been following is defined as follows. First, recall that any rational number can be written as a fraction in which the numerator and denominator have no common factors. We will refer to this unique reduced representation of the number and we will call the denominator n_x.

$$h(x) = \begin{cases} \dfrac{1}{n_x} & x \in rationals \\ 0 & x \in irrationals \end{cases}$$

It is straightforward to show that h is discontinuous on the rationals. Given x let $\varepsilon = \frac{1}{2}\left(\frac{1}{n_x}\right)$, $\delta > 0$. Since the irrationals are dense in the reals, there is an irrational number, y, in the interval $(x - \delta, x + \delta)$. Thus:

$$|h(x) - h(y)| = \left|\frac{1}{n_x} - 0\right| = \frac{1}{n_x} > \frac{1}{2n_x}.$$

Thus h is discontinuous on the rationals, which is not necessarily bad, as we already know the rationals are a set of measure zero. It is a little more subtle to see that the function is continuous on the irrational numbers however. Let x be an irrational number and let $\varepsilon > 0$. There is a number n so that $\frac{1}{n} < \varepsilon$. So, for any rational number, y, whose denominator, in lowest terms is greater than n, we have the following:

$$|h(x) - h(y)| = \left|\frac{1}{n_x} - 0\right| = \frac{1}{n_x} < \varepsilon.$$

So we need only worry about those rationals whose denominator is less than n. Fortunately there are only finitely many of these. As a result we can measure there distances from x. We then take δ to be half the least of these numbers. Thus, for any y within δ of x $h(y)$ will be within ε of $h(x)$. By definition, h is continuous on the irrationals. So the Riemann integral of h exists, and it is straightforward to show that the integral of this function is in fact zero.

Now we will continue expanding the functions we can integrate with the Lebesgue integral. We have already defined our integral for simple functions and for bounded measurable functions on sets of finite measure. We will now use our previous definition to expand our range of functions to unbounded functions on any measurable set, finite measure or not. We must, for now at least, give up something, however. We must restrict ourselves for the moment to non-negative functions. We will define the integral of a non-negative function, f, defined on a measurable set E:

$$\int_E f \;=\; \sup_{h \le f} \int_E h.$$

The supremum in the above integral is taken over all bounded functions h, with $f(x) \ge h(x)$, for all $x \in E$ and $h(x) = 0$ outside of a bounded interval.

We will also define the terminology, integrable. A non-negative measurable function is said to be integrable over the measurable set E if the integral over E is finite. As has been the case before, we can show that our integral is still a linear operator. We also get the next two of our integral convergence theorems, Fatou's Lemma and the Monotone Convergence Theorem.

Theorem (Fatou's Lemma): *If $\langle f_n \rangle$ is a sequence of non-negative measurable functions and $f_n(x) \to f(x)$ almost everywhere on a set E, then*

$$\int_E f \;\le\; \underline{\lim} \int_E f_n.$$ [4]

---

[4] Ibid. 86

Theorem (Monotone Convergence Theorem): *Let $\langle f_n \rangle$ be an increasing sequence of non-negative measurable functions, and let $f = \lim f_n$ almost everywhere. Then,*

$$\int_E f \quad = \quad \lim \int_E f_n.$$ [5]

These theorems tell us that over a great range of functions we can pass a limit "through" an integral sign. This is a very useful thing to be able to do.

The phrase "almost everywhere" or a.e. which appears above means that the set where the criteria listed fail to hold is a set of measure zero. For instance, the smog function is zero a.e. This idea is a very important difference between Lebesgue and Riemann integration. In the former we in almost every case can simply skip over or ignore sets of measure zero, even if the sets in question are not only infinite, but in the case of the Cantor set, uncountably infinite. We saw with the salt and pepper function however that this is not the case with Riemann integration.

We are now at the stage of finally getting our full integral over all measurable functions. To do this we will note that a function can be either positive or negative, but not both at the same time. Using this we will define the functions f^+ and f^- as we did in Chapter 1:

$$f^+(x) = \begin{cases} f(x) & f(x) \geq 0 \\ 0 & f(x) < 0 \end{cases}$$
$$f^-(x) = \begin{cases} -f(x), & f(x) \leq 0 \\ 0 & f(x) > 0 \end{cases}$$

---

[5] Ibid. 87

It is not difficult to see that from this definition $f(x) = f^+(x) - f^-(x)$. This definition also has the advantage that it turns a general measurable function into a linear combination of two measurable functions we can integrate with what we have. We now define the integral of a measurable function *f* over a measurable set *E*:

Definition: *A measurable function is said to be integrable over a measurable set E if both f^+ and f^- are integrable over E. In this case we define*

$$\int_E f \;=\; \int_E f^+ - \int_E f^-.$$

We now have an integral which will integrate any measurable function, except for those where both the positive and negative parts (f^+ and f^-, respectively) have infinite integrals. Our integral fails in this case because we are left with a difference of the form $\infty - \infty$. We cannot work with differences like this.

We will now consider the last of our integral convergence theorems, the Lebesgue Convergence Theorem. In this theorem we require simply that our sequence of functions be bounded by some integrable function. We also require that the sequence of functions be convergent to the measurable limit function almost everywhere. This is a much less restrictive premise than for our other convergence theorems. Of course, this is due in part to the Lebesgue Theorem using the previous work in its proof.

Theorem (Lebesgue Convergence Theorem): *Let g be integrable over E and let $\langle f_n \rangle$ be a sequence of measurable functions such that $|f_n(x)| \le g(x)$ on E and for almost all x in E we have $f(x) = \lim f_n(x)$. Then*

$$\int_E f \quad = \quad \lim \int_E f_n.$$ [6]

We will conclude this chapter by considering the relationship between the integral we have just developed and the derivative. This will conclude with a theorem similar to, but more precise than the fundamental theorem of calculus.

Before we can look too closely at the relationship between integration and differentiation, we must of course define the derivative. Unlike the integral, the derivative we define will be precisely the same as the standard derivative from first semester calculus. We will however construct "derivates" which are analogous to the right and left hand derivatives we learned as freshmen. In fact, for those functions which had right and left hand derivatives, the derivates will be the right and left derivatives. With this in mind, we now define:

$$D^+ f(x) = \overline{\lim_{h \to 0^+}} \frac{f(x+h) - f(x)}{h}$$
$$D^- f(x) = \overline{\lim_{h \to 0^-}} \frac{f(x+h) - f(x)}{h}$$
$$D_+ f(x) = \underline{\lim_{h \to 0^+}} \frac{f(x+h) - f(x)}{h}$$
$$D_- f(x) = \underline{\lim_{h \to 0^-}} \frac{f(x+h) - f(x)}{h}$$

The first and the second derivates are called the upper right and upper left derivates, respectively. The third and the fourth are called the lower right and lower left derivates, respectively. Notice that the derivates are defined as limits superior and inferior of non-empty sets of numbers. Thus, they always exist. If the upper and lower right derivates are equal, they are equal to the right hand derivative. Similarly, the upper

[6] Ibid. 91

and lower left hand derivate give the left hand derivative if they are equal to one another. If all four derivates are equal, then they all are equal to the derivative of the function, denoted f'.

One result from calculus we can loosen up slightly is that if a function is continuous and any one of its derivates is everywhere non-negative, then f is non-decreasing on the interval.

We can now state a pair of results which are very useful, both in the path our development will take, and in their own place. The first result is that an increasing function has a derivative almost everywhere. This is an incredible piece of information to have based on nothing more than that a function never decreases.

An example of this is the "$\frac{1}{q^3}$" function, defined on the unit interval. We will define this function to be

$$f(x) = \sum_{\substack{\text{allrational} \\ q<x}} \frac{1}{n_q^{\ 3}}$$

Where n_q is the reduced denominator of q, as in the smog function. We observe that the function will be a jump at each rational number. The jump will be the reciprocal of the denominator of the rational cubed. Thus the function is not continuous. There are at most n_q-1 rationals with a denominator of n_q, thus the value of the function is bounded by:

$$f(x) = \sum_{\substack{\text{allrational} \\ q<x}} \frac{n_q}{n_q^{\ 3}} = \sum_{\substack{\text{allrational} \\ q<x}} \frac{1}{n_q^{\ 2}} \leq \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

So the function is well defined for all x in the unit interval. Since this function is increasing, it has a derivative almost everywhere. Since the function is discontinuous at each rational, where it jumps, it cannot have a derivative on the rationals. We cannot say if the derivative exists everywhere else or not (at least not based on this result), nor what the derivative is where it does exist. It is interesting however, that the derivative does in fact exist except on a set of measure zero.

We can combine the result that monotone functions have derivatives almost everywhere with the fact from, Chapter 1 that a function of bounded variation can be written as a difference of increasing functions to state that a function of bounded variation has a derivative almost everywhere.

We can also state that since any absolutely continuous function is of bounded variation, that any absolutely continuous function has a derivative almost everywhere. Applying the contrapositive of this gives us that the famous function, due to Wierstrass, which is continuous, but has a derivative nowhere, is not absolutely continuous.

The second result is a weaker form of part of the Fundamental Theorem of Calculus. It states that given an increasing function f on a closed interval [a, b]

$$\int_a^b f'(x)dx \leq f(b) - f(a).$$

Recall, that we defined, in Chapter 1, the indefinite integral of a function f to be

$$F(x) = \int_a^x f(t)dt.$$

With this definition and the above results we can make several statements about integrals. The first of these is that the indefinite integral of a function is absolutely continuous and further is of bounded variation. Another is that if f is integrable on [a, b], then

$$\left(\int_a^x f(t)dt\right)' = f(x) \text{ a.e.}$$

A quick sequence of lemmas and propositions, based upon the properties of absolute continuity and bounded variation will lead us to the result with which we will end our chapter.

Theorem: *A function F is an indefinite integral if and only if it is absolutely continuous*[7].

An example of a function which is not the anti-derivative of another function is the Cantor ternary function, based upon the Cantor set. The function is defined as follows.

