Basic Analysis on Countable Sets

Jeffrey Lassahn

A THESIS
Presented to the Department of Mathematics
of the University of Oregon
June 8, 1994

Contents

Acknowledgements

I would like to take this opportunity to thank some of the many people who helped this thesis get written, including Professor Dick Koch, who suggested and managed the whole project, Professors Peter Gilkey and Bill Kantor who read it, thought about it and provided suggestions and commentary, Professor Kenneth Ross for $\mathbf{Q} \times \mathbf{Q}$ and Alex West for the constant stream of advice and counterexamples.

1 Introduction

It is usual in an introduction to analysis to begin with a discussion of the topology and completeness properties of the real numbers. This set is presented as the foundation for all further work, and ideas such as continuity and convergence which are important to all of analysis depend critically on the details its properties. As a result, while the real numbers are originally introduced as one set of numbers among many, and while subjects such as algebra routinely explore properties of sets such as the rational numbers, the algebraic numbers, and other more specialized objects, analysis is usually done only on a few specific sets; if work is being done in one dimension, functions are assumed to be between sets of real numbers, while in two or more dimensions the choices expand to allow $\mathbf{R}^n$ or the complex numbers or some combination of these. Other possibilities are dismissed early on. The notion that analysis cannot be done on the rational numbers because they are not complete is generally presented, accepted and forgotten within the first few days of the first introductory class taught on the subject.

The reasons for this are not unconvincing, of course. Limits and convergent sequences are some of the most powerful tools of analysis, and their properties are much clearer in complete spaces. Notions such as compactness make no sense at all on sets like $\mathbf{Q}$ except under conditions so restrictive that their power as analytical tools is eliminated. The whole modern notion of the real numbers was created essentially so that the methods of analysis would have a space to work in.

Although the real numbers provide a convenient place to do much of analysis, they come with their own collection of problems. Their algebraic structure is far from simple; while some sets like the integers or the rationals can be described easily and completely by their algebraic properties, the reals are difficult to characterize. They are not the closure of any finite set under any binary operations. There cannot even be any notation which can be used to write out an arbitrary real number, since such a notation would map each number to a finite string from some alphabet and the set of all such strings is countable.

The uncountability of the real numbers produces other problems as well. Even operations which seem trivial on finite or countable sets can become the basis for deep, even inherently undecidable questions for uncountable sets. Can all infinite subsets of $\mathbf{R}$ be put into one-to-one correspondence with either $\mathbf{Z}$ or $\mathbf{R}$? No set can be found which won't do this, but is that the same as saying there are none? This can not be proven. Several important results in measure theory apply to objects which only exist if the axiom of choice is assumed. While analysis becomes definite about what domains its functions should take it becomes ambiguous about questions of what sort of logic is acceptable for deriving results.

I would like to propose that for some purposes the logical definiteness gained by working in countable sets may outweigh the loss of the standard methods from analysis, and that it is a reasonable and interesting problem to look for different methods which produce analogous results on countable spaces to those given by standard analysis and topology on uncountable spaces. In particular, completeness and connectedness are usually only properties of uncountable spaces. Certainly countable spaces like the integers with the usual metric are complete, but in an uninteresting way. There are also sets smaller than the real numbers which can be made connected, including such things as the single point, and the indiscrete topology on any countable set, but again these seem to lack some desirable structure. It would be nice to have analogous concepts which say useful things about countable metric spaces.

This thesis actually has two related goals. One is to develop some simple analogies to standard analytic concepts which can be used on incomplete sets. In particular, it is partially an attempt to give some description of how completeness is actually related to other properties of a space, and to remove this concept from ideas which can be developed without it. This will among other things give an examination of continuity which separates the notion of total boundedness (which is the generalization of boundedness from finite dimensional metric spaces to arbitrary metric spaces) from that of completeness even though the two are usually intertwined in the hybrid idea of compactness. It will also examine how much of basic calculus can be reconstructed without the intermediate value and extreme value theorems which come directly from the properties of completeness.

In addition to looking for extensions of analytic concepts to more general domains, I am also attempting to give some sort of impression of what calculus might look like if the real numbers were never considered a valid concept at all. Many of the definitions and results given in the text have logically equivalent forms which depend explicitly on the properties of the real numbers. For instance, the notion of strong continuity on a dense subset of $\mathbf{R}$, which I define in Chapter 2 in terms of distances between sets, can be shown to have the equivalent formulation: f is strongly continuous if and only if it is the restriction of a continuous function F defined everywhere on $\mathbf{R}$. This is not the definition I choose to work with, even though it makes the demonstration of certain results trivial. Of course strongly continuous functions map bounded sets to bounded sets, since the domain of the function can be extended to a compact set, which maps to another compact set because of a theorem from standard analysis. This is a perfectly reasonable proof of Theorem 2.5, but it is unenlightening. In general, of the proofs given the only ones which depend on the properties of the real numbers are those such as Theorems 2.3, 3.1 and 4.3 which explicitly describe the relationship between the concepts presented here and those of standard analysis. In other cases, even when $\mathbf{R}$ appears explicitly in the theorem I am often not depending on its completeness or cardinality. In particular some theorems use the terminology $X \subseteq \mathbf{R}$ to specify the domain X of some function. Unless specific facts about X are given, it can be assumed to be arbitrary, and the statement is simply a short way of specifying any set such as $\mathbf{Q}$ which has the ordering and some arithmetic properties of $\mathbf{R}$.

Either one of these goals extends into large enough areas of mathematics that this thesis can only address a few preliminary points, and must ignore many more. In particular, the question of what can be done to topology to produce something more consistent with this version of analysis is almost totally ignored. The chapters have also been separated by interludes which are intended to informally present extensions to ideas in the main text which are generally conjectural. A major theme of the interludes is that when the real numbers have been removed from consideration there are many different possible sets which can take their place. None of these have all of the characteristics of $\mathbf{R}$ and each has its own character. The exploration and construction of countable sets which have different analytic properties seems to me to be an interesting project on its own merits and these interludes are my way of encouraging others to play with the possibilities.

Some Unusual Countable Sets

In the introduction I made some fuss about difficulties arising from the formal manipulation of uncountable sets. There are many examples of problems both logical and conceptual which these sets present, not the least of which arise from the sheer difficulty of describing such objects. As previously mentioned, any notation consisting of finite strings of characters from a finite alphabet can distinguish only a countable number of objects. This includes essentially all reasonable forms of communication; even pictures can be included as long as they are not expected to be examined at infinitely fine detail. There simply is no way to write down an arbitrary real number.

By similar reasoning, any objects which could reasonably be called algorithms can only be part of a set which is at most countable. There cannot be an algorithm which approximates each real number. For that matter, the number of proofs is countable, as is the number of possible conjectures; there cannot be a theorem, or even a conjecture specific to each real number.

These facts suggest a picture of the real numbers as a set with a small minority of its members available for examination and use, while the majority sit out of reach. It is even possible to examine the set of numbers whose properties are actually available for inspection. Perhaps this can be described as the set of all expressions which when evaluated over the positive integers produce convergent sequences, or the set of propositions which can be evaluated for every rational number and for which the rationals where they are true form a Dedekind cut. Neither of these sets consists of all of $\mathbf{R}$, since both the number of propositions and the number of expressions are countable. It is not clear what other properties they have, or even if they are actually the same set.

For that matter, it is not immediately clear that such things are useful at all. Why bother defining a set by merely observing that the notation we usually use only produces a subset of the objects we try to think about, if in practice this observation has no consequences for the way calculation is carried out? A set with membership determined by a rule such as "if you can write it down it's in the set" would seem to be so nebulous that it couldn't possibly be useful.

The Paradox of Algorithms

There are, however, interesting results which can be discovered with the aid of such sets. For instance, consider the set of all decimal expansions which can be produced by algorithms. That is, suppose there is a procedure which can be described by a finite set of instructions which when given a positive integer n returns after a finite number of calculations the nth digit of some real number between 0 and 1. The set of these algorithms must be countable, since they can be listed by a procedure like alphabetization.

Suppose such a list exists and contains all possible algorithms from this set. A real number can be approximated by a decimal expansion which has its nth digit constructed from the nth digit of the number d produced by the nth algorithm in the list by making

$$s_n = \begin{cases} 3 & \text{if } d \neq 3; \\ 2 & \text{if } d = 3. \end{cases}$$

The number which s_n converges to will be different than that produced by any algorithm in the list. But isn't the procedure just described the outline of an algorithm? Shouldn't it already be on the list? If it is then the number produced by it would be the same as the number produced by some algorithm on the list, and if it is not then any listing of algorithms must not contain all possible ones, so the set of these algorithms can not be countable. There appears to be a contradiction here.

One way to resolve this problem as it stands is to observe that what was presented was not an algorithm of finite length. In order to do the calculation we must have access to every algorithm in the list, and if this infinite list must be included to specify what the algorithm does, its description is no longer finite. Perhaps this requirement for infinite length can be avoided by having the algorithm we are constructing generate the list of algorithms internally. Each algorithm is a string in some alphabet of symbols, so our new algorithm can simply start producing strings in order of increasing length, examine each one to see if it is the appropriate type of algorithm, then calculate based on this information. By this method, it can give any number of digits in its decimal expansion after only finding a finite number of other algorithms and evaluating each one for a single value of n, so it should still terminate, and it now has a finite description (programmers of the language LISP might even consider implementing such an algorithm). It would appear that our paradox has returned.

One plausible conclusion to be made from this is that I am defining much too loosely what counts as a computational operation. Is it reasonable to expect an algorithmic process to evaluate the computations of another algorithmic process? In the field of computer science, such algorithms are used all the time. The entire concept behind a programmable computer is that it is an implementation of a formal algorithm which processes other algorithms. The more doubtful question is whether there is an algorithmic method for recognizing other algorithms. If there is no strictly formal analysis of text strings which can determine whether they will behave in some predictable way when executed as algorithms, then the proposed method cannot be guaranteed to complete its calculation successfully and the paradox is resolved.

This conclusion corresponds nicely to a result from the theory of computer science that states that no algorithm can exist which can evaluate other algorithms and reliably determine whether they terminate. It is usually demonstrated by entirely different methods that are proposed here, but it is perhaps surprising to see how such a result can arise from questions about cardinality.