One way of defining the Cantor set is that it is those elements of [0, 1] which have a ternary (i.e. base 3) expansion without any ones in it. However a ternary expansion is not necessarily unique. In fact precisely the "endpoints" of the Cantor set, the endpoints of the removed open intervals, have two expansions, one without any ones in it. It can be shown that this definition is equivalent to the definition given earlier. Those element's of the unit interval not in the Cantor set must have a one in their ternary expansion.

Given any element of [0, 1] not in the Cantor set, call it x, we can consider the set consisting of the index of the terms in the ternary expansion which are one. This set is a set of natural numbers. It is non-empty, otherwise the point would be in the Cantor set by definition. By the well ordering property of the natural numbers, there is a least element in the set. We will call this element n. Now we will define $f(x)$ to be $\sum_{i=1}^{n} \frac{a_i}{2^i}$, where a_i is $\frac{1}{2}b_i$, where b_i is the i^{th} term in the ternary expansion of x for $i < n$ and $a_n = 1$.

Notice that for any term in a given interval in the complement of the Cantor set, the first term which is one in the ternary expansion will be the same. Further all previous terms will be the same. Thus on a given interval of the complement of the Cantor set, $f(x)$ will be a constant. For those points in the Cantor set the function will use the same sum, using the expansion without ones, but now since there is no term which is one it will be the sum from one to infinity. The series is bounded above by $\sum_{n=1}^{\infty} \frac{1}{2^n} = 1$, and is bounded below by $\sum_{n=1}^{\infty} 0 = 0$, so we know that not only does the series converge for all x in the Cantor set, but that $0 \le f(x) \le 1$ and f is non-decreasing.

We showed that the function was constant on each component of the complement of the Cantor set. Thus the function must do all of its growing on the Cantor set itself. But we have already shown that the function has a value of zero at zero and a value of one at one. We stated that the function was continuous. Thus, the function grows by a value of one on a set of measure zero. Given any ε, such that $0 < \varepsilon < 1$, and any $\delta > 0$, we can get a finite collection of intervals with measure less than δ so that they contain the entire Cantor set, and hence the sum of the jumps of the function over that interval will be one, which is greater than ε, hence the Cantor ternary function is not absolutely continuous.

---

[7] Ibid. 110

We now define a very pathological function f using our fat Cantor set, FC, with measure of a half. On the FC we set $f(x)=0$. For each interval (a, b) in the complement of FC we define:

$$f(x) = (x-a)^2(b-x)^2 \sin\left(\frac{\pi}{2}\frac{b-a}{(x-a)}\right)\sin\left(\frac{\pi}{2}\frac{b-a}{(b-x)}\right)$$

This function has very similar properties to the function $g(x) = x^2 \sin\frac{1}{x}$ at the origin. As a result, f is of bounded variation. On the intervals of the complement of the fat Cantor set, the function is continuous, and has a well defined, but very messy derivative. On the endpoints of the Cantor set, we must look at the difference quotient of the function and take the limit as our h goes to zero to evaluate the derivative. Let a be the left endpoint of the component, and $h > 0$. Then

$$|f(x)| = |x-a|^2|b-x|^2\left|\sin\left(\frac{\pi}{2}\frac{(b-a)}{(x-a)}\right)\right|\left|\sin\left(\frac{\pi}{2}\frac{(b-a)}{(b-x)}\right)\right| \le |x-a|^2, \text{ so}$$

$$\left|\frac{f(a+h)-f(a)}{h}\right| = \left|\frac{f(a+h)-0}{h}\right| \le \frac{h^2}{h} = h.$$

Which, of course, goes to zero. Thus the derivative of f is zero at the left endpoints of the components of the complement of FC. The proof that the derivative of f is zero at the right endpoints will follow the same pattern.

If we next consider a point x in the Cantor set, but not an endpoint, the analysis will be much the same. We must look at the point $(x + h)$, if it is in the Cantor set, our difference quotient is zero. If it is not then, it is in an interval in the complement of the

Cantor set, with left endpoint a, to the right of x. We have $x < a < x + h$. Now we again consider our difference quotient. We will also define h' to be the distance from a to $x+h$, so $h' < h$.

$$f'(x) = \frac{f(x+h) - f(x)}{h} = \frac{f(x+h') - 0}{h}$$
$$\leq \frac{|x+h-a|^2}{h} = \frac{h'}{h} < \frac{h^2}{h} = h.$$

Thus our function has a derivative, which is zero on all points of the fat Cantor set. The derivative of the function on the complement does not have a limit as x tends to an endpoint of the components of the complement. The function wobbles between two values more and more rapidly. This means that the derivative is not continuous at the endpoints, or on any other point in the fat Cantor set.

Since we have a Cantor set with a measure of one half, and that is the set of discontinuities, the Riemann integral for f' does not exist. The requirements for the Lebesgue integral to exist are that the function is finite a.e. and measurable. Both requirements are met, and so f' does indeed have a Lebesgue integral.

The strange property of f, therefore, is that it has a derivative <u>at every point</u>, yet the Riemann integral of the derivative does not exist. The Lebesgue integral does exist. In this case we can even compute it by using symmetry to notice that in any interval in the complement of the fat Cantor set, the integral can be computed on that interval. We can then decompose [0,1] to FC and [0,1]-FC. The integral

$$\int_0^x f'(t)dt = \int_{[0,x]\cap FC} f'(t)dt + \int_{[0,x]\cap FC^C} f'(t)dt.$$

The first integral on the right is zero, and the second can be computed by further subdividing $[0,x]\cap FC^c$ into its open intervals. On these we can use an improper Riemann integral to evaluate the integral on the interval. For any point on the fat Cantor set, the integral is simply zero.

CHAPTER III

THE RIESZ REPRESENTATION THEOREM

We defined the Riemann integral, and stated the Fundamental Theorem of Calculus in the first chapter, along with many definitions which form much of the groundwork for the rest of the thesis. Next we traced the development of the Lebesgue integral in the second chapter. In the next chapter we will develop the variation integral and develop a few results about it. In this chapter, following again the development in Royden's *Real Analysis,* we will prove a few results leading up to one theorem. We will then spend most of the chapter proving this theorem. The theorem we will be proving is the Riesz Representation Theorem, which is one of the most important theorems in functional analysis.

To begin with, we will define a convex function. The geometric idea of a convex curve is that a curve is convex if given two points on the curve, the strait line connecting the two points lies above the curve. If the strait line always falls below the curve, the curve is concave. Our rigorous definition will be nothing more than taking the above definition and re-writing it in a mathematical formula:

Definition: *A function φ is convex on an interval [a, b] if for each x, y in (a, b), and for each λ, with $0 \leq \lambda \leq 1$, we have:*

$$\varphi(\lambda x + (1-\lambda)y) \leq \lambda\varphi(x) + (1-\lambda)\varphi(y).$$

Many useful results follow concerning convex functions. One of these results we are interested in, for the Riesz Representation Theorem, is the following lemma.

Lemma: *If φ is convex on (a, b) and if x, y, x', y' are points in (a, b) with* $x \leq x' \leq y'$ *and* $x' \leq y \leq y'$, *then the chord over* (x', y') *has larger slope than the chord over (x, y), that is:*

$$\frac{\varphi(y)-\varphi(x)}{y-x} \leq \frac{\varphi(y')-\varphi(x')}{y'-x'}.$$ [1]

With this result in hand, we will now move on to proper functional analysis. Many of the definitions we need here were ones we presented in Chapter 1. Among these are the idea of a norm, a linear space, and completeness. We now must define the spaces in which we are working.

A function *f* with the property that $\int_0^1 |f|^p < \infty$ is said to be p-integrable. The set of all p-integrable functions on the interval [*0, 1*] is called L^p. We observe that L^p is a linear space for any *p*. To demonstrate this we note that

$$\int |\alpha f|^p = \int |\alpha|^p |f|^p = |\alpha|^p \int |f|^p,$$

and observe that $|\alpha|^p$ is a number. If the function is *p*-integrable, then the integral of the absolute value of the function to the power of *p* is finite. The product of two numbers is finite. Hence αf is *p*-integrable. We can show in a cute little proof that the sum of two *p*-integrable functions is *p*-integrable. To do this consider, two *p*-integrable functions, *f* and *g*. At any point *x* in the unit interval, one of the two functions must be greater. Thus if *f* (*x*) is the greater, then

---

[1] Royden, 113

$$\left|f(x)+g(x)\right|^p \le \left|f(x)+f(x)\right|^p = \left|2f(x)\right|^p = 2^p\left|f(x)\right|^p$$

or if $g(x)$ is the greater, then

$$\left|f(x)+g(x)\right|^p \le \left|g(x)+g(x)\right|^p = \left|2g(x)\right|^p = 2^p\left|g(x)\right|^p$$

Notice that regardless of whichever term is greater, the right hand side of each of the above inequalities is still a positive number. If we have a number, x, which is greater than another number, y, and we add a positive number, z, to x, we can say that $x+z>y$. This leads to the conclusion that

$$\left|f+g\right|^p \le 2^p\left|f\right|^p + 2^p\left|g\right| \;= 2^p\left(\left|f\right|^p + \left|g\right|^p\right).$$

But, as we noted in the previous chapter, the Lebesgue integral is a linear operator, so

$$\begin{aligned}\int\left|f+g\right|^p &\le \int 2^p\left(\left|f\right|^p+\left|g\right|^p\right)\\ &= 2^p\int\left|f\right|^p + 2^p\int\left|g\right|^p\end{aligned}$$

We know, by assumption that both of the integrals in the bottom line are finite, and that the product of a constant times those integrals is also finite. Thus the integral of the sum of two p-integrable functions is finite. Hence the sum of p-integrable functions is p-integrable. Thus the space of p-integrable functions is a linear space.

We will now define a norm on these functions. For any p greater than or equal to one and less than infinity, we define the norm of a function f in L^p to be:

$$\left\|f\right\| = \left\|f\right\|_p = \sqrt[p]{\int_0^1\left(\left|f\right|^p\right)}.$$

We will show that the above function does in fact describe a norm on L^p. Clearly, this function maps Lebesgue p-integrable functions to the non-negative reals. Next we must make a small observation. If a function is non-zero only on a set of measure zero, then the norm of the function as defined above will be zero. Thus we must work with equivalence classes of functions. Two functions are considered to be the equivalent if they are equal almost everywhere. Since $|f(x)|$ is a non-negative number for all x, the integral of f is zero only if the function is zero up to equivalence.

What remains is to show the Minkowski inequality, that is $\|f+g\|_p \le \|f\|_p + \|g\|_p$ We start this by observing that if either of the functions is zero, the statement is trivial. Thus we can without loss of generality define $\alpha = \|f\|_p \ne 0$, and $\beta = \|g\|_p \ne 0$. We can normalize these functions and get $f_0(x) = \dfrac{|f(x)|}{\alpha}$ and $g_o(x) = \dfrac{|g(x)|}{\beta}$. Now we define λ to be $\lambda = \dfrac{\alpha}{\alpha+\beta}$, leaving $1-\lambda = \dfrac{\beta}{\alpha+\beta}$. Next we consider,

$$\begin{aligned}
|f(x)+g(x)|^p &\le \left(|f(x)|+|g(x)|\right)^p = \left(\alpha f_0(x)+\beta g_0(x)\right)^p \\
&= (\alpha+\beta)^p\left[\left(\frac{\alpha f_0(x)}{(\alpha+\beta)}\right)+\frac{\beta g_0(x)}{(\alpha+\beta)}\right]^p = (\alpha+\beta)^p\left(\lambda f_0(x)+(1-\lambda)g_0(x)\right)^p \\
&\le (\alpha+\beta)^p\left(\lambda f_0{}^p(x)+(1-\lambda)g_0{}^p(x)\right)
\end{aligned}$$

The last line comes about because the function y^p is convex on the unit interval. If we now integrate the ends of the inequality, the first term will of course be the norm of $f+g$ to the power p.

$$\begin{aligned}
\left(\|f+g\|_p\right)^p &\le (\alpha+\beta)^p\left[\lambda\|f_0\|^p+(1-\lambda)\|g_0\|^p\right] \\
&= (\alpha+\beta)^p[\lambda+(1-\lambda)] = (\alpha+\beta)^p.
\end{aligned}$$

The jump from the second line to the third was because $\|f_0\| = 1 = \|g_0\|$. Thus the p-norm is a norm, as can be seen by taking the p$^{\text{th}}$ root of both sides of the inequality.

If p is equal to positive infinity, we can again describe a norm as follows:

$$\|f\|_\infty = \inf\{M : m\{t : f(t) > M\} = 0\}.$$

As quite often happens in mathematics, a few words or explanation will clear up much confusion over what a mathematical statement "means." The norm described above of a function is the infimum of all the numbers, M, such that the measure of points at which the function takes on values which are greater than M is of measure zero. In more descriptive words if the "essential supremum" of a function is M, the function can take on values greater than M, but only on a set of measure zero, and again M is the smallest such number.

Notice that for continuous functions, the essential supremum is the same as the supremum. Hence on $C\,[\,0,\ 1\,]$, this norm is exactly the same sup norm we dealt with throughout Chapter 1.

The proofs that the above norms are norms involved demonstrating that the Minkowski inequality holds. There is another inequality, the Hölder inequality which states:

Theorem: *If p and q are non-negative, extended real numbers, such that* $\frac{1}{p} + \frac{1}{q} = 1,$ *and if $f \in L^p$ and $g \in L^p$., then*

$$\int |fg| \leq \|f\|_p \|g\|_q.$$ [2]

[2] Ibid. 121

A major result of functional analysis, the Riesz-Fisher Theorem can be stated quite succinctly, "The L^p. Spaces are complete." We will need a proposition about complete spaces to prove this result.

Proposition: *A normed linear space X is complete if and only if every absolutely summable series is summable.*[3]

Theorem: *The L^p. Spaces are complete.*[4]

Proof: We will start by noting that using the result of the above proposition, we need only show that any absolutely summable series of functions in L^p is summable.

We start with an absolutely summable sequence of functions, $\langle f_n \rangle$. Since the sequence is absolutely summable, we have, by definition that $\sum_{i=1}^{\infty} \|f_n\|_p = B < \infty$.

We next define a sequence of functions $\langle g_n \rangle$ by $g_n(x) = \sum_{i=1}^{n} |f_i(x)|$.

Using the Minkowski inequality in the L^p norm, we get $\|g_n\|_p \leq \sum_{i=1}^{n} \|f_n\|_p$. Since the right hand side of the inequality is bounded by B so it the left. Thus

$$\int (g_n)^p \leq B^p < \infty.$$

For each x the function values $g_n(x)$ form an increasing series. Thus they must converge to an extended real number. The number to which they converge we will call $g(x)$. Since each g_n is positive, and converge to measurable g we can use Fatou's lemma to show

$$\int \lim g_n \leq \underline{\lim} B^p = B^p.$$

---

[3] Ibid. 124

[4] Ibid. 125

This tells us that g is integrable and hence is finite a.e. For any x where g is finite, $\sum_{i=1}^{n} f_i(x)$ is absolutely summable. The sum we will call $s(x)$. For those x's where g is infinite, we set $s(x)=0$. Notice that if we define $s_n(x) = \sum_{i=1}^{n} f_i(x)$ then s is the limit of $\langle s_n \rangle$ a.e. Since $|s_n(x)| \leq g(x)$ for all n, we have $|s(x)| \leq g(x)$. This gives us that s is integrable, and the triangle inequality gives us $|s_n(x) - s(x)|^p \leq 2^p (g(x))^p$. This gives us the dominating function we need to invoke the Lebesgue convergence theorem and state that

$$\lim \int \left(|s_n - s|^p \right) = \int \lim |s_n - s|^p = \lim \left(\|s_n - s\|_p \right)^p = 0.$$

Thus we have $\sum_{i=1}^{\infty} f_n(x) = s(x)$. So given an absolutely summable series of functions, we have found a function to which the series converges.

QED.

With this result proven, we are now just a proposition, a couple of lemmas, and a few definitions away from the major goal of this chapter.

The following proposition states that in L^p given a function there is a step function and a continuous function so that these functions are arbitrarily close to our original function in the L^p. norm. In fact, if $p=\infty$, this is exactly the same as one of the theorems we gave in Chapter 2.

Proposition: *Given* $f \in L^p, 1 < p \leq \infty$, and $\varepsilon > 0$, *there is a step function* φ *and a continuous function* ψ *such that* $\|f - \varphi\|_p < \varepsilon$, and $\|f - \psi\|_p < \varepsilon$.[5]

---

[5] Ibid. 128

We will now make a couple of definitions.

Definition: *A linear functional F on a space X is a mapping from X to the real numbers, such that* $F(\alpha f + \beta g) = \alpha F(f) + \beta F(g)$ *for all* $f, g \in X$. *A linear functional is bounded if there is an M so that* $|F(f)| \le M\|f\|$ *for all* $f \in X$. *The smallest number M for which the preceding holds is the norm of F.*

We can, of course show that the purported norm above is in fact a norm on the space of linear bounded functionals. The easiest one of the properties to show is that if the norm of a functional is zero, then the functional must be the zero functional. If M=0 then for any f, $|F(f)| \le 0 \cdot \|f\|$. Since the absolute value of a number is non-negative, the functional must map every function to zero. Hence the functional is the zero functional. Next, showing that $\|\alpha F\| = |\alpha|\|F\|$ follows trivially.

$$\|\alpha F\| = \sup \frac{|\alpha F(f)|}{\|f\|} = \sup \frac{|\alpha||F(f)|}{\|f\|},$$

but since we can pull a multiplicative constant through a supremum or infimum, the above is equal to

$$|\alpha| \sup \frac{|F(f)|}{\|f\|} = |\alpha| \cdot \|F\|.$$

and we are done. The final thing we must show is the Minkowski inequality, that the norm of a sum is less than or equal to the sum of the norms. This will follow easily from the fact that the numerator in the definition of the norm is a real number, and we have the triangle inequality on the real numbers. When we put this together with the fact that, as we have shown previously, the supremum of a sum is less than or equal to the sum of the suprema, the result falls out nicely.

$$\|F+G\| = \sup\frac{|F(f)+G(f)|}{\|f\|}$$
$$\leq \sup\frac{|F(f)|+|G(f)|}{\|f\|}$$
$$= \sup\left(\frac{|F(f)|}{\|f\|}+\frac{|G(f)|}{\|f\|}\right)$$

$$\leq \sup\left(\frac{|F(f)|}{\|f\|}\right)+\sup\left(\frac{|G(f)|}{\|f\|}\right)$$
$$= \|F\|+\|G\|.$$

This gives us all we need to show that we have a norm on the space of linear functionals.

We will now state and prove the lemmas we need and then proceed to state and prove the Riesz Representation Theorem.

Proposition: *Each function g in L^q defines a bounded linear functional F on L^p, by*

$$F(f) = \int fg.$$

We have $\|F\| = \|g\|_q$.[6]

Proof: First, we observe that the Hölder inequality tells us that the integral is bounded and that $\|F\| \leq \|g\|_q$. We will also recall from Chapter 2 that the Lebesgue integral is a linear operator. Thus we have that *F* is a bounded linear functional, since the integral clearly maps every function to a number.

[6] Ibid. 131

We next choose a specific function f and apply the functional F. We will then divide the result by the norm of f. This will give us a lower bound for the norm of F. We choose $f(x) = |g(x)|^{q/p} \operatorname{signum} g(x)$. Observe that,

$$
\begin{aligned}
&|f|^p = \left(|g|^{q/p}\right)^p = |g|^q \\
&|g| = |f|^{p/q}, \text{ so } |g|^q = |f|^p \\
&fg = |g(x)|^{q/p} (\operatorname{signum} g(x)) g(x) = |g(x)|^{1+q/p} = |g|^p.
\end{aligned}
$$

We will refer back to these relations in the next lemma as well as here.