Suggestions of an Incompleteness Theorem

Another line of argument with similar form uses the fact that the number of real numbers is larger than the number of theorems about them. Suppose that there is a formal system which allows proofs of the correctness of mathematical calculation (choose your favorite one, like the axioms of set theory combined with the predicate calculus). It is possible to assemble a list of pairs of strings in this formal system, the first of which is an expression which returns a decimal digit for each positive integer, and the second of which is a proof that the first actually does this. These can be found by a strictly formal examination of all strings; simply check that the second string is a series of statements which can be derived from previous statements and the first string by the transformation rules of the system, and that the concluding statement is of the form $<$ first string $>\in \{1 \ldots 9\}$ for all $n \in \mathbf{N}$ however this is best expressed in the particular formal system.

If the formal system is powerful enough that it can represent the logic of Cantor's diagonal proof, and the typographical processes used in its rules of inference can be represented as operations on sets in the system, then a construction which evaluates all strings for the above properties and composes the expressions it finds with the function

$$f(x) = \begin{cases} 3 & \text{if } x \neq 3; \\ 2 & \text{if } x = 3. \end{cases}$$

would constitute an expression which returns a decimal digit for each positive integer, but this fact has no proof in the formal system.

Again, results of this sort have been proven by arguments much different from the ones briefly sketched here, and actually showing that this argument leads to valid conclusions requires much more careful work, but these examples are suggestive of a method of reasoning which uses the countability of constructable subsets of uncountable sets to help discover surprising results.

2 Continuity

Continuity and Strong Continuity

The concept of a continuous function is useful in many mathematical contexts. This has led to a large number of characterizations of continuity, all of which are equivalent, but each emphasizes different features of the functions which are of interest in different circumstances.

Definition 2.1 *For a function $f : X \to Y$ between metric spaces X and Y the following are equivalent and any can be taken as the definition for continuity of the function f.*

(a) For any $A \subseteq X$, $f(Cls(A)) \subseteq Cls(f(A))$.

(b) For any open set $B \subseteq Y$, $f^{-1}(B)$ is open in X.

(c) For any closed set $B \subseteq Y$, $f^{-1}(B)$ is closed in X.

(d) For any $x \in X$ and $\epsilon > 0$, there is some $\delta > 0$ such that $d(x, z) < \delta$ implies $d(f(x), f(z)) < \epsilon$.

(e) For any sequence a_n which converges to $a \in X$, the sequence $f(a_n)$ converges to $f(a)$.

These characterizations were designed to correspond to the intuitive idea of a continuous function on a complete metric space such as the real numbers, but give somewhat questionable results on metric spaces which are not complete.

Example 2.1 *A function $f : \mathbf{Q} \to \mathbf{Q}$ is defined as follows:*

$$f(x) = \begin{cases} 1 & \text{if } x^2 < 2; \\ 0 & \text{if } x^2 > 2. \end{cases}$$

This function is continuous on $\mathbf{Q}$.

A definition of continuity which allows this sort of function to be continuous is not going to be particularly helpful for doing work on $\mathbf{Q}$. In general, these definitions allow continuous functions to have steplike changes in value anywhere that their domains are not complete. Since many interesting countable subsets of $\mathbf{R}^n$ including $\mathbf{Q}$ are not complete *anywhere*, any attempt to extend notions of analysis to countable sets must begin with a rethinking of continuity.

One property which continuous functions should have is that pieces of their domains which touch should have images which touch. This notion can be formalized. A is called **disconnected** from B if $Cls(A) \cap B = \emptyset$ and $A \cap Cls(B) = \emptyset$.

Two sets are **connected** to each other if they are not disconnected. This does not assume anything about the individual sets; in particular, two sets can be connected to each other even when each one can be divided into parts which are disconnected.

This gives a new equivalent characterization for continuity.

Theorem 2.1 *A function between metric spaces $f : X \to Y$ is continuous if and only if for any $A \subseteq X$ connected to $B \subseteq X$, $f(A)$ is connected to $f(B)$ in Y.*

Proof. Choose any $A \subseteq X$ and $B \subseteq X$ such that A is connected to B. This means either $Cls(A) \cap B \neq \emptyset$ or $A \cap Cls(B) \neq \emptyset$. Without loss of generality I can assume $Cls(A) \cap B \neq \emptyset$. If f is continuous, $f(Cls(A)) \subseteq Cls(f(A))$, so in particular $f(Cls(A) \cap B) \subseteq Cls(f(A))$. This means that $Cls(f(A))$ contains some element of $f(B)$, so $f(A)$ is connected to $f(B)$.

Conversely, if f is not continuous, there must exist some $A \subseteq X$ such that $f(Cls(A)) \nsubseteq Cls(f(A))$. There must be some point $b \in Cls(A)$, $f(b) \notin Cls(f(A))$. A is connected to b, since $b \in Cls(A)$, but $f(A)$ is disconnected from $f(b)$, because $f(b) \notin Cls(f(A))$ by hypothesis and $Cls(f(b)) = f(b) \notin f(A)$. ∎

Of course, since this characterization of continuity is equivalent to the others, it will allow the same functions (such as Example 2.1) to be continuous. It does, however, make clearer why such functions are continuous. In particular, the function in Example 2.1 is continuous around $\sqrt{2}$ even though its image is the disconnected sets $\{0\}$ and $\{1\}$ because the set $(-\infty, \sqrt{2}]$ is disconnected from $[\sqrt{2}, \infty)$ in $\mathbf{Q}$. Since countable metric spaces tend to be disconnected everywhere, they tend to impose only weak conditions on their continuous functions. Perhaps if connectedness in this characterization of continuity is replaced with a slightly weaker notion, a natural analogue to continuity will appear which is more useful on countable spaces.

Definition 2.2 *Two subsets A and B of a metric space are said to be* **close** *if and only if there are nonempty totally bounded sets $\alpha \subseteq A$ and $\beta \subseteq B$ such that $d(\alpha, \beta) = 0$.*

The requirement that the sets α and β be totally bounded prevents sets from being close if their distance only goes to zero at infinity. For example the sets $\{(x, 0)\}$ and $\{(x, y) : y = 1/x\}$ in $\mathbf{R}^2$ are not close, even though the distance between them is zero.

A criterion similar to continuity can now be constructed.

Definition 2.3 *A function between metric spaces $f : X \to Y$ is* **strongly continuous** *if and only if for any two sets $A \subseteq X$ and $B \subseteq X$ which are close, $f(A)$ and $f(B)$ are close in Y.*

This new notion of strong continuity has some striking similarities to continuity of the usual sort. In particular, the two are equivalent on complete metric spaces, such as $\mathbf{R}^n$. On spaces which aren't complete, such as $\mathbf{Q}^n$ or open subsets of $\mathbf{R}^n$, strong continuity is a more restrictive condition than continuity; some properties which continuous functions preserve on complete spaces are preserved on all metric spaces by strong continuity. Under strong continuity completeness can often be thought of as an incidental property of a space instead of being inextricably intertwined with many other characteristics of the space. This conceptual separation is in large part a result of the replacement of connectedness, which is strongly affected by completeness, by closeness, which is not. Unfortunately, closeness and connectedness are not equivalent even in the case where spaces are assumed to be complete, (for example $[-1, 0)$ is close to $(0, 1]$ in $\mathbf{R}$,) but there is a particularly simple characterization for closeness in complete spaces.

Theorem 2.2 *Two subsets of a complete metric space X are close if and only if their closures intersect.*

Proof. If A and B are close in X, they have totally bounded subsets α and β such that $d(\alpha, \beta) = 0$. This means there are sequences $a_n \in \alpha$ and $b_n \in \beta$ such that $\lim_{n\to\infty} d(a_n, b_n) = 0$. Since compact is equivalent to complete and totally bounded, $Cls(\alpha)$ is compact. This means the sequence a_n has a convergent subsequence, a_k. $Cls(\beta)$ is also compact, so the subsequence b_k has a convergent subsubsequence, b_l. Let $a = \lim_{l\to\infty} a_l$ and $b = \lim_{l\to\infty} b_l$. For any $\epsilon > 0$, l can be made large enough so that $d(a_l, a) < \epsilon$, $d(b_l, b) < \epsilon$, and $d(a_l, b_l) < \epsilon$ simultaneously. This implies $d(a, b) < 3\epsilon$, for all $\epsilon > 0$, so it must be true that $a = b$. But $a \in Cls(A)$ and $b \in Cls(B)$, so $Cls(A) \cap Cls(B) \neq \emptyset$.

Conversely, if there is some $x \in Cls(A) \cap Cls(B)$, then there are sequences $a_n \in A$ and $b_n \in B$ so that $\lim_{n\to\infty} a_n = x = lim_{n\to\infty} b_n$. All convergent sequences are totally bounded, so $\alpha = \{a_n\}$ and $\beta = \{b_n\}$ are totally bounded subsets of A and B and $d(\alpha, \beta) \leq d(\alpha, x) + d(x, \beta) = 0$, so A and B must be close. ∎

Even if closeness and connectedness are not equivalent on complete spaces, they are comparable in some weaker sense. Connectedness is preserved by continuous functions, and closeness is preserved by strongly continuous functions. In complete spaces, the two types of continuity are equivalent, so any function that preserves connectedness preserves closeness.

Theorem 2.3 *If X and Y are complete metric spaces, a function $f : X \to Y$ is strongly continuous if and only if it is continuous.*

Proof. Choose any two sets A and B in X which are close. The closures of A and B must intersect. This means there must be some $x \in Cls(A) \cap Cls(B)$. If f is continuous, $f(x) \in Cls(f(A))$ and $f(x) \in Cls(f(B))$. Thus $Cls(f(A)) \subseteq Y$ and $Cls(f(B)) \subseteq Y$ intersect, so f must be strongly continuous.

Suppose $f : X \to Y$ is not continuous. There is some $A \subseteq X$ such that $f(Cls(A)) \nsubseteq Cls(f(A))$. A point b can be chosen so that $b \in Cls(A)$, but $f(b) \notin Cls(f(A))$. $Cls(b) \cap Cls(A) = b$, so b is close to A, but $Cls(f(b)) \cap Cls(f(A)) = \emptyset$ so $f(b)$ is not close to $f(A)$ and f cannot be strongly continuous. ∎

Properties of Strong Continuity

In cases where strong continuity is not equivalent to continuity, strongly continuous functions may still preserve interesting properties of their domains. In fact, as the name suggests, strong continuity implies continuity on any metric space, so every property of a metric space which is preserved by continuous functions is also preserved by strongly continuous functions.