By the first equation above, we have that f is p-integrable, and hence $f \in L^p$. We have by the third equation that

$$F(f) = \int fg = \int |g|^q = \left(\|g\|_q\right)^q = \|f\|_p \|g\|_q.$$

Thus we have that the norm of F is:

$$\|F\| \geq \frac{|F(f)|}{\|f\|_p} = \frac{\|f\|_p \|g\|_q}{\|f\|_p} = \|g\|_q.$$

Since we already knew $\|F\| \leq \|g\|_q$, we have $\|F\| = \|g\|_q$.

QED.

This is a very profound proposition. It tells us that if we take *any* q-integrable function, we can define a bounded linear functional on L^p. This tells us that L^p and L^q are very closely related. We will soon show that each linear functional can be identified with a q-integrable function. The term given to the relationship the spaces share is that they are the dual spaces of one another.

Lemma: *Let g be an integrable function on* [*0, 1*], *and suppose that there is a constant M, such that*

$$\left|\int fg\right| \leq M\|f\|_p$$ [7]

for all bounded measurable functions f. Then g is in L^q, *and* $\|g\|_q = M$.

Proof: We will begin by proving the lemma for the case where $1 < p < \infty$. Since $\frac{1}{p} + \frac{1}{q} = 1$, if $p=1$, then $q=\infty$, we will need to prove this case separately.

We will define a sequence of bounded measurable functions $\langle g_n(x)\rangle$ as follows

$$g_n(x) = \begin{cases} g(x) & \text{if } |g(x)| \leq n. \\ 0 & \text{if } |g(x)| > n \end{cases}$$

For each *n*, $g_n(x)$ is clearly bounded by *n*. It is apparent that since *g* was measurable, as it must be in order to be integrable, that $g_n(x)$ is also measurable for each *n*. We now define a sequence $\langle f_n(x)\rangle$, analogous to the function *f* in the previous proposition.

$$f_n(x) = |g_n(x)|^{q/p}\,\text{signum}(g_n(x)).$$

We have the following relations:

$$|f_n(x)|^p = |g_n(x)|^p$$

$$\|f_n\|_p = \left(\|g_n\|_q\right)^{q/p}$$

$$|g_n(x)|^q = f_n(x)g_n(x) = f_n(x)g(x)$$

---

[7] Ibid. 131

Where the last equality in the third equation follows from the fact that if

$g_n(x) \le n$, $g(x) = g_n(x)$, otherwise $g_n(x) = 0 = f_n(x)$, and the statement is trivial.

Observe that,

$$\left(\|g_n\|_q\right)^q = \int |f_n g_n| = \int f_n g \le M\|f\|_p$$

by the hypothesis of the lemma, and

$$M\|f\|_p = M\left(\|g_n\|_q\right)^{q/p}.$$

Since $\dfrac{1}{p} + \dfrac{1}{q} = 1$, $q - \dfrac{q}{p} = 1$ and

$$\frac{\left(\|g_n\|_q\right)^q}{\left(\|g_n\|_q\right)^{q/p}} = \|g_n\| \le M$$

$$\|g_n\|^q \le M^q.$$

Since g is integrable, it is finite a.e. and thus $\langle g_n(x)\rangle$ converges to $g(x)$ a.e. Thus $\left\langle |g_n(x)|^q \right\rangle$ converges a.e. to $|g(x)|^q$, and thus we have the conditions to apply Fatou's lemma, which gives

$$\int |g|^q \le \underline{\lim} \int |g_n|^q \le \underline{\lim} M^q = M^q$$
$$\text{so, } g \in L^q, \|g\|_q \le M.$$

We now need only consider the case $p=1, q=\infty$. We let $\varepsilon > 0$ and set $E = \{x \mid |g(x)| \ge M + \varepsilon\}$. Where M is given by the hypothesis of the lemma. Next we define f to be the characteristic function on the set E. We have

$$\|f\|_1 = \int |\chi_E| = mE.$$

Thus we have that

$$MmE = M\|f\|_1 \geq \left|\int fg\right|$$
$$\left|\int fg\right| \geq (M+\varepsilon)mE$$

Where the first inequality is by hypothesis, and the second by the construction of E. If we combine the two inequalities we get that $mE=0$. Thus by definition of essential supremum, $\|g\|_\infty \leq M$.

QED.

Using the above lemma, we will now prove the Riesz Representation Theorem, which says that every bounded linear functional on L^p, can be represented as the integral of f multiplied by a function g in L^q. This is the converse of the preceding proposition above and tells us that the set of linear functionals on L^p is exactly L^q.

The proof of the theorem we will give follows that given by Royden, as have almost all the proofs in this chapter. However this is an important theorem, and Royden's proofs tend to consist of a small number of statements, each of which requires a proof of its own.

Theorem: (Riesz Representation Theorem) *Let F be a bounded linear functional on* L^p, $1 \leq p < \infty$. *Then there is a function g in* L^q *such that*

$$F(f) = \int fg.$$

We also have $\|F\| = \|g\|_q$.[8]

---

[8] Ibid. 132

Proof: This proof will start in much the same way as our construction of the Lebesgue integral did in the previous chapter, that is by starting with simple functions and using approximation theorems to expand the range of functions to cover the entire domain.

We will consider first a set of functions, χ_s= the characteristic function on the interval [*0, s*]. For each *s* in the unit interval, *F* will map χ_s to a number, as *F* is a functional. Thus as *s* ranges over the unit interval we will define a function $\Phi(s) = F(\chi_s)$. We will show that this function is absolutely continuous.

Let ε > 0. Consider a finite collection of intervals $\{(s_i, s'_i)\}_{i=1}^{n}$ with

$\sum_{i=1}^{n}(s'_i - s_i) < \left(\frac{\varepsilon}{\|F\|}\right)^p = \delta$. If we set

$f(x) = \sum_{i=1}^{n}\left(\chi_{x'_i}(x) - \chi_{x_i}(x)\right) signum\left(\Phi(s'_i) - \Phi(s_i)\right)$, then $F(f) = \sum_{i=1}^{n}\left(\Phi(s'_i) - \Phi(s_i)\right)$.

Notice that $f(x)$ always has an absolute value of one within the collection of intervals, $(s_i, s_i']$ and zero outside. Thus, $\left(\|f\|_p\right)^p < \delta$, and we have

$$F(f) \le \|F\| \cdot \|f\|_p \le \|F\| \delta^{1/p}, \text{ so}$$

$$F(f) \le \|F\| \cdot \frac{\varepsilon}{\|F\|} = \varepsilon.$$

Notice that *F(f)* is the growth of the function Φ over the collection of intervals. Thus Φ is absolutely continuous.

We can now apply the major result from Chapter 2 and say that there is a function *g* which is the derivative a.e. of Φ, such that $\Phi(s) = \int_0^s g$. But since

$F(\chi_s) = \Phi(s)$, we conclude that

$$F(\chi_s) = \int_0^s g = \int_0^1 g\chi_s.$$

The last equality above follows from the fact that we can break the last integral up into a sum of the integral from 0 to s and from s to 1. The characteristic function, being one on the former, has no effect there. It does however change the integrand to zero on the interval (s, 1]. Thus the second integral becomes zero, and we have what we started with. This equation tells us that a bounded linear functional defined on the characteristic function of [0, s] can be written as the integral of the product of the characteristic function and some integrable function, g.

The next step we will take is based upon the fact that the Lebesgue integral is a linear operator. This will let us expand our function to step functions. From step functions, we can use Littlewood's second principle,(again) to expand to bounded and measurable functions in L^p. From there we will use the ubiquitous analysis method of adding and subtracting something to a difference. We then follow the standard operating procedure and apply the triangle inequality to the result from the previous step. This will allow us to expand our domain to the fullest extent.

But, we must go one step at a time. First we will recall from Chapter 2 that a step function, ψ, can be written in its canonical representation as $\sum_i c_i\chi_i$. Notice, the E_i's we are using are intervals, but not necessarily half-open intervals. The biggest possible difference between the different types of intervals is the endpoints of the intervals. These form an at most countable set. The measure of this set is zero, and the set will have no contribution to the integral.

Now we will apply the linearity of the Lebesgue integral.

$$\begin{aligned}\int g\psi &= \int g\left(\sum_i c_i \chi_i\right)\\ &= \int \sum_i g c_i \chi_i\\ &= \sum_i \int c_i g \chi_i\\ &= \sum_i c_i \int g \chi_i\\ &= \sum_i c_i F(\chi_i)\end{aligned}$$

Now we will work our way through the same steps using the fact that we know F is a linear functional.

$$\begin{aligned}\sum_i c_i F(\chi_i) &= \sum_i F(c_i\, \chi_i)\\ &= F\left(\sum_i c_i\, \chi_i\right)\\ &= F(\psi).\end{aligned}$$

We know from a previous theorem that given a bounded measurable function, we can find a step function so that $|f(x) - \psi(x)| > \varepsilon$, only on a set of measure less than ε, and that the step function will be bounded by the same bound as f. We can use this fact to construct a sequence of step functions $\langle \psi_N \rangle$ so that the ε in question converges to zero. This property is known as convergence in measure. One consequence of convergence in measure is that a subsequence of the sequence we constructed, call it $\langle \psi_n \rangle$ to avoid double subscripts, will converge to our function f almost everywhere.

Since we know that each of our ψ_n*'s* are bounded by M, the bound on f, we know that the sequence $\langle |f - \psi_n|^p \rangle$ is bounded on the unit interval by M^p. We also know that the sequence converges almost everywhere to zero. These are the criteria we need to apply the Bounded Convergence Theorem. This theorem tells us that the integral of our sequence, which is the p^{th} power of the norm of $(f - \psi_n)$, converges to zero. Hence the norm of $(f - \psi_n)$ converges to zero.

Now we apply the fact that F is a bounded linear functional to get that

$$|F(f) - F(\psi_n)| = |F(f - \psi_n)|.$$

The next step is to apply the definition of the norm of a functional to get

$$|F(f - \psi_n)| \le \|F\| \cdot \|f - \psi_n\|_p.$$

We will now observe that the norm of F is a fixed number, and that the norm of the elements of our sequence goes to zero. Hence

$$|F(f - \psi_n)| = |F(f) - F(\psi_n)| \to 0.$$

Notice that the equality comes about because, once again, F is a linear functional. This tells us that $\lim\limits_{n\to\infty} F(\psi_n) = F(f)$.

Also we know that $g\psi_n$ is bounded by $|g|M$. This gives us, by the Lebesgue convergence theorem,

$$\int fg = \lim \int \psi_n g.$$

Which now holds for all bounded and measurable functions f in L^p.

If we now observe once more that $|F(f)| \le \|F\| \|f\|_p$ for every p-integrable function, then we take $M = \|f\|_p$, the lemma on page 57 gives us that g is q-integrable and that $\|g\|_q \le \|F\|$.

We will now proceed with the *coup d'grace*. We will let f be any p-integrable function, bounded or not. We also choose ε > 0. Since f is p-integrable we can get a step function η on [*0, 1*], so that $\|f-\eta\|_p < \varepsilon$. Since all step functions are, by definition, bounded, we know $F(\eta) = \int \eta g$. This leads to the following chain of inequalities:

$$\begin{aligned}
\left|F(f) - \int fg\right| &= \left|F(f) - \int \eta g + \int \eta g - \int fg\right| \\
&\leq \left|F(f) - \int \eta g\right| + \left|\int \eta g - \int fg\right| \\
&\leq \left|F(f) - F(\eta)\right| + \left|\int (\eta - f)g\right| \\
&\leq \left|F(f-\eta)\right| + \left|\int (\eta - f)g\right| \\
&\leq \|F\| \cdot \|f-\eta\| + \|g\|_q \cdot \|f-\eta\| \\
&= \left(\|F\| + \|g\|_q\right)\|f-\eta\| \\
&\leq \left(\|F\| + \|g\|_q\right)\varepsilon.
\end{aligned}$$

If we now note that $\left(\|F\| + \|g\|_q\right)$ is a number, as both F and g are already determined, and that ε is arbitrarily small, the only conclusion to be reached is that

$$\left|F(f) - \int fg\right| = 0,$$

and hence that

$$F(f) = \int fg.$$

The final part of the theorem, that $\|g\|_q = \|F\|$, comes directly from the previous lemma.

QED.

CHAPTER IV

THE VARIATION INTEGRAL

We now have all the tools we will need to complete our goal. We will, in this chapter, develop our new integral, as we developed the Lebesgue integral in the second chapter. We will then give and prove a theorem, giving us a way of relating linear functionals on our space with our new integral. This will accomplish the analog of what we accomplished with the Riesz Representation Theorem in Chapter Three.

In Chapters Two and Three the space of functions we were working on we the Lebesgue integrable functions. The conclusion of the Riesz Representation Theorem was that for each Lebesgue q-integrable function, g, we could define a linear functional on the p-integrable functions defined by $F(f) = \int fg$. We conversely found that every linear functional on the p-integrable functions was defined by some function g as above. Thus there is a bijective correspondence between the functions in L^q, and the functionals on L^p.

We will develop such a theorem in this chapter. We will work with a much stronger norm, but our space of functions will be more restricted. We will only work on the space of absolutely continuous functions defined on the interval [0,1], which we will refer to as AC [0,1]. A slight restriction we will need to place on this set is that for any f in AC [0,1], $f(0) = 0$.

We showed in Chapter One that this set is a linear space.

A normed space of function consists of two things, the first of which is the set of objects in the space. This we have just defined. The other thing involved is the norm we put on the space. The norm we will put on AC [0,1] is the bounded variation norm. The value of the norm of a function using this norm is the total variation T, of the function on the interval [0, 1]. We defined this in Chapter One.

We must now show that this does in fact form a norm on the space. Clearly the norm maps any function in AC [0,1] to the non-negative reals. Further, we now show that, for the total variation to be zero, the function must be the zero function. Suppose the function had a value of y, not zero at a point, p. We could then pick the partition, $P = \{0, x, 1\}$. The variation of f over this partition is

$$\begin{aligned} V_P &= |f(x) - f(0)| + |f(1) - f(x)| \\ &= |y - 0| + |f(1) - y|. \end{aligned}$$

The first term in this sum is a positive number and the second term is non-negative. Hence the variation $V_0^1 f = \sup_\sigma \sum_\sigma |f(x_i) - f(x_{i-1})|$ must be greater than zero.

The norm is defined as the supremum over all partitions of the variation of the function. Given any scalar, we can pull a scalar out of each term in the sum, add, then multiply the scalar back in. That is

$$\sum |\alpha f(x_i) - \alpha f(x_{i-1})| = |\alpha| \sum |f(x_i) - f(x_{i-1})|.$$

Using this fact we can easily show that:

$$\|\alpha f\|_{BV} = V_0^1(\alpha f) = |\alpha| V_0^1(f).$$

The only thing remaining, in order to show that the bounded variation norm is a norm, is the Minkowski inequality. We will show this below, making use of the fact that the supremum of a sum is less than or equal to the sum of the suprema, as we have shown before.

$$\begin{aligned}
V(f+g) &= \sup_{\sigma} \sum_{\sigma} \left| f(x_i) + g(x_i) - \left(f(x_{i-1}) + g(x_{i-1}) \right) \right| \\
&= \sup_{\sigma} \sum_{\sigma} \left| f(x_i) - f(x_{i-1}) + g(x_i) - g(x_{i-1}) \right| \\
&\leq \sup_{\sigma} \sum_{\sigma} \left| f(x_i) - f(x_{i-1}) \right| + \sup_{\sigma} \sum_{\sigma} \left| g(x_i) - g(x_{i-1}) \right| \\
&= \|f\|_{BV} + \|g\|_{BV}.
\end{aligned}$$

When we were dealing with Lebesgue measurable functions, we approximated them with simple functions in order to construct the Lebesgue integral. We cannot do this here, as simple functions, and their subset of step functions are not, in general continuous, let alone absolutely continuous. Further, if f is absolutely continuous, and s is a step function, then

$$\|f - s\|_{BV} = \|f\|_{BV} + \|s\|_{BV}$$

and we cannot possibly hope to have a sequence of step functions converging to f in the BV norm, except in the trivial case of f being the zero function.

To get around this, we will in developing our new integral, approximate our functions with polygonal approximations to the functions. A polygonal approximation can be thought of as taking a partition of the interval, evaluating the function at the points of the partition, and "connecting the dots" with straight line segments. More rigorously, if we have a partition, $\sigma = \left\{ 0 = x_0 < x_i < \cdots < x_{n-1} < x_n = 1 \right\}$, then we define the polygonal approximation, $pf_\sigma = f(x_{i-1}) + (f(x_i) - f(x_{i-1}))\left(\frac{x - x_{i-1}}{x_i - x_{i-1}}\right)$, for all $x \in (x_{i-1}, x_i]$, and $f(0) = 0$.

We can show that polygonal approximations are, in fact absolutely continuous. The function, pf_σ, will on each interval $(x_{i-1}, x_i]$, have slope m_i. Since there are only a finite number of subintervals in σ, we can take the largest slope, in absolute value and call it m. Then given ε > 0, our δ will simply be any number less than $\frac{\varepsilon}{|m|}$. Now, if we take a collection of subintervals, $\{(x_i, y_i)\}_{i=1}^n$, where $\sum_{i=1}^n (y_i - x_i) < \delta$, we get that

$$\begin{aligned}\sum_{i=1}^n |f(y_i) - f(x_i)| &\le \sum_{i=1}^n |m||y_i - x_i| \\ &= |m|\sum_{i=1}^n |y_i - x_i| < |m|\delta = |m|\frac{\varepsilon}{|m|} = \varepsilon.\end{aligned}$$

Hence, our polygonal approximations will be absolutely continuous, and thus elements of $AC[\,0, 1\,]$.

The next major step in our construction will be to show that the closure of the space of polygonal functions under the BV norm is $AC[\,0, 1\,]$. That is, given a Cauchy sequence of polygonal functions in $AC[\,0, 1\,]$, there is an absolutely continuous function to which the sequence converges, and that $pf_\sigma \to f$ in the BV norm.

Recall that as we showed in Chapter One, any absolutely continuous function is of bounded variation. We next show that the supremum of a function is less than or equal to the total variation. Consider a function $f \in AC[\,0, 1\,]$. In the unit interval there is a point x for which $|f(x)| = \|f\|_\infty$. We will find the variation of f over the partition, $\sigma = \{0, x, 1\}$.

The variation over this partition is:

$$\begin{aligned}&|f(0)-f(x)|+|f(x)-f(1)|=\\&|0-f(x)|+|f(x)-f(1)|=\\&|f(x)|+|f(x)-f(1)|=\\&\|f\|_\infty+|f(x)-f(1)|\geq\|f\|_\infty\end{aligned}$$

Since the variation of a function is the supremum of the variation over all partitions, and we have found one partition whose variation was greater than the sup norm of the function, we are done.

An easy consequence to see of this result is that if a sequence is Cauchy in the BV-norm, it is Cauchy in the sup norm. To see this consider a Cauchy sequence of absolutely continuous functions, $\langle f_n \rangle$, in AC[0, 1] under the BV norm. By the definition of Cauchy, given ε >0, there is a natural number, N, so that if n, $m > N$, then $\|f_n - f_m\|_{BV} < \varepsilon$. But notice that $\|f_n - f_m\|_\infty < \|f_n - f_m\|_{BV} < \varepsilon$, so the sequence is also Cauchy under the sup norm, using the same N as the sequence in the BV norm. Similarly, a sequence of functions which converge in the BV norm converge in the sup norm.