Theorem 2.4 *If $f : X \to Y$ is strongly continuous, it is continuous.*

Proof. Suppose f is not continuous. There must be some $A \subseteq X$ so that $f(Cls(A)) \nsubseteq Cls(f(A))$. There must be some $a_0 \in Cls(A)$ such that $f(a_0) \notin Cls(f(A))$. Since the closure of a set consists of all points with distance zero from it, $d(f(a_0), f(A)) > 0$, but $d(a_0, A) = 0$. This means that a_0 is close to A, but $f(a_0)$ is not close to $f(A)$, so f is not strongly continuous. ∎

Strongly continuous functions must, therefore, map connected sets to connected sets, compact sets to compact sets, and so forth. It should not be surprising that strongly continuous functions also preserve properties which continuous functions in general do not. Closeness is one obvious example. The images of close sets under continuous functions are in general not close unless the domain of the function is complete. Compactness, which in metric spaces is simply the combination of completeness and total boundedness, is preserved by all continuous functions, but strongly continuous functions preserves total boundedness as a separate property.

Theorem 2.5 *If A is totally bounded and $f : A \to B$ is strongly continuous, then $f(A)$ is totally bounded.*

Proof. Suppose $f(A)$ contains a sequence b_n and there is an $\epsilon > 0$ so that $d(b_i, b_j) > \epsilon$ for all i and j, and f is strongly continuous. There must be some sequence $a_n \in f^{-1}(b_n)$ and some $\epsilon' > 0$ so that $d(a_i, a_j) > \epsilon'$ for all i and j. Otherwise, there would be some two subsets of a_n which were close, while all pairs of subsets of b_n would be separated by at least ϵ and not be close.

But a space cannot have a sequence a_n where $d(a_i, a_j) > \epsilon'$ if it can be covered by finitely many balls of radius $\epsilon'/2$ or less, because each of these balls can contain at most one member of such a sequence, and there are infinitely many members of the sequence. Thus, if the sequence b_n exists, A cannot be covered by a finite number of balls with radius less than $\epsilon'/2$, so A cannot be totally bounded.

If $f(A)$ is not totally bounded, it must have a sequence b_n such that $d(b_i, b_j) > \epsilon$ for all i and j. This sequence can be constructed by taking ϵ small enough that $f(A)$ cannot be covered by balls of radius ϵ, then choosing an arbitrary $b_1 \in f(A)$. The ball of radius ϵ around b_1 does not cover $f(A)$, so a b_2 can be chosen outside of it. If there is a partial sequence $b_1 \dots b_n$ such that $d(b_i, b_j) > \epsilon$ for all $i, j \le n$, then the union of balls of radius ϵ centered on the members of $b_1 \dots b_n$ must not cover all of $f(A)$, so a point outside of this union can be chosen as b_{n+1}. This point must have distance at least ϵ from all of the $b_1 \dots b_n$, so there is now a larger partial sequence $b_1 \dots b_{n+1}$ such that $d(b_i, b_j) > \epsilon$ for all $i, j \le n + 1$. By induction, there must be an entire sequence with this property.

Since any B which is not totally bounded allows the creation of a sequence b_n, which in turn forces A not to be totally bounded, the image of a totally bounded set under f must be totally bounded. ■

Completeness, however, is not preserved independently of total boundedness. For example, there are strongly continuous functions $f : \mathbf{R} \to (0, 1)$.

I would now like to develop some tools for working with strongly continuous functions, including some basic properties of sums and products and such, as well as the behavior of sequences and their images. It will be useful to first examine set products. Unlike continuity which depends only on the topology imposed on a space, strong continuity depends on the metric, so the question of what metrics should be used for product spaces becomes important. There are several obvious and useful metrics to place on the product of metric spaces. If X is the set product of metric spaces $X_1 \dots X_n$, with each X_i having its own metric d_i, the metric d_X on X could be chosen as the Euclidian one $d_X(p, q) = \sqrt{\sum d_i(p_i, q_i)^2}$, or as $d_X(p, q) = \max_{all\ i}(d_i(p_i, q_i))$, or $d_X(p, q) = \sum d_i(p_i, q_i)$, all of which are useful under different circumstances. Because of the inequality

$$\max_{all\ i}(d_i(p_i,q_i)) \leq (\sum d_i(p_i,q_i)^k)^{1/k} \leq \cdots \leq \sum d_i(p_i,q_i) \leq n\cdot\max_{all\ i}(d_i(p_i,q_i))$$

where $i = 1,\ldots,n$, if any of these distance functions makes two sets close the others must as well. Two metrics which specify the same sets as close must specify the same functions as strongly continuous, so results about strong continuity proved for one of these metrics hold for each of the others. I will assume unless stated otherwise that product spaces have one of these metrics. Another fact which is used in the following proof is that if two sequences have a distance which tends to zero in the limit, their images under a strongly continuous function must as well, since not only must those two sequences have images which are close, but all of their subsequences must as well.

Theorem 2.6 *If $f : A \to C$ and $g : B \to D$ are strongly continuous, the function $h : A \times B \to C \times D$ defined by $h(x,y) = (f(x), g(y))$ is strongly continuous.*

Proof. Choose two close sets $X \subseteq A \times B$ and $Y \subseteq A \times B$. There must be some sequences $x_n = (a_{x_n}, b_{x_n}) \in X$ and $y_n = (a_{y_n}, b_{y_n}) \in Y$ such that x_n and y_n are both totally bounded and

$$\lim_{n\to\infty} \max(d(a_{x_n}, a_{y_n}), d(b_{x_n}, b_{y_n})) = 0.$$

This means that $\lim_{n\to\infty} d(a_{x_n}, a_{y_n}) = 0$ and $\lim_{n\to\infty} d(b_{x_n}, b_{y_n}) = 0.$

Since f and g are strongly continuous,

$$\lim_{n\to\infty} d(f(a_{x_n}), f(a_{y_n})) = 0 \quad \text{and} \quad \lim_{n\to\infty} d(g(b_{x_n}), g(b_{y_n})) = 0, \quad \text{so}$$

$$\lim_{n\to\infty} \max(d(f(a_{x_n}), f(a_{y_n})), d(g(b_{x_n}), g(b_{y_n}))) = 0.$$

But $(f(a_{x_n}), g(b_{x_n})) = h(x_n)$ and $(f(a_{y_n}), g(b_{y_n})) = h(y_n)$ so $h(x_n)$ and $h(y_n)$ must have distance zero which makes $h(X)$ close to $h(Y)$. ■

This result is particularly useful for working with binary operations on strongly continuous functions. Some calculations of this type are used to prove the following list of useful facts about strong continuity.

Theorem 2.7 *If A, B and C are arbitrary metric spaces, and f and g are strongly continuous, the following are true.*

(a) If $f : A \to B$ is constant, it is strongly continuous.

(b) If $f : A \to B$ and $g : B \to C$ then $h : A \to C = g \circ f$ is strongly continuous.

(c) If $f : A \to A$ is the identity map, it is strongly continuous.

If A is an arbitrary metric space, B is a vector space, and $f : A \to B$ and $g : A \to B$ are strongly continuous, the following are true.

(d) For any scalar c, $c\cdot f$ is strongly continuous.

(e) $f + g$ is strongly continuous.

(f) $f - g$ is strongly continuous.

If A is an arbitrary metric space, B is an arbitrary subfield of the complex numbers, and $f : A \to B$ and $g : A \to B$ are strongly continuous, the following are true.

(g) $f{\cdot}g$ is strongly continuous.

(h) $1/f$ is strongly continuous if for some $\epsilon > 0$ and every $a \in A$, $d(f(a), 0) > \epsilon$.

(i) f/g is strongly continuous if for some $\epsilon > 0$ and every $a \in A$, $d(g(a), 0) > \epsilon$.

Proof.

(a) If f is constant, its image in B is the single point, b. Any two nonempty subsets of b are close, so the image of close sets under f must be close.

(b) For two arbitrary close sets $\alpha \subseteq A$ and $\beta \subseteq A$, $f(\alpha)$ and $f(\beta)$ must be close, since f is strongly continuous. Since g is also strongly continuous, $g(f(\alpha))$ must be close to $g(f(\beta))$. This means that α close to β implies $h(\alpha)$ close to $h(\beta)$.

(c) For two arbitrary close sets, $\alpha \subseteq A$ and $\beta \subseteq A$, $f(\alpha) = \alpha$, so it is close to $f(\beta) = \beta$.

(d) $c{\cdot}f$ is $h \circ f$ where $h(x) = c{\cdot}x$. Because of part(b), showing that h is strongly continuous is sufficient.

Let $\alpha \subseteq B$ and $\beta \subseteq B$ be arbitrary close sets. There must be totally bounded sequences $a_n \in \alpha$ and $b_n \in \beta$ such that $\lim_{n\to\infty} d(a_n, b_n) = 0$. $d(h(a_n), h(b_n)) = d(c{\cdot}a_n, c{\cdot}b_n) = c{\cdot}d(a_n, b_n)$, so $\lim_{n\to\infty} d(h(a_n), h(b_n)) = 0$. $h(\alpha)$ must be close to $h(\beta)$, making h is strongly continuous.

To prove strong continuity for the binary operations begin by defining $h : A \times A \to B \times B$ (where B is a vector space in (e) and (f) but a field in (g)) so that $h(a_1, a_2) = (f(a_1), g(a_2))$. The functions from $B \times B$ to B $p(x, y) = x + y$ and $t(x, y) = x{\cdot}y$ can be constructed, and showing that p and t are strongly continuous is sufficient to show that $p \circ h = f + g$ and $t \circ h = f{\cdot}g$ are strongly continuous.

(e) Given two arbitrary close sets $\alpha \subseteq B \times B$ and $\beta \subseteq B \times B$, there must be totally bounded sequences $a_n = (x_{a_n}, y_{a_n}) \in \alpha$ and $b_n = (x_{b_n}, y_{b_n}) \in \beta$ so that $\lim_{n\to\infty} d(x_{a_n}, x_{b_n}) + d(y_{a_n}, y_{b_n}) = 0$.

$$d(x_{a_n}, x_{b_n}) + d(y_{a_n}, y_{b_n}) = \|x_{a_n} - x_{b_n}\| + \|y_{a_n} - y_{b_n}\| \geq \|(x_{a_n} - x_{b_n}) + (y_{a_n} - y_{b_n})\|$$
$$= \|(x_{a_n} + y_{a_n}) - (x_{b_n} + y_{b_n})\| = d(p(x_{a_n}, y_{a_n}), p(x_{b_n}, y_{b_n})).$$

Thus, $\lim_{n\to\infty} d(p(a_n), p(b_n)) = 0$ so $p(\alpha)$ is close to $p(\beta)$ and p is strongly continuous.