As we showed in Chapter One, the space C [0, 1] is complete under the sup norm. We will use this fact and follow the same steps to show that the closure of the space of polygonal functions under the BV norm is AC [0, 1].

Theorem: *A Cauchy sequence of polygonal functions on* [0, 1] *under the BV norm converge to an absolutely continuous function.*

Proof: We will start by considering a Cauchy sequence of polygonal functions, $\langle p_n \rangle$, under the BV norm. As we have noted, this means $\langle p_n \rangle$ is Cauchy under the sup norm.

Thus, since polygonal functions are continuous, and using that $C[0, 1]$ is complete under the sup norm, there is a function *f*, which is the limit of $\langle p_n \rangle$ in the sup norm.

We now have three things to prove in order to establish this theorem. The first is that f is of bounded variation. Next, we will show that f is absolutely continuous. Finally, we will show the $\langle p_n \rangle$ converges to f *in the BV Norm.*

In showing that the function is of bounded variation, we will consider $\langle p_n \rangle$. The norms of these functions form a Cauchy sequence of real numbers. Since the sequence $\langle \|p_n\|_{BV} \rangle$ is Cauchy, it is bounded, as all Cauchy sequences are bounded. We choose a bound B and note that for all n, $\|p_n\|_{BV} \le B$.

Next, we let $\varepsilon > 0$, and fix a partition σ. This partition has M points. Since we know $\langle p_n \rangle$ converges to f in the sup norm, there is a natural number N, so that

$$|f(x) - p_n(x)| < \frac{\varepsilon}{4M}$$

for all $n > N$, and for all x in the unit interval.

We now choose $n > N$. Then

$$\sum_\sigma |f(x_i) - f(x_{i-1})| = \sum_\sigma |f(x_i) - p_n(x_i) + p_n(x_i) - p_n(x_{i-1}) + p_n(x_{i-1}) - f(x_{i-1})|$$

When we apply the triangle inequality of the reals, we get

$$\begin{aligned}\sum_\sigma |f(x_i) - f(x_{i-1})| &\le \sum_\sigma |f(x_i) - p_n(x_i)| + \sum_\sigma |p_n(x_i) - p_n(x_{i-1})| + \sum_\sigma |p_n(x_{i-1}) - f(x_{i-1})| \\ &\le \sum_{i=1}^M \left|\frac{\varepsilon}{4M}\right| + \sum_\sigma |p_n(x_i) - p_n(x_{i-1})| + \sum_{i=1}^M \left|\frac{\varepsilon}{4M}\right|\end{aligned}$$

$$= \frac{\varepsilon}{4} + \sum_{\sigma} |p_n(x_i) - p_n(x_{i-1})| + \frac{\varepsilon}{4}$$

$$< \frac{\varepsilon}{2} + B < \infty.$$

Since both B and ε are fixed, f is of bounded variation.

We must now show that the function f is absolutely continuous. We start by letting $\varepsilon > 0$. Since $\langle p_n \rangle$ is Cauchy in the BV norm, we can get a number N_1, so that $\|p_n - p_m\|_{BV} < \frac{\varepsilon}{4}$, for all $n,m > N_1$. We next choose such an n. Since p_n is absolutely continuous, there is a $\delta > 0$ so that for any finite collection of intervals $\{[x_i, y_i]\}_{i=1}^{k}$, with total length less than δ, we get $\sum_{i=1}^{k} |p_n(y_i) - p_n(x_i)| < \frac{\varepsilon}{4}$. We continue by choosing such a collection of intervals. Since $\langle p_n \rangle$ converges to f in the sup norm, we can get a natural number N_2, so that for any $x \in [\,0, 1\,]$, $|f(x) - p_n(x)| < \frac{\varepsilon}{4k}$. We now choose $m = N_1 + N_2$. This gives

$$\sum_{i=1}^{k} |f(y_i) - f(x_1)|$$
$$= \sum_{i=1}^{k} \left|(f(y_i) - p_m(y_i)) + (p_m(y_i) - p_n(y_i)) + (p_n(y_i) - p_n(x_i)) + (p_n(x_i) - p_m(x_i)) + (p_m(x_i) - f(x_i)\right|$$
$$= \sum_{i=1}^{k} \left|(f(y_i) - p_m(y_i)) + (p_m(y_i) - p_n(y_i) - p_m(x_i) + p_n(x_i)) + (p_n(y_i) - p_n(x_i)) + (p_m(x_i) - f(x_i))\right|$$

This monster does reduce to something somewhat more manageable when we break it into pieces using the classical triangle inequality.

$$\sum_{i=1}^{k} |f(y_i) - f(x_i)| \le \sum_{i=1}^{k} |f(y_i) - p_m(y_i)| + \sum_{i=1}^{k} |p_m(y_i) - p_n(y_i) - p_m(x_i) + p_n(x_i)|$$
$$+ \sum_{i=1}^{k} |p_n(y_i) - p_n(x_i)| + \sum_{i=1}^{k} |p_m(x_i) - f(x_i)|$$

The first and last sums add up to less than $\frac{\varepsilon}{4}$ each because of the sup norm (or uniform) convergence. The third term adds up to less than $\frac{\varepsilon}{4}$ because of the absolute continuity of p_n. The second term can be rearranged to look like

$$\sum_{i=1}^{k} \left| \left(p_m(y_i) - p_n(y_i)\right) - \left(p_m(x_i) - p_n(x_i)\right) \right|.$$

This term is a part of the variation on a partition of the function $(p_m - p_n)$. This will make it less than or equal to the norm of the function, which is less than $\frac{\varepsilon}{4}$, since $\langle p_n \rangle$ is Cauchy. By adding these pieces together, we get a total jump in the function over the collection of intervals of less than ε. Thus f is absolutely continuous.

Our last step in the proof of this theorem is to show $\langle p_n \rangle$ converges to f in the BV norm. That is we must show that, $\|f - p_n\|_{BV} \to 0$. So given ε > 0, we must come up with an N so that for any $n > N$, then $\|f - p_n\|_{BV} < \varepsilon$. As we know we have a Cauchy sequence of polygonal functions, we will choose the N, so that for $n, m > N$, $\|p_n - p_m\|_{BV} < \frac{\varepsilon}{3}$. We next choose any partition we want. It will have k points in it. We must show now that

$$\sum_{i=1}^{k} \left| \left(f(x_i) - p_n(x_i)\right) - \left(f(x_{i-1}) - p_n(x_{i-1})\right) \right| < \varepsilon.$$

We will now choose, based on uniform convergence, an m so that for any x in the unit interval, $\sum_{i=1}^{k} |f(x) - p_m(x)| < \frac{\varepsilon}{3k}$. Another application of the triangle inequality will now yield

$$
\begin{aligned}
&\sum_{i=1}^{k}\left|\left(f(x_i)-p_n(x_i)\right)-\left(f(x_{i-1})-p_n(x_{i-1})\right)\right| \\
&=\sum_{i=1}^{k}\left|\left(f(x_i)-p_n(x_i)\right)-\left(f(x_{i-1})-p_n(x_{i-1})\right)+\left(p_m(x_i)-p_m(x_i)\right)+\left(p_m(x_{i-1})-p_m(x_{i-1})\right)\right| \\
&=\sum_{i=1}^{k}\left|\left(f(x_i)-p_m(x_i)\right)+\left(f(x_{i-1})-p_m(x_{i-1})\right)+\left[\left(p_m(x_i)-p_n(x_i)\right)-\left(p_m(x_{i-1})-p_n(x_{i-1})\right)\right]\right| \\
&\leq\sum_{i=1}^{k}\left|\left(f(x_i)-p_m(x_i)\right)+\left(f(x_{i-1})-p_m(x_{i-1})\right)\right|+\sum_{i=1}^{k}\left|\left(p_m(x_i)-p_n(x_i)\right)-\left(p_m(x_{i-1})-p_n(x_{i-1})\right)\right|
\end{aligned}
$$

The second term on the last line is an underestimate of $\left\|p_n - p_m\right\|_{BV}$. We already know this to be less than $\frac{\varepsilon}{3}$. The first term we can break into

$$
\begin{aligned}
&\sum_{i=1}^{k}\left|\left(f(x_i)-p_m(x_i)\right)-\left(f(x_{i-1})-p_m(x_{i-1})\right)\right| \\
&\leq\sum_{i=1}^{k}\left|\left(f(x_i)-p_m(x_i)\right)\right|+\sum_{i=1}^{k}\left|\left(f(x_{i-1})-p_m(x_{i-1})\right)\right| \\
&\leq\frac{\varepsilon}{3}+\frac{\varepsilon}{3}=\frac{2\varepsilon}{3}.
\end{aligned}
$$

When we throw in the $\frac{\varepsilon}{3}$ from the previous term, we get the sum of all terms to be less than ε. Thus our sequence converges in the BV norm.

QED.

Having gone from a Cauchy sequence of polygonal functions in the BV norm to an absolutely continuous function, we must now turn the process around and go from an absolutely continuous function, *f*, to a sequence of polygonal approximations which converge to *f*.

Theorem: *Every function f, in AC* [0, 1] *is the limit of a sequence of polygonal approximations to that function.*

Proof: We start with a function *f* in *AC* [0, 1] and an ε > 0. By our results in Chapter Two, we know there is a function $f' \in L^1$, so that $\int_0^x f'(t)dt = f(x)$. We have now moved from the sometimes bizarre space of the bounded variation norm to the more farmiliar space of Lebesgue integrable functions under the L^1 norm. We know therefore that there is a step function s_n, so that $\|f'-s_n\|_1 < \frac{\varepsilon}{2^n}$.

It is not too difficult to see that the integral of a step function is a polygonal function. We will now define a function $\Phi_n(x)$ to be this integral. That is

$$\Phi_n(x) = \int_0^x s_n(t)dt.$$

For an absolutely continuous function, we have that

$$\|f\|_{BV} = \int |f'|.$$

This being the case, we can state that for each *n*, the is an s_n and a Φ_n so that

$$\|f - \Phi_n\|_{BV} = \int |f'-s_n| < \frac{\varepsilon}{2^n}.$$

This, of course tells us that $\langle \Phi_n \rangle$ converge to f in the BV norm.

QED.

We will next show that if we take a sequence of polygonal approximations $\langle pf_\sigma \rangle$, on a directed partially ordered set of partitions, the sequence will converge to f in the BV norm.

Theorem: $\lim\limits_{\sigma} pf_\sigma = f$.

Proof: We start by letting ε > 0. We continue by taking a sequence $\langle q_n \rangle$ of polygonal functions converging to f. There is a term in this sequence, q_n, so that $\|f - q_n\|_{BV} < \frac{\varepsilon}{2}$.

Next we consider the set σ of the points at which q_n has corner points. Clearly σ is a partition of the unit interval. Let σ' be a partition beyond σ. We want to show that $pf_{\sigma'}$, the polygonal approximation to *f* with corners on σ', lies within ε of *f* in the BV norm. That is $\|f - pf_{\sigma'}\|_{BV} < \varepsilon$.

Since the BV norm is a norm, we can apply the Minkowski inequality to $\|f - pf_{\sigma'}\|_{BV}$ to yield

$$\begin{aligned} \|f - pf_{\sigma'}\|_{BV} &\le \|f - q_n\|_{BV} + \|q_n - pf_{\sigma'}\|_{BV} \\ &< \frac{\varepsilon}{2} + \|q_n - pf_{\sigma'}\|_{BV} \end{aligned}$$

It now remains to show $\|q_n - pf_{\sigma'}\|_{BV} < \frac{\varepsilon}{2}$. We know q_n is an approximation to *f*. Notice that for $x \in \sigma', pf_{\sigma'}(x) = f(x)$. So $\|q_n - pf_{\sigma'}\|_{BV}$ is an approximation of $\|q_n - f\|_{BV}$ on the partition $\sigma \cup \sigma'$. Thus, $\sum_{\sigma \cup \sigma'} |q_n - pf_{\sigma'}| \le V_0^1(q_n - f)$, and we get that

$$\|q_n - pf_{\sigma'}\|_{BV} \le \|q_n - f\|_{BV} < \frac{\varepsilon}{2}.$$

QED.

At this point we refer back to the Riesz Representation Theorem. If we have any linear functional F defined on the L^p functions, we know that $F(f) = \int fg$, for some *g* in L^q, where $\frac{1}{p} + \frac{1}{q} = 1$. Let us now consider a new set function,

$$\mu(A) = \int_A g.$$

Except for the fact that the integral might be negative, we have all the properties of a measure for the set. In fact, we have a "signed measure."

Recall that we could approximate any measurable function f with a simple function ψ, so that $\|f - \Psi\|_1 < \varepsilon$. Further, we could decompose ψ into a sum of numbers and characteristic functions, the canonical representation, where

$$\Psi(x) = \sum \alpha_i \chi_{E_i}.$$

If we now apply a linear functional F to f, we will get

$$\begin{aligned} F(f) \approx F(\Psi) &= \sum \alpha_i F(\chi_{E_i}) \\ &= \sum f(t_i)\mu(E_i). \end{aligned}$$

But, if we recall that μ is a measure, the last line represents an integral with respect to the μ measure in the limit. This is what we will do to define our new integral.

Our first step is to decompose our functions into some simpler functions with which we can work. We cannot use step functions since we are working in a space of absolutely continuous functions. Step functions, non-trivial ones anyway, are discontinuous. Further, as we have noted, we cannot use them to approximate functions under the BV norm anyway.

With some thought, we can see that we can decompose a polygonal function based on its partition σ. The functions will decompose into linear combinations of functions $\Phi_{a,b}$, where a and b are consecutive points in σ. We define $\Phi_{a,b}$ as follows

$$\Phi_{a,b} = \begin{cases} 0, & \text{if } x \in [0,a] \\ \dfrac{x-a}{b-a}, & \text{if } x \in (a,b] \\ 1, & \text{if } x \in (b,1] \end{cases}$$

Also, notice that for a polygonal approximation to a function f,

$$pf_\sigma(x) = \sum_\sigma \left(f(x_i) - f(x_{i-1}) \right) \Phi_{x_{i-1},x_i}(x), \text{ where } x \in (x_{i-1}, x_i).$$

We can show that the Φ functions add convexly. That means that there are real numbers λ_1, λ_2 so that

$$\lambda_1 \Phi_{a,b} + \lambda_2 \Phi_{b,c} = \Phi_{a,c}.$$

The easiest way to prove this is simply to give the values for λ_1 and λ_2.

$$\lambda_1 = \frac{b-a}{c-a},$$
$$\lambda_2 = \frac{c-b}{c-a}.$$

Notice that for $x \in [a,b]$, $\Phi_{b,c}(x) = 0$, and we have

$$\begin{aligned} \Phi_{a,c}(x) &= \frac{b-a}{c-a} \Phi_{a,b}(x) \\ &= \frac{b-a}{c-a} \left(\frac{x-a}{b-a} \right) = \frac{x-a}{c-a}. \end{aligned}$$

This is exactly what it should be. For $x \in [b,c]$, we have $\Phi_{a,b}(x) = 1$, so

$$\begin{aligned} \Phi_{a,c}(x) &= \frac{b-a}{c-a} + \frac{c-b}{c-a} \Phi_{b,c}(x) \\ &= \frac{b-a}{c-a} + \frac{c-b}{c-a} \left(\frac{x-b}{c-b} \right) \\ &= \frac{b-a}{c-a} + \frac{x-b}{c-a} = \frac{x-a}{c-a} \end{aligned}$$

which is again what is expected.

If we now consider a linear functional on the space AC [0, 1], F, and apply it to a function $\Phi_{a,b}$, we will have defined a convexly additive signed set function (or convex measure) on [a, b]. We will denote it by μ. To demonstrate this

$$\begin{aligned}\mu((a,c]) &= F(\Phi_{a,c}) \\ &= F(\lambda_1\Phi_{a,b} + \lambda_2\Phi_{b,c}) \\ &= \lambda_1 F(\Phi_{a,b}) + \lambda_2 F(\Phi_{b,c}) \\ &= \lambda_1\mu((a,b]) + \lambda_2\mu((b,c]).\end{aligned}$$

Now we approximate an absolutely continuous function f with pf_σ, decompose this, and apply F to it to get

$$\begin{aligned}F(f) \approx F(pf_\sigma) &= \sum_\sigma \Delta_i f\left(F(\Phi_{x_{i-1},x_i}\right) \\ &= \sum_\sigma \Delta_i f\mu((x_{i-1},x_i]).\end{aligned}$$

If we now take the limit of the above sums as we let $pf_\sigma \to f$, as we have shown we can for any absolutely continuous function, our sum goes to our new variation integral.

$$\begin{aligned}F(f) = \lim_\sigma F(pf_\sigma) &= \lim_\sigma \sum_\sigma \Delta_i fF\left(\Phi_{x_{i-1},x_i}\right) \\ &= \lim_\sigma \sum_\sigma \Delta_i f\mu((x_{i-1},x_i]) \\ &= v\int \Delta f d\mu.\end{aligned}$$

We have, in one fell swoop, now both defined our integral, and given our Integral Representation Theorem. Notice also that the integral we have defined can be evaluated. Everything works on intervals and can be numerically calculated. While this integral does not have the advantage of being evaluated simply by taking its anti-derivative, it can be evaluated. This is not something which can be said about the Lebesgue integral in most cases.

That we can evaluate the integral follows from the fact that like the Riemann integral, the limit of the sums as the norm of the partition goes to zero is the integral, as we show below.

We start with an absolutely continuous function f. We know that the limit of pf_σ is f, as we take the limit of a directed partition. So there is a partition σ_1, with N points, so that if $\sigma_1 \subset \sigma$, then $\left\| pf_{\sigma_1} - pf_\sigma \right\|_{BV} < \frac{\varepsilon}{3}$, because the directed set, $\langle pf_\sigma \rangle$ is Cauchy. Since f is absolutely continuous, there is a $\delta' > 0$, so that if $\sum_{i=1}^{n} (y_i - x_i) < \delta'$, then $\sum_{i=1}^{n} \left| f(y_i) - f(x_i) \right| < \frac{\varepsilon}{3}$. Let $\delta = \frac{\delta'}{2N}$. We next choose any partition σ_2 so that

$$\max_{x_i \in \sigma_2} \left\{ \left(x_i - x_{i-1} \right) \right\} < \delta.$$

We now define the union of the two partitions $\sigma_1 \cup \sigma_2 = \sigma$. Now note,

$$\left\| f - pf_{\sigma_2} \right\|_{BV} \le \left\| f - pf_\sigma \right\|_{BV} + \left\| pf_\sigma - pf_{\sigma_2} \right\|_{BV}.$$

Since σ is a refinement of σ_1, the first term in the right hand side of the above inequality is less than $\frac{\varepsilon}{3}$.

Now we define the set A to be the set of intervals defined by σ, $[x_{i-1}, x_i]$, so that either of the two endpoints are in σ_1. Notice that $\left\| pf_\sigma - pf_{\sigma_2} \right\|_{BV} = \sum_A \left\| pf_\sigma - pf_{\sigma_2} \right\|_{BV}$ because for any interval not in A, both endpoints are in σ_2, and $pf_\sigma = pf_{\sigma_2}$ on the sub-interval.

Since we know the norm of the partition σ is less than δ', we know the length of each interval in A is at most $\frac{\delta'}{2N}$, and we know there are at most $2N$ intervals in A. So we know the total width of all the intervals in A is less than $2N\frac{\delta'}{2N} = \delta'$.