(f) $f - g = f + (-1){\cdot}g$. If g is strongly continuous, then $(-1){\cdot}g$ is as well by part (d), and if f is also strongly continuous $f + (-1){\cdot}g$ must be because of part (e).

(g) Given two arbitrary close sets $\alpha \subseteq B \times B$ and $\beta \subseteq B \times B$, there must be totally bounded sequences $a_n = (x_{a_n}, y_{a_n}) \in \alpha$ and $b_n = (x_{b_n}, y_{b_n}) \in \beta$ so that $\lim_{n\to\infty} \max(d(x_{a_n}, x_{b_n}), d(y_{a_n}, y_{b_n})) = 0$. This implies that $\lim_{n\to\infty} d(x_{a_n}, x_{b_n}) = 0$ and $\lim_{n\to\infty} d(y_{a_n}, y_{b_n}) = 0$.

$$\begin{aligned}
d(t(a_n), t(b_n)) &= \|x_{a_n} \cdot y_{a_n} - x_{b_n} \cdot y_{b_n}\| \\
&= \|(x_{a_n} - x_{b_n}){\cdot}(y_{a_n} - y_{b_n}) + x_{b_n}{\cdot}(y_{a_n} - y_{b_n}) + y_{b_n}{\cdot}(x_{a_n} - y_{b_n})\| \\
&\leq \|x_{a_n} - x_{b_n}\|{\cdot}\|y_{a_n} - y_{b_n}\| + \|x_{b_n}\|{\cdot}\|y_{a_n} - y_{b_n}\| + \|y_{b_n}\|{\cdot}\|x_{a_n} - x_{b_n}\| \\
&= d(x_{a_n}, x_{b_n}){\cdot}d(y_{a_n}, y_{b_n}) + \|x_{b_n}\|{\cdot}d(y_{a_n}, y_{b_n}) + \|y_{b_n}\|{\cdot}d(x_{a_n}, x_{b_n}).
\end{aligned}$$

Since $\lim_{n\to\infty} d(x_{a_n}, x_{b_n}) = 0$, $\lim_{n\to\infty} d(y_{a_n}, y_{b_n}) = 0$ and both x_{b_n} and y_{b_n} are bounded, $\lim_{n\to\infty} d(t(a_n), t(b_n)) = 0$. This makes $t(\alpha)$ close to $t(\beta)$, so t must be strongly continuous.

(h) Given two arbitrary close sets α and β, there must be two sequences $a_n \in \alpha$ and $b_n \in \beta$ which are both totally bounded and so that $\lim_{n\to\infty} d(a_n, b_n) = 0$.

$$d\left(\frac{1}{f(a_n)}, \frac{1}{f(b_n)}\right) = \left\|\frac{1}{f(a_n)} - \frac{1}{f(b_n)}\right\| = \left\|\frac{f(b_n) - f(a_n)}{f(b_n)\cdot f(a_n)}\right\|.$$

If $\|f(x)\| > \epsilon$ for all x,

$$d\left(\frac{1}{f(a_n)}, \frac{1}{f(b_n)}\right) \leq \frac{d(f(a_n), f(b_n))}{\epsilon^2}$$

so $\frac{1}{f(\alpha)}$ must be close to $\frac{1}{f(\beta)}$ and $1/f$ is strongly continuous.

(i) $f/g = f\cdot(1/g)$. If g is strongly continuous and $d(g, 0) > \epsilon$ for some $\epsilon > 0$, $1/g$ must be strongly continuous by (h), and so $f\cdot(1/g)$ is strongly continuous by (g). ∎

It is now easy to construct some examples of strongly continuous functions. For instance, all polynomials in **R**, **Q**, or **C** are strongly continuous, since polynomials are sums of terms constructed from products of the identity with itself and constants, and both sums and products preserve strong continuity.

The function $1/x$ in either $\mathbf{R} - \{0\}$ or $\mathbf{Q} - \{0\}$ is continuous but not strongly continuous. This can be easily seen by observing $1/x$ on any bounded region around the origin becomes arbitrarily large, while strongly continuous functions of bounded sets are bounded. $1/x$ is strongly continuous on any closed subset of $\mathbf{R} - \{0\}$, including all sets of the form $(-\infty, -\epsilon] \cup [\epsilon, \infty)$, however.

By similar reasoning, rational functions can be seen to be strongly continuous where they have no vertical asymptotes. This is true of functions like $x\cdot(x-1)/x$ even though its denominator is not bounded away from zero, because $x\cdot(x-1)/x = (x-1)$ on its entire domain and $(x-1)$ is strongly continuous on any subset of **R**. In general, a strongly continuous function is strongly continuous on any restriction of its domain, so functions like $f(x) = \sin(x)$, $x \in \mathbf{Q}$ are strongly continuous because the extension of f to **R** is strongly continuous, even though determining the strong continuity of f on **Q** directly is not easy.

The notion of a **strong homeomorphism**, meaning a bijective strongly continuous function with a strongly continuous inverse, is also possible. The unit circle, $\{(x, y) : x^2 + y^2 = 1\}$ and the square $\{(x, y) : |x| + |y| = 1\}$ are strongly homeomorphic by the map $f(x, y) = (x, y)\cdot \frac{|x|+|y|}{\sqrt{x^2+y^2}}$, for example, and in general every homeomorphism between two complete spaces is a strong homeomorphism. There are homeomorphisms which are not strong, however. In particular, total boundedness is invariant under strong homeomorphisms so any homeomorphism $h : \mathbf{R} \leftrightarrow (0, 1)$ cannot be strong.

Continuity at a Point and Uniform Continuity

The usual definitions of continuity are local, in the sense that they can be written so that each point in the domain of a function is tested independently for continuity. It is then possible to speak of functions which are continuous on parts of their domain, continuous everywhere but at a specific point, and so on. Ideally strong continuity would also have this ability, so it may be worthwhile to look for definitions of strong continuity at a specific point.

Definition 2.4 (proposed) *A function $f : A \to B$ is* **strongly continuous at a point** *$x \in A$ if and only if for any set $\alpha \subseteq A$ which is close to x, $f(\alpha)$ is close to $f(x)$.*

This resembles the definition of strong continuity, with the additional condition that one of the sets must be the point we are testing. Requiring that the set α be close to x, however, is just asking that $d(x, \alpha) = 0$, or equivalently that there is a sequence $a_n \in \alpha$ which converges to x. If f is strongly continuous at x, any subsequence of $f(a_n)$ must have distance zero from $f(x)$, which is true exactly when $f(a_n)$ converges to $f(x)$. This is just the definition for continuity at a point, so *our proposed definition of strong continuity at x is equivalent to the usual continuity at x.*

In fact there can be no local definition of strong continuity such that making f strongly continuous at every point forces f to be strongly continuous. In example 2.1, every point in the domain has an interval about it where f is constant, so any generalization of continuity at a point must either claim that some functions which are constant close to x are still not continuous at x, or claim that this function is continuous at every point. In essence, one thing that the definition of strong continuity gives up is the idea that any function which fails to be continuous fails at some specific point in the domain. I will still use the notion of continuity at a point occasionally, but with the implicit warning that continuity at every point implies continuity in the usual sense, but does not imply strong continuity. The term **pointwise continuous** can be used as a synonym for continuous to emphasize this distinction.

Independent of any mathematical difficulties related to the definitions of closeness and strong continuity, there is the more aesthetic question of why the requirement of total boundedness is included for close sets. This appears in some ways to be an artificial addition, and perhaps invites the question of whether results such as theorem 2.5 are at all significant mathematically or simply built intentionally into the definition. It is no difficult matter to make boundedness invariant by including requirements about boundedness explicitly in the definitions. In fact, theorem 2.5 does not use the boundedness of close sets in any important way, and is equally true when rewritten as

Theorem 2.8 *If A is totally bounded and $f : A \to B$ takes any two sets with distance zero from each other to images which also have distance zero from each other, then $f(A)$ is totally bounded.*

The proof is exactly analogous to that of theorem 2.5.

If total boundedness is not included in the definition specifically to make the images of bounded sets under strongly continuous functions bounded, why is it included? Most of the properties of strong continuity which were proven in the theorems so far are equally valid without the restriction of total boundedness (the exceptions are theorems 2.2 and 2.3, and theorem 2.7 part (g).) In fact, by excluding total boundedness from the definition of "close", yet another strong form of continuity can be defined, and it turns out to be exactly equivalent to uniform continuity.

Theorem 2.9 *A function $f : A \to B$ takes pairs of sets with distance zero to images with distance zero if and only if it is uniformly continuous.*

Proof. The function f takes pairs of sets with distance zero to images with distance zero if and only if for any two sequences $a_n \in A$ and $b_n \in B$ such that $\lim_{n\to\infty} d(a_n, b_n) = 0$ in A, $\lim_{n\to\infty} d(f(a_n), f(b_n)) = 0$ in B, because any pair of sets α and β have distance zero exactly when they contain sequences $a_n \in \alpha$ and $b_n \in \beta$ so that $\lim_{n\to\infty} d(a_n, b_n) = 0$.

Suppose f is not uniformly continuous. There is some $\epsilon > 0$ such that no $\delta > 0$ exists which forces $d(f(x), f(y)) < \epsilon$ for every x and y where $d(x, y) < \delta$. This means that for any $\delta > 0$ there are some a_1 and b_1 such that $d(a_1, b_1) < \delta$, but $d(f(a_1), f(b_1)) \geq \epsilon$. Since δ is arbitrary, there must also be a_n and b_n so $d(a_n, b_n) < \delta/n$ but $d(f(a_n), f(b_n)) \geq \epsilon$ for any n. Thus a_n and b_n form a pair of sequences where $\lim_{n\to\infty} d(a_n, b_n) = 0$ but $\lim_{n\to\infty} d(f(a_n), f(b_n)) \neq 0$.

Conversely, suppose there is some pair of sequences a_n and b_n so that $\lim_{n\to\infty} d(a_n, b_n) = 0$ but $\lim_{n\to\infty} d(f(a_n), f(b_n)) \neq 0$. There must be some $\epsilon > 0$ such that for any N there is an $n > N$ where $d(f(a_n), f(b_n)) \geq \epsilon$. For any $\delta > 0$ there must also be some N such that for any $n > N$, $d(a_n, b_n) < \delta$. This implies that there is no $\delta > 0$ such that $d(a_n, b_n) < \delta$ implies $d(f(a_n), f(b_n)) < \epsilon$, so f cannot be uniformly continuous. ∎

Strong continuity is now revealed to be weaker than uniform continuity and stronger than pointwise continuity, but has features of both. On complete spaces, strong continuity and pointwise continuity are equivalent while on totally bounded spaces strong continuity and uniform continuity are equivalent. Of course, spaces which are both complete and totally bounded make all three equivalent, which agrees with the result from basic analysis that continuous functions on compact spaces are uniformly continuous. This observation also strengthens the analogy between compactness for continuous functions and boundedness for strongly continuous functions, since in either case both conditions taken together imply uniform continuity.

Interlude: Number Systems

In the last chapter the idea of continuity was extended in such a way that functions on the rational numbers (as well as some other sets such as the algebraic numbers) would behave much like functions on the real numbers. Ideally, one might expect, this sort of thing could be done with other concepts from analysis and topology until a structure was in place which made the rational numbers just like the real numbers in all important respects. Any philosophical queasiness about uncountable cardinalities could then be sidestepped in good conscience by concentrating on the rational numbers, or perhaps the algebraic numbers, or even some more general but still countable number system.

A limited form of this plan is the essential purpose of this thesis, but it is undertaken with the realization that such attempts cannot be completely successful. The set of real numbers is different from its countable subsets, and those differences are not trivial or accidental. Complete spaces have properties which cannot be effectively mimicked in their incomplete subsets, and every countable metric space containing the rational numbers must be incomplete.

In usual treatments of continuity, for instance, there are generally theorems presented which claim that continuous functions on closed intervals attain minimum and maximum values, and if they are positive somewhere and negative elsewhere, they must assume the value zero in between. Such statements are simply *not true* on incomplete spaces, even when stronger sorts of continuity are required. If a function on the rational numbers, for instance, assumes a maximum at point x, a new space $\mathbf{Q} - \{x\}$ can be constructed which is dense in $\mathbf{Q}$, and has a function identical to the first one except that it doesn't attain its maximum at x. It is also easy to find functions which look obviously continuous on $\mathbf{Q}$ but don't have the expected zeros, such as $f(x) = x^2 - 3$.

A useful way to think of this problem is that number systems are created sometimes in order to allow certain operations to be well defined, and other times to allow certain sorts of equations to have solutions. The natural numbers can be defined as the smallest set which contains 1 and allows addition between any two of its elements. If the condition is then added that all equations of the form $x + a = b$ can be solved for x, the negative integers and zero must be included.

Multiplication is already well defined on the integers, but the equations $a{\cdot}x + b = 0$ do not always have solutions. Solving them requires the rational numbers. The search for solutions to equations like $\sum a_n x^n = 0$ leads to the algebraic and complex algebraic numbers.

Each of the number systems just listed is countable. This is a general fact about numbers created to solve specific sets of equations. If a notation is given which can represent all equations in the set, then the size of the set of acceptable equations must be countable. If the domain of the free variable is assumed to be countable, and it is found not to provide a solution for some of the equations, at most a countable number of new elements must be added to it to make them solvable. Now there may be new equations that can be formed from the new elements in the domain, and these may require additional elements to be added, but this process only has to be repeated a countable number of times, so the final set produced will still be countable.

As a result, any countable list of functions can be extended to a countable domain over which theorems about intermediate and extreme values will be true. For example, every polynomial in the algebraic numbers has a maximum and minimum value on any closed, bounded interval. If the polynomial is positive at $x = a$ and negative at $x = b$ then it will be zero on $[a, b]$.

The set of continuous functions on $\mathbf{R}$, or even the set of strongly continuous functions on $\mathbf{Q} \cap [0, 1]$ are not countable. There is not any countable set over which they will assume their mean or extreme values. In fact, this provides a new definition of the real numbers analogous to the previous definitions of other number systems.

Theorem. *The real numbers are the smallest ordered set which contains the rational numbers and in which every strongly continuous function that takes on both positive and negative values must somewhere be zero.*

The real numbers are obviously a set which meets these demands, but is it the smallest? This can be shown to be true by constructing a pair of sequences of rational numbers which converge to some real number x_0, l_n converging from below and r_n converging from above. This can be done for any x_0. A function f can now be assembled by taking $f(l_n) = -1/n$ and connecting these points by linear segments, then taking $f(r_n) = 1/n$, connecting these by linear segments and making $f(x) = 1$ for large x and $f(x) = -1$ for small x. f is strongly continuous, and if x_0 is not in the domain of f then f assumes both positive and negative values but is never zero. If f is required to have a zero then x_0 must be in the domain of f. Since x_0 can be any real number, any set which meets our definition must contain every real number.

3 Differentiation

Derivatives

Derivatives are usually only calculated for functions on complete spaces such as $\mathbf{R}^n$ or $[0,1] \subseteq \mathbf{R}$. In fact, the usual definitions and techniques for differentiation are often applicable to other spaces, such as $\mathbf{Q}$. The one dimensional derivative

$$f'(x) = \lim_{t \to 0} \frac{f(x+t) - f(x)}{x - t}$$

is well defined for many functions on the rational numbers, and often gives reasonable values for derivatives. Rational polynomials, for example, are all differentiable on all of $\mathbf{Q}$ by this definition, and $\frac{d}{dx} x^n = n x^{n-1}$. Also,

$$\begin{aligned} (f+g)' &= f' + g', \\ (f \cdot g)' &= f \cdot g' + g \cdot f', \\ (f \circ g)' &= (f' \circ g) \cdot g', \qquad \text{and} \\ \left(\frac{1}{f}\right)' &= \frac{-f'}{f^2}, \end{aligned}$$

whenever f and g are differentiable over appropriate domains. None of these facts depends on details of the number system such as completeness, so they will be true on any dense subset of $\mathbf{R}$ where the arithmetic operations are defined.

Difficulties with differentiation on incomplete spaces do occur for other reasons, however. This definition of differentiation calculates derivatives for functions which look like they shouldn't be differentiable.

Example 3.1 *The function defined in example 2.1,*

$$f(x) = \begin{cases} 1 & \text{if } x^2 < 2; \\ 0 & \text{if } x^2 > 2. \end{cases}$$

is not only continuous on the rational numbers, it is differentiable, and $f'(x) = 0$ for all x.

There are also functions from an incomplete set to itself which fail to be differentiable because although all of their values are in the set, the derivative at some point is not.

Example 3.2 *Construct a function $f : [-1,1] \cap \mathbf{Q} \to \mathbf{Q}$ by setting $f(\frac{1}{n}) = \frac{s_n}{n}$ and $f(\frac{-1}{n}) = \frac{-s_n}{n}$ where $s_1 = 1$, $s_2 = 1.4$, $s_3 = 1.41$, and $\lim_{n \to \infty} {s_n}^2 = 2$. The values of $f(x)$ for $x \neq \frac{1}{n}$ can be found by connecting these points with straight line segments and setting $f(0) = 0$.*

The derivative at $x = 0$ is $\lim_{n \to \infty} s_n$, which does not exist in $\mathbf{Q}$.

Failures of this second sort are a fundamental property of incomplete sets. Essentially the procedure used in example 3.2 can be used in any incomplete field to produce a function whose derivative at zero would be the (nonexistent) limit of some Cauchy sequence s_n. These examples must either be accepted as not differentiable, or their derivatives must be considered to be functions into a superset of their codomains.

Example 3.1 demonstrates a problem with differentiation which is much harder to ignore. Much of the usefulness of the derivative comes from the fact that functions can be categorized as constant, increasing or decreasing by finding whether the derivatives are zero, positive, or negative. The calculation of integrals as antiderivatives assumes that two functions with the same derivative must differ by a constant.

Strong Differentiability

In order to eliminate this problem, a new condition must be imposed on differentiable functions. One such restriction is as follows.

Definition 3.1 *A function f with domain A is* **strongly differentiable** *if and only if $f' = lim_{t\to 0}\frac{f(x+t)-f(x)}{t}$ exists for all $x \in A$ and the function*

$$g(x,t) = \begin{cases} \frac{f(x+t)-f(x)}{t} & \text{if } t \neq 0 \\ f'(x) & \text{if } t = 0 \end{cases}$$

is strongly continuous.

This notion is stronger than differentiability, even on complete sets. In particular, it forces f' to be continuous. In fact, on closed subsets of $\mathbf{R}$, that additional condition is sufficient.

Theorem 3.1 *A function $f : [a,b] \subseteq \mathbf{R} \to \mathbf{R}$ is strongly differentiable if and only if it has a continuous derivative everywhere on $[a,b]$.*

Proof. If f is strongly differentiable, its derivative must exist everywhere, and since $f'(x) = g(x,0)$ and g is continuous, f' must be continuous as well.

Since g is defined on a closed subset of $\mathbf{R}^2$, it will be strongly continuous whenever it is continuous. But g will be continuous at (x_0, y_0) if for any $\epsilon > 0$ there is some $\delta > 0$ so that for any (x,t) where $d(x_0,x) + d(t_0,t) < \delta$, $d(g(x_0,t_0), g(x,t)) < \epsilon$. If $t_0 \neq 0$, then any $\delta < d(t_0, 0)$ is sufficient, since the function g is continuous whenever f is continuous and $t \neq 0$ by the arithmetic properties of continuity.

If $t_0 = 0$ and $t \neq 0$, $d(g(x_0,t_0), g(x,t)) \leq d(g(x_0,t_0), g(x_0,t)) + d(g(x_0,t), g(x,t))$. There is some δ_1 such that $d(t_0,t) < \delta_1$ implies $d(g(x_0,t_0), g(x_0,t)) < \epsilon/2$ because g must be continuous with respect to t near 0 when f is differentiable. There must also be a δ_2 such that $d(x_0,x) < \delta_2$ implies $d(g(x_0,t), g(x,t)) < \epsilon/2$ because g is continuous when $t \neq 0$. Choosing $\delta = \min(\delta_1, \delta_2)$ forces $d(g(x_0,t_0), g(x,t)) < \epsilon$.

If $t_0 = 0$ and $t = 0$, it is sufficient that f' is continuous. ■

This definition is restrictive enough to force strongly differentiable functions to have all the properties I am interested in, but it may be somewhat too restrictive. The function

$$f(x) = \begin{cases} x & \text{if } x^2 < 2, \\ \frac{1}{x} & \text{if } x^2 > 2. \end{cases}$$

on the rational numbers is not strongly differentiable, but it is differentiable everywhere, increases where its derivative is positive, decreases where its derivative is negative, and behaves as expected under the fundamental theorem of calculus (which is proven later and uses strong differentiability.) In fact, for all of these theorems, strong continuity and piecewise strong differentiability is all that is required. It should be noted, however, that strong continuity of $g(x_0,t)$ and $g(x,t_0)$ for fixed x_0 and t_0 is a weaker condition than strong differentiability, and there are functions with that property for which Theorem 3.2 does not hold.

Theorem 3.2 *If $f : A \to B$ is a strongly differentiable function, $A \subseteq \mathbf{R}$ is not the union of two subsets which are not close, and $B \subseteq \mathbf{R}$, the following are true.*

(a) If $f' \geq 0$, f is nondecreasing.
(b) If $f' \leq 0$, f in nonincreasing.
(c) If $f' = 0$, f is constant.
(d) If $f' > 0$, f is strictly increasing.
(e) If $f' < 0$, f is strictly decreasing.

Proof. A function is nondecreasing exactly when it is nondecreasing over every bounded subset of its domain. The same is true for nonincreasing, decreasing, increasing, and constant functions. It suffices to show results for arbitrary bounded subsets of A.

(a) The function $g(x,t) = (f(x+t)-f(x))/t$ for $t \neq 0$ and $g(x,0) = f'(x)$ is strongly continuous by hypothesis, and since its domain is bounded it must also be uniformly continuous. This means that for any $\epsilon > 0$, there is some $\delta > 0$ so that $f(x+t)-f(x)/t > f'(x)-\epsilon \geq -\epsilon$ whenever $|t| < \delta$, and this δ is independent of x.

If $f(a) = c$, $f(a+\delta/2) > c - \epsilon{\cdot}\delta/2$. This in turn implies that $f(a+\delta/2+\delta/2) > f(a+\delta/2) - \epsilon{\cdot}\delta/2$, so $(f(a+\delta) > c-\epsilon{\cdot}\delta$. In general, if $f(a+x) > c-\epsilon{\cdot}x$, $f(a+x+t) > c-\epsilon{\cdot}(x+t)$ whenever $0 < t < \delta$. Applying this procedure inductively shows that $f(a+x) > c - \epsilon{\cdot}x$ whenever $x > a$ as long as for every point $a+x$ there is some point $a+x+t$ where $0 < t < \delta$. If there were no such point, A could be divided into subsets which weren't close by choosing $A_1 = \{y \in A : y \leq a+x\}$ and $A_2 = \{y \in A : y > a+x\}$. Thus, for any $a \in A$ and $a+x \in A$, $f(a+x) > f(a) - \epsilon{\cdot}x$. Since ϵ is arbitrary, this forces $f(a+x) > f(a)$.

The proof of (b) is analogous to that of (a).

(c) If $f' = 0$, then the function is nondecreasing by (a), and nonincreasing by (b), so f must be constant.

(d) If f is nondecreasing, but not strictly increasing, there must be some x and δ so that for any $0 < t < \delta$ $f(x+t) = f(x)$. But then $f'(x+t) = 0$, contrary to the assumption $f' > 0$. (e) follows by the same argument. ∎

The usual properties for arithmetic combination of differentiable functions generalize in the obvious way.

Theorem 3.3 *If f_1 and f_2 are strongly differentiable then so are $f_1 + f_2$, $f_1{\cdot}f_2$ and $f_2 \circ f_1$, and $1/f_1$ is strongly differentiable wherever $f_1 > \epsilon$ for some fixed positive ϵ.*

Proof. Assuming

$$g_n(x,t) = \begin{cases} \frac{f_n(x+t)-f_n(x)}{t} & \text{if } t \neq 0 \\ f_n'(x) & \text{if } t = 0 \end{cases},$$

if $f = f_1 + f_2$ then $g = g_1 + g_2$, so g is strongly continuous, and $f_1 + f_2$ is strongly differentiable.

If $f = f_1{\cdot}f_2$, then

$$g(x,t) \quad = \quad \frac{f_1(x+t){\cdot}f_2(x+t) - f_1(x){\cdot}f_2(x)}{t},$$

$$= \quad f_1(x+t)\cdot\left(\frac{f_2(x+t)-f_2(x)}{t}\right)+f_2(x)\cdot\left(\frac{f_1(x+t)-f_1(x)}{t}\right)$$

whenever $t \neq 0$, so $g(x,t) = f_1(x+t)\cdot g_2(x,t) + f_2(x)\cdot g_1(x,t)$, which is strongly continuous whenever f_1, f_2, g_1 and g_2 are strongly continuous. g_1 and g_2 are strongly continuous by hypotheses, and f_1 must be strongly continuous since $f_1(t) = t\cdot g_1(x_0,t) + c$ for any fixed x_0 if $c = f(x_0)$. Since g_1, t and c are strongly continuous f_1 must be, as must $f_2 = t\cdot g_2(x_0,t) + f_2(x_0)$.

Since uniform continuity implies strong continuity and strong continuity only requires that bounded sets have close images, showing that g is uniformly continuous on arbitrary bounded sets when $f = f_2 \circ f_1$ is sufficient to show that g is strongly continuous.

For any fixed bounded domain, g_1 and g_2 are uniformly continuous. This means that for any $\epsilon_1 > 0$ and $\epsilon_2 > 0$ there are $\delta_1 > 0$ and $\delta_2 > 0$ such that $|p_1| < \epsilon$ when $p_1 = g_1(x,t) - f_1'$ and $t < \delta_1$, and similarly for p_2.

Since

$$\begin{aligned}
g(x,t) &= \frac{f_1(f_2(x+t)) - f_1(f_2(x))}{t} \\
&= \frac{f_1(t\cdot(f_2'(x)+p_2)+f_2(x)) - f_1(f_2(x))}{t} \\
&= \frac{t\cdot(f_2'(x)+p_2)\cdot(f_1'(b(x))+p_1)}{t} \\
&= f_1'(f_2(x))\cdot f_2'(x) + p_1\cdot f_2'(x) + p_2\cdot f_1'(f_2(x)) + p_1\cdot p_2
\end{aligned}$$

(this las equality holds even for $t = 0$),

$|g(x,t) - g(x_0,t_0)| < |f_1'(f_2(x))\cdot f_2'(x) - f_1'(f_2(x_0))\cdot f_2'(x_0)| + \epsilon\cdot|f_2'(x) + f_1'(f_2(x))|$ as long as t and $t_0 < \min(\delta_1,\delta_2)$. Since f_1' and f_2' are bounded and $f_1'(f_2(x))\cdot f_2'(x)$ is uniformly continuous, $g(x,t)$ must also be uniformly continuous wherever $t \leq \min(\delta_1,\delta_2)/2$. It must also be uniformly continuous wherever $t \geq \min(\delta_1,\delta_2)/2$ since f is uniformly continuous and t is bounded away from 0. Since a function which can be divided into a finite number of uniformly continuous pieces must be uniformly continuous, g uniformly continuous on bounded domains, and so strongly continuous everywhere.

$1/f_1$ must be strongly differentiable wherever f_1 and $1/x$ are. f_1 is strongly differentiable everywhere, and $1/x$ is whenever $|x| > \epsilon$ for any $\epsilon > 0$. ■

Of course, since strong differentiability is an additional condition imposed on the functions, rather than a change in the meaning of the derivative, the derivatives of these functions can be calculated by the usual methods. Also, any functions which are continuously differentiable on **R** are strongly differentiable when restricted to dense subsets of **R**. The usual warnings apply about functions such as $1/x$ which are only defined on some open subset of **R**, which sometimes must be restricted to a closed set before they become strongly differentiable.

The work on differentiation in the previous chapter assumes that the sets on which it operates are one dimensional. Exactly what I even mean by that may not be entirely clear. Ideally, sets which are dense in $\mathbf{R}^n$ could be thought of as n-dimensional, and more generally, if some set were dense in a more complicated space it would inherit the dimension of that space.

This approach has two problems. First, it assumes a well developed theory of the dimension of complete spaces, which is exactly the sort of knowledge I am trying to avoid using for such definitions. Even worse, the usual methods for finding the dimensions of complete spaces already assign dimensions to some of their dense subsets, and these are often different than the ones I would like to assign. In particular, sets which are totally disconnected often have dimension zero, and some methods for calculating topological dimension make every countable set zero dimensional.

A new definition of dimension is necessary. It cannot rely on standard topological techniques, because this would make dimension a homeomorphism invariant, but I would like the spaces $\mathbf{Q}$ and $\mathbf{Q} \times \mathbf{Q}$ to have different dimensions even though they are homeomorphic[1]. Making dimension invariant under strong homeomorphism (strongly continuous bijection) would seem desirable, however.

One way to go about this is to assign finite sets dimension zero, then examine larger sets which are not the union of two sets unless those sets are close (this is essentially a generalization of connectedness) using the following method.

Definition. *A set X has dimension $n+1$ if there is a set $Y \subseteq X$ of dimension n such that any set C which makes Y not close to $X - C$ also separates $X - C$ into subsets which are not close, but there are no sets of smaller dimension than n which do this.*

This makes $\mathbf{Q}$ 1 dimensional, since any set which can be removed to make a single point not close to the rest of $\mathbf{Q}$ must contain some interval around the point, and this will then separate $\mathbf{Q}$ into two sections which are not close. It does have the unfortunate property of assigning a space which appears to have dimension n near one point and dimension m near another a dimension of $\min(n, m)$, so one stray point makes the entire set zero dimensional.

Differentiation in n-space

Since there is now some evidence that dimension can still mean something in this context, the question of how to take derivatives in spaces of higher than one dimension can be pursued. Ideally a generalization of the strategy used in the last chapter of creating a function from the difference quotients and forcing it to be continuous as the difference becomes small should work. So perhaps the derivative of a function on an n dimensional vector space can be found by finding a linear transformation $D_{x,t}$ where x and t are vectors such that

$$\frac{|f(x+t) - f(x) - D_{x,t}(t)|}{|t|} = 0$$

[1]The fact that $\mathbf{Q}$ and $\mathbf{Q} \times \mathbf{Q}$ are homeomorphic is not easy to demonstrate, and I am indebted to Dr. Kenneth Ross of the University of Oregon for proving it to me.

for every x and t and insisting that the resulting set of linear transformations form a continuous function on $x \times t$. Unfortunately, in general this method does not uniquely determine D at any point, so constructing a unique continuous extension of it is not really a well defined operation. In order to completely determine D at each point the difference quotient must involve n points. that is, the equations

$$\frac{|f(x+t_1) - f(x) - D_{x,t_1,\ldots,t_n}(t_1)|}{|t_1|} = 0$$

$$\frac{|f(x+t_2) - f(x) - D_{x,t_1,\ldots,t_n}(t_2)|}{|t_2|} = 0$$

$$\vdots$$

$$\frac{|f(x+t_n) - f(x) - D_{x,t_1,\ldots,t_n}(t_n)|}{|t_n|} = 0$$

must all be satisfied simultaneously. This makes D a function into the set of linear transforms from a $n{\cdot}n + n$ dimensional space produced by taking the product of $n+1$ copies of the domain of f. As long as $t_i \neq t_j$ for $i \neq j$ and $t_i \neq 0$ for all i, this function is well defined. Insisting that it be continuously extendable to those places where it isn't defined and evaluating it where $t_i = 0$ for all i may give a reasonable definition for the derivative. If this is the case, it will certainly have to be suplimented with notions like the matrix of partial derivatives in order to be useful for calculations, since continuity is difficult to demonstrate on such abstract spaces and the dimension of the space where the difference quotient exists becomes large rapidly with increasing n.

4 Integration

When working with derivatives, it was clear that the fundamental definition of the derivative of a function at a point extended without change to sets which were not complete. Some new sorts of pathological behavior had to be accounted for, but the basic calculation was mostly unexamined and unmodified.

Integration does not generalize so easily. The usual elementary definition of the integral involves taking supremums and infimums over intervals, as well as over sets of partitions. The notion of a supremum only makes sense in a complete space, so integration by this process on sets like $\mathbf{Q}$ will in general result in undefined values, even for functions where a reasonable integral should exist. In practice, integrals do not have to be computed using supremums or arbitrary partitions. If a function is integrable, any sequence of partitions with characteristic length tending towards zero can be evaluated on arbitrary points within their segments, and they will still tend toward the correct value for the integral. Unfortunately, this process will proceed and give values for the integrals of functions which are not integrable in the usual sense. Under these conditions, the exact method used to select the partitions and sample points can change the calculated value of the integral. Obviously, some method must exist for finding which functions can be assigned a meaningful and unique integral.

One strategy for deciding which functions should be integrable is to select a collection of partitions which can be assigned topological properties, and insist that the approximations to the integral given for each partition form a continuous function. This eliminates from consideration those functions for which small changes in the structure of the partition produce radically different values for the integal.

If X is a subset of $\mathbf{R}$ which contains points A and B and where $A \leq x \leq B$ for any $x \in X$, A partition of X is a finite list of points in X which contains both A and B. Partitions can also be thought of as being numbered so that if $P_0 \dots P_N$ is a partition, $P_0 \leq P_1 \leq \cdots \leq P_N$. Partitions can be assigned a measure by defining the length of the partition P_N

$$|P_N| = \sqrt{\sum_{i=1}^{N} (P_i - P_{i-1})^2}.$$

The distance between two partitions can be calculated by extending the one with the smaller number of elements by adding points B (which doesn't change its length) until the two have the same number of elements. The distance between them is

$$d(P_N, Q_N) = \sqrt{\sum_{i=1}^{N} ((P_i - P_{i-1}) - (Q_i - Q_{i-1}))^2}.$$

This has the usual properties of a metric, and $d(P,Q) \leq |P| + |Q|$ so the length function can be thought of as $d(P, P_0)$ even though there is no actual partition which acts as P_0. I will artificially add the object P_0 to the space of partitions, and give it the properties $d(P, P_0) = |P|$ and $|P_0| = 0$, even though it is not a partition.

A refinement of a partition is another partition which contains all the points of the original at least as many times, and possibly others.

Theorem 4.1 *If Q_N is a refinement of P_N, then $|Q_N| \leq |P_N|$.*

Proof. If Q_N contains only points which are in P_N (even if it contains more instances of some of them) then $|Q_N| = |P_N|$. If not, Q_N must be a refinement of P'_N where P'_N is a refinement of P_N which contains only points from P_N, and therefore has the same length. Q_N must differ from P'_N only be having some points which are between adjacent points of P'_N. If Q_N has points $q_1 \ldots q_n$ between P'_i and P'_{i-1} then

$$(P'_i - P'_{i-1}) = (P'_i - q_n) + (q_n - q_{n-1}) + \cdots + (q_2 - q_1) + (q_1 - P'_{i-1})$$

and so

$$(P'_i - P'_{i-1})^2 \geq (P'_i - q_n)^2 + (q_n - q_{n-1})^2 + \cdots + (q_2 - q_1)^2 + (q_1 - P'_{i-1})^2$$

Substitution of this inequality into the expression for $|P'_N|$ for each pair of points which are separated by elements of Q_N gives $|P'_N|^2 \geq |Q_N|^2$, so $|Q_N| \leq |P_N|$. ■

Although there is no partition of length zero, partitions can have arbitrarily small lengths. In fact, sequences of partitions have lengths which converge to zero if and only if the values $\max_{all\ i} |P_i - P_{i-1}|$ converge to zero.

Theorem 4.2 *For any $\epsilon > 0$ there is some $\delta > 0$ such that any partition P_N where $|P_i - P_{i-1}| < \delta$ whenever $1 \leq i \leq N$ has $|P_N| < \epsilon$.*

Proof. Suppose P_N has $|P_i - P_{i-1}| < \delta$ whenever $1 \leq i \leq N$. A new partition Q_M can be constructed from P_N by removing points so that $\delta \leq |Q_i - Q_{i-1}| < 2{\cdot}\delta$ whenever $1 \leq i \leq M$ and P_N is a refinement of Q_M. Each term in the expression for $|Q_N|$ is less than $(2{\cdot}\delta)^2$, and there are at most $\lfloor \frac{P_N - P_0}{\delta} \rfloor$ of them. this makes

$$|Q_M| < \sqrt{(2{\cdot}\delta)^2{\cdot}\frac{P_N - P_0}{\delta}}$$

so choosing $\delta = \epsilon^2/4{\cdot}(P_N - P_0)$ forces $|P_N| \leq |Q_M| < \epsilon$. ■

Of course, if there is a $|P_i - P_{i-1}| > \delta$ for some δ, $|P_N| \geq \sqrt{\delta^2}$, so the condition is necessary as well as sufficient.

The integral can now be defined in terms of continuity on the set of all partitions.

Definition 4.1 *The* **integral by continuity** *of a function $f : X \to \mathbf{R}$ is the value which must be assigned to the point P_0 attached to the space of all partitions over X, so that the function*

$$g(P_N, f) = \sum_{i=0}^{N-1} f(P_i){\cdot}(P_{i+1} - P_i)$$

is continuous at P_0, if a unique such value exists. Otherwise the integral is undefined.

Of course, X must be a bounded subset of $\mathbf{R}$ in order for the space of partitions to be defined, and X must be dense in order for there to be partitions arbitrarily close to P_0. The integral by continuity can be written

$$\oint_A^B f \qquad \text{or} \qquad \oint_A^B f(x)\,dx$$

when the domain of integration is the region of the domain of f bounded by A and B.

In the definition, the function f was given the real numbers as codomain. This is done as a matter of convenience of notation, as f can be assumed to go into whatever subset of $\mathbf{R}$ seems appropriate without changing the content of the definition. The value of the integral itself may also be required to lie in some proper subset of $\mathbf{R}$, in particular, the function

$$F(x) = \oint_0^x f(t)\,dt$$

might reasonably be expected to have the same codomain as f. As in the case of differentiation, however, it is possible to construct functions f which have integrals F that are only well defined into supersets of the codomain of f whenever that set is not complete. In this case, as long as the codomain has the arithmetic properties of closure under addition, subtraction and multiplication $g(P_N, f)$ is defined for every partition except possibly P_0. This means that the existence of the continuous integral depends on only one possibly undefined limit process in the codomain, instead of the usual arbitrarily large number.

In the case where both the domain and codomain are complete (that is, closed subsets of $\mathbf{R}$), the integral by continuity is applied to the same cases where the Riemann integral works. The two should agree if both are defined, and also be defined for similar types of functions. In fact, on intervals in $\mathbf{R}$ the two are equivalent.

Theorem 4.3 *A function $f : \mathbf{R} \to \mathbf{R}$ is Riemann integrable on $[A, B]$ if and only if it is integrable by continuity.*

Proof. If a function is Riemann integrable, for and $\epsilon > 0$ there is some $\delta > 0$ such that for any partition with $\max |P_i - P_{i-1}| < \delta$

$$\left| \int f - \sum_{i=1}^{N} f(x_i)\cdot(P_i - P_{i-1}) \right| < \epsilon$$

where $x_i \in [P_{i-1}, P_i]$. This means that $g(P, f)$ will be continuous at P_0 if and only if $g(P_0, f) = \int f$.

Conversely, if f is integrable by continuity, there is some $\delta > 0$ for each $\epsilon > 0$ so that any partition with $\max |P_i - P_{i-1}| < \delta$ has sum $|g(P, f) - \oint f| < \epsilon/4$. Choose any partition Q with $\max |Q_i - Q_{i-1}| < \delta/2$. A new partition U can be constructed with one more point than Q by assigning $U_0 = Q_0$, $U_{N+1} = Q_N$ and choosing $U_i \in [Q_{i-1}, Q_i]$ for $1 \leq i \leq N$ such that $f(U_i) > \sup_{x\in[Q_{i-1},Q_i]} f(x) - \epsilon/4\cdot(B - A)$. Now $g(U, f) > \mathcal{U}(U) - \epsilon/4$, where $\mathcal{U}(U)$ is the Riemann upper sum of U, and a partition L can be constructed by analogous methods so that $g(L, f) < \mathcal{L}(L) + \epsilon/4$. Since $\max |U_i - U_{i-1}| < \delta$ and $\max |L_i - L_{i-1}| < \delta$, $|g(U, f) - g(L, f)| < \epsilon/2$, so $\mathcal{U}(U) - \mathcal{L}(L) < \epsilon$ for any ϵ, and f must be Riemann integrable. ∎

Even in cases where the domain and codomain of f are not complete, it is true that

$$\oint_A^B f + \oint_B^C f = \oint_A^C f$$

if all of these integrals exist, although cases can be constructed where f is not integrable on $[A, B]$ or $[B, C]$ but is integrable on $[A, C]$.

It is always true when addition and scalar multiplication are defined that if f and g are integrable,

$$\oint (f+g) = \oint f + \oint g \qquad \text{and} \qquad \oint (c\cdot f) = c\cdot \oint (f).$$

The Fundamental Theorem of Calculus

In complete spaces, the operations of differentiation and integration behave much like inverses of each other. For cases where all of the operations in question are defined,

$$\frac{d}{dx}\int_A^x f(t)\,dt = f(x) \qquad \text{and} \qquad \int_A^B f' = f(B) - f(A).$$

Ideally this should also be true for more general domains. Some care must be taken, however, since the mere existence of derivatives and integrals is not always sufficient.

Example 4.1 *Suppose $f : \mathbf{Q} \to \mathbf{Q}$ is defined by*

$$f(x) = \begin{cases} 1 & \text{if } x^2 < 2; \\ 0 & \text{if } x^2 > 2. \end{cases}$$

This function is differentiable everywhere, and $f' = 0$. This makes

$$\oint_0^2 f' = \oint_0^2 0 = 0 \neq f(2) - f(0).$$

As might be expected, something stronger than the existance of derivatives at every point is necessary here. For arbitrary differentiable functions taking a derivative loses not only information about the value of $f(A)$, but also the magnitude of any steplike features such as those in example 4.1. Basically, since not all functions with derivative 0 everywhere differ by only a constant, integration cannot uniquely reconstruct them. Functions which are strongly differentiable are determined up to a constant by their derivatives, by Theorem 3.2, so the fundamental theorems should apply to them.

Theorem 4.4 *If f is integrable by continuity on X, A is the least element of X and f is continuous on X, then*

$$\frac{d}{dx}\oint_A^x f = f(x)$$

whenever $x \in X$.

Proof. For any $\epsilon > 0$ there must be some $\delta_0 > 0$such that for any partition of X, P_N such that $\max |P_i - P_{i-1}| < \delta_0$,

$$-\epsilon + \sum_{i=0}^{N-1} f(P_i)\cdot(P_{i+1} - P_i) < \oint_X f < \epsilon + \sum_{i=0}^{N-1} f(P_i)\cdot(P_{i+1} - P_i).$$

In particular, $P_i = A + i\cdot\delta$ for $i < N$ and $P_N = \max(X)$ is such a partition if $\delta < \delta_0$.

In order to be integrable by continuity, there must be some M so that $|f(x)| < M$ for any $x \in X$ (otherwise, by selecting points in the partition where f is arbitrarily large, partitions of arbitrarily small mesh could have arbitrarily large sums), so $\epsilon_2 \geq \epsilon + M{\cdot}\delta$ can be made arbitrarily small by selecting ϵ and δ to be small. If $n = \lfloor (x-A)/\delta \rfloor$ then

$$-M{\cdot}\delta + \oint_A^x f < \oint_A^{A+n\cdot\delta} f \leq \oint_A^x f$$

so

$$-\epsilon_2 + \sum_{i=0}^{n-1} \delta{\cdot}f(A+i{\cdot}\delta) < \oint_A^x f < \epsilon_2 + \sum_{i=0}^{n-1} \delta{\cdot}f(A+i{\cdot}\delta)$$

and

$$-\epsilon_2 + \sum_{i=0}^{n} \delta{\cdot}f(A+i{\cdot}\delta) < \oint_A^{x+\delta} f < {\cdot}\epsilon_2 + \sum_{i=0}^{n} \delta{\cdot}f(A+i{\cdot}\delta)$$

Using this to approximate the derivative,

$$\frac{1}{\delta}{\cdot}\left(\sum_{i=0}^{n}\left(\frac{-\epsilon_2}{n} + \delta{\cdot}f(A+i{\cdot}\delta)\right) - \sum_{i=0}^{n-1}\left(\frac{\epsilon_2}{n-1} + \delta{\cdot}f(A+i{\cdot}\delta)\right)\right)$$
$$< \frac{1}{\delta}{\cdot}\Big(\oint_A^{x+\delta} f - \oint_A^x f\Big) <$$
$$\frac{1}{\delta}{\cdot}\left(\sum_{i=0}^{n}\left(\frac{\epsilon_2}{n} + \delta{\cdot}f(A+i{\cdot}\delta)\right) - \sum_{i=0}^{n-1}\left(\frac{-\epsilon_2}{n-1} + \delta{\cdot}f(A+i{\cdot}\delta)\right)\right)$$

which simplifies to

$$\frac{-\frac{\epsilon_2}{n-1} + \delta{\cdot}f(A+n{\cdot}\delta)}{\delta} < \frac{1}{\delta}{\cdot}\Big(\oint_A^{x+\delta} f - \oint_A^x f\Big) < \frac{\frac{\epsilon_2}{n-1} + \delta{\cdot}f(A+n{\cdot}\delta)}{\delta}$$

and

$$-\frac{\epsilon_2}{\delta{\cdot}(n-1)} + f(A+n{\cdot}\delta) < \frac{1}{\delta}{\cdot}\Big(\oint_A^{x+\delta} f - \oint_A^x f\Big) < \frac{\epsilon_2}{\delta{\cdot}(n-1)} + f(A+n{\cdot}\delta).$$

If δ is made small, then $\delta{\cdot}(n-1) \approx (x-A)$ and since f is continuous $f(A+n{\cdot}\delta) \approx f(x)$. Thus, for fixed x and ϵ_2, as $\delta \to 0$

$$-\frac{\epsilon_2}{(x-A)} + f(x) < \frac{d}{dx}\oint_A^x f < \frac{\epsilon_2}{(x-A)} + f(x).$$

and since ϵ_2 can be arbitrarily small,

$$\frac{d}{dx}\oint_A^x f = f(x)$$

for any $x > A$.

For the case $x = A$, notice that f is continuous, so for any $\epsilon > 0$ there is some $\delta > 0$ such that $|f(x+\delta) - f(x)| < \epsilon$. This means

$$\left|\oint_A^{A+\delta} f - f(x){\cdot}\delta\right| < \epsilon{\cdot}\delta.$$

Varying δ with ϵ fixed gives

$$\left|\lim_{\delta\to 0}\left(\frac{1}{\delta}\cdot\oint_A^{A+\delta} f\right) - f(A)\right| < \epsilon$$

and since ϵ is arbitrary, this makes

$$\frac{d}{dx}\oint_A^x f = f(x)$$

true for $x = A$. ∎

This theorem is not the one which requires the extra conditions; the only restriction on f is that it be continuous (not even strong continuity is necessary.) Strong differentiability, and with it strong continuity, are required for the other theorem of the pair.

Theorem 4.5 *If f is strongly differentiable on $X \cap [A, B]$ and f' is integrable by continuity on $X \cap [A, B]$ then*

$$\oint_A^B f' = f(B) - f(A).$$

Proof. Since f' is integrable by continuity, for any $\epsilon > 0$ there is some $\delta_1 > 0$ such that for any $0 < \delta < \delta_1$

$$-\epsilon + \sum_{i=0}^{N} \delta\cdot f'(A + i\cdot\delta) < \oint_A^B f' < \epsilon + \sum_{i=0}^{N} \delta\cdot f'(A + i\cdot\delta)$$

where $N = \lfloor (B - A)/\delta \rfloor$. Also, since f is strongly differentiable and its domain is bounded, the approximation function g for f' must be uniformly continuous. This means that there must be some $\delta_2 >$ so that for any $\delta < \delta_2$

$$-\epsilon + \frac{f(A + i\cdot\delta + \delta) - f(A + i\cdot\delta)}{\delta} < f'(A + i\cdot\delta) < \epsilon + \frac{f(A + i\cdot\delta + \delta) - f(A + i\cdot\delta)}{\delta}$$

where δ_2 does not depend on $A + i\cdot\delta$. Combining these results gives

$$\left|\oint_A^B f' - \sum_{i=0}^{N} \frac{\delta\cdot(f(A + i\cdot\delta + \delta) - f(A + i\cdot\delta))}{\delta}\right| < \sum_{i=0}^{N} \delta\cdot\epsilon + \epsilon$$

or, since $N\cdot\delta \leq (B - A)$,

$$\left|\oint_A^B f' - (-f(A) + f(A + N\cdot\delta + \delta))\right| < \epsilon\cdot(B - A) + \epsilon$$

Since ϵ is arbitrary, the term on the left must be 0, and since f is continuous and $B \leq A + N\cdot\delta + \delta \leq B + \delta$ this must mean

$$\oint_A^B f' = f(B) - f(A).$$

∎

These results allow the usual techniques of integration to be used. If a strongly differentiable function F can be found which has derivative $F' = f$, then the integral of f can be found directly by evaluating F. The formula for integration by parts follows directly from the product rule for differentiation.

Interlude: More Sets of Numbers

The past few chapters have gone through arcane contortions to produce definitions of the operations of calculus which make sense on sets of numbers which aren't complete. This seems like a reasonable project, but in execution something seems to be missing; for both integration and differentiation there is a remark buried away in the text to the effect that these operations still cannot be guaranteed to be defined unless they are taken to give arbitrary real numbers as values. Certainly the processes used to approximate these operators as presented now give meaningful values for any closeness of approximation desired, but at the final step there is still a limit process, which means that the correct value can still fall anywhere in the real numbers, not just in some incomplete subset. It is perhaps a nice thing to have definitions which fail only in predictable ways, and to know that when they do converge to a solution that the solution will have reasonable properties, but we are still left with the fundamental problem that sometimes