If we next apply the Minkowski inequality to $\left\|pf_\sigma - pf_{\sigma_2}\right\|_{BV}$, we get

$$\left\|pf_\sigma - pf_{\sigma_2}\right\|_{BV} = \sum_A \left\|pf_\sigma - pf_{\sigma_2}\right\|_{BV} \le \sum_A \left\|pf_\sigma\right\|_{BV} + \sum_A \left\|pf_{\sigma_2}\right\|_{BV}.$$

But by absolute continuity, we have that both sides of the right hand side of the above inequality are less than $\frac{\varepsilon}{3}$. Thus given any partition σ_2 with norm smaller than δ, we have that

$$\begin{aligned}\left\|f - pf_{\sigma_2}\right\|_{BV} &\le \left\|f - pf_\sigma\right\|_{BV} + \sum_A \left\|pf_\sigma\right\|_{BV} + \sum_A \left\|pf_{\sigma_2}\right\|_{BV} \\ &< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.\end{aligned}$$

Thus we can define our integral as the limit of a directed partition, or for the sake of computability as the limit of the norm of the partition.

The beauty of the Riemann integral is that you can simply take the values of the anti-derivatives at the endpoints of the interval of integration. This simple fact is the climax of first semester freshman calculus. In second semester calculus, we learn that we are not always able to find an anti-derivative. For instance laboratory results are often noisy, and can seldom be represented as a closed form function. Even when it they can be, it is often easier to find an integral some other way.

By using a very fine partition, we can evaluate to any desired precision any Riemann integral we are given, even if all we have are experimental data points. One of the main advantages of the variation integral is, as we have shown, it shares this property.

In closing, we will present two examples of the variation integral. We will give the convexly additive set functions μ and ν, which will integrate to give the functionals *T* and

D, which give the value of the function and the value of its derivative at the point $x = \frac{1}{2}$, respectively.

$$\mu((a,b]) = \begin{cases} 1, & b \le \frac{1}{2} \\ \dfrac{\frac{1}{2} - a}{b - a}, & a < \frac{1}{2} < b \\ 0, & a \ge \frac{1}{2} \end{cases}$$

Clearly, this will give a linear functional on $AC\,[\,0,\ 1\,]$, since

$$\begin{aligned} T(pf_\sigma + pg_\sigma) &= \sum_\sigma (\Delta_i f + \Delta_i g)\mu((x_{i-1}, x_i]) \\ &= \sum_{i=1}^{x_i < \frac{1}{2}} (\Delta_i f + \Delta_i g) \cdot 1 + \left[(f(b) + g(b)) - (f(a) + g(a))\right] \frac{\frac{1}{2} - a}{b - a} + 0 \\ &= f(a) + g(a) - f(0) - g(0) + \left[(f(b) + g(b)) - (f(a) + g(a))\right] \frac{\frac{1}{2} - a}{b - a} \\ &= f(a) + (f(b) - f(a)) \frac{\frac{1}{2} - a}{b - a} + g(a) + (g(b) - g(a)) \frac{\frac{1}{2} - a}{b - a} \\ &= T(pf_\sigma) + T(pg_\sigma). \end{aligned}$$

Where a is the last element of σ less than or equal to $\frac{1}{2}$, and b is the very next point in the partition.

Since every absolutely continuous function is the limit of pf_σ, the above establishes that the functional is linear. We can similarly show that T is linear under scalar multiplication. If we observe the way the sum in the integral collapsed we will note

$$T(pf_\sigma) = f(a) + \left(\frac{\frac{1}{2} - a}{b - a} \right) (x - a).$$

But this is the equation of the chord between $(a, f(a))$ and $(b, f(b))$. Since f is (absolutely) continuous, and a and b straddle one half, the continuity of f forces the value of the functional to $f\left(\frac{1}{2}\right)$ in the limit.

We next will show that the measure μ is indeed convex. That is that

$$\lambda_1 \mu((a,b]) + \lambda_2 \mu((b,c]) = \mu((a,c]).$$

For the case where a is greater than one half, and c is less than one half, convexity is very easy to establish. Thus, the main cases we are concerned with are when one half falls between a and b or between b and c. The algebraic calculations are very similar in both cases. So we will suppose the latter case. Notice we are still using the same lambdas from earlier in the chapter.

$$\begin{gathered} \lambda_1 \mu((a,b]) + \lambda_2 \mu((b,c]) = \\ \frac{b-a}{c-a} \mu((a,b]) + \frac{c-b}{c-a} \mu((b,c]) = \\ \frac{b-a}{c-a} \cdot 1 + \frac{c-b}{c-a} \left(\frac{\frac{1}{2} - b}{c-b} \right) = \\ \frac{b-a}{c-a} + \left(\frac{\frac{1}{2} - b}{c-a} \right) = \left(\frac{\frac{1}{2} - a}{c-a} \right) = \mu((a,c]) \end{gathered}$$

We conclude this functional, by looking at a calculation of a specific example of the variation integral. We will evaluate $F(x^2)$ using the partition $\sigma = \{0, \frac{1}{3}, \frac{2}{3}, 1\}$. Notice the third sub-interval will not contribute to the integral as $\mu\left(\left(\frac{2}{3},1\right]\right)=0$. So our approximation to the integral is

$$\nu\int_0^1 x^2 d\mu \approx \left[\left(\tfrac{1}{3}^2 - 0^2\right)\right]\mu\left(\left(0,\tfrac{1}{3}\right]\right) + \left[\left(\tfrac{2}{3}^2 - \tfrac{1}{3}^2\right)\right]\mu\left(\left(\tfrac{1}{3},\tfrac{2}{3}\right]\right)$$
$$= \frac{1}{9} + \frac{1}{6} = \frac{5}{18}.$$

If we had taken a finer partition and computed the integral again, we would have come closer to the functional's value, which is one fourth.

Our final example is the functional D, the (right hand) derivative evaluated at the point $x = \frac{1}{2}$. The convexly additive measure we will use here is

$$\nu\left((a,b]\right) = \begin{cases} 0, & b \le \dfrac{1}{2} \\ \dfrac{1}{b-a}, & a \le \dfrac{1}{2} < b \\ 0, & a > \dfrac{1}{2} \end{cases}$$

As the (right hand) derivative is not defined for all absolutely continuous functions, the functional cannot be linear and bounded. However, we can still demonstrate that this measure gives the point evaluation of the derivative, and that the measure is convexly additive.

We let a be the last element of the partition to be less than or equal to $x = \frac{1}{2}$ again and let b be the next point in the partition, as in the previous example. We then apply D to pf_σ. The only non-zero term in the sum will be the one from the interval straddling the point $x = \frac{1}{2}$. Thus we get

$$D(pf_\sigma) = \left(f(b) - f(a)\right)\left(\frac{1}{b-a}\right) = \frac{f(b)-f(a)}{b-a}$$

Which is the standard difference quotient, and in the limit becomes the right hand derivative of f at $x = \frac{1}{2}$.

We next consider convexity. As ν is zero except on the interval containing one half, the cases with $c \le \frac{1}{2}$ and $a > \frac{1}{2}$ are trivial. We will consider the case therefore where $a \le \frac{1}{2} < b$. The other case, where $b \le \frac{1}{2} < c$ is almost identical.

$$\begin{gathered}
\lambda_1 \nu((a,b]) + \lambda_2 \nu((b,c]) = \\
\left(\frac{b-a}{c-a}\right)\frac{1}{b-a} + \frac{c-b}{c-a}\cdot 0 = \\
\frac{1}{c-a} = \nu((a,c])
\end{gathered}$$

This shows that the function is convex. Finally we give a numerical example of our integration. We again use the x squared function. We will use a slightly different partition though, $\sigma = \left\{0, \frac{1}{3}, \frac{3}{5}, 1\right\}$. When we expand our sum, only the middle term will remain

$$\begin{aligned}
D(pf_\sigma) &= \left[\left(\frac{1}{3}\right)^2 - 0\right]\cdot 0 + \left[\left(\frac{3}{5}\right)^2 - \left(\frac{1}{3}\right)^2\right]\left(\frac{1}{\frac{3}{5}-\frac{1}{3}}\right) + \left[1 - \left(\frac{3}{5}\right)^2\right]\cdot 0 \\
&= \left[\left(\frac{3}{5}\right)^2 - \left(\frac{1}{3}\right)^2\right]\left(\frac{1}{\frac{3}{5}-\frac{1}{3}}\right) = \frac{14}{15}.
\end{aligned}$$

This number is very close to the true derivative at one half, which is one. As the partition gets finer, the answer will come closer and closer to one.

BIBLIOGRAPHY

H. L. Royden, *Real Analysis, 3rd Edition,* Macmillan, New York, 1988

W. Rudin, *Principles of Mathematical Analysis, 3rd Edition,* McGraw-Hill, New York 1976

J. R. Edwards and S. G. Wayment, *Representations for Transformations Continuous in the BV Norm,* Transactions of the American Mathematical Society, **154,** 1971 (251 – 265)

VITA

Gerard Leonard Ornas, Jr. (a.k.a. Lenny) was born in New Braunfels, Texas on 26 February 1969, the son of then SSgt (now SMsgt, retired) Gerard Leonard Ornas and Patricia Ann Ornas (formerly Zunker). After completing high school at Aviano American High School, Aviano Italy on 12 June 1987, Lenny began his undergraduate studies. After several years of study interrupted once, Lenny received his Bachelor of Science in Physics from the University of Texas at Austin in May of 1995. Seven months later he received his Bachelor of Science in Mathematics, also from the University of Texas at Austin. Six weeks after this he began his studies for his Master of Science in Mathematics at Southwest Texas State University, San Marcos, Texas. In September of 1996 Lenny was hired by the Department of Mathematics at SWTSU as a Graduate Teaching Assistant.

Permanent Address: 162 Sunfire Trail
Seguin, Texas, 78155

This thesis was typed by Lenny Ornas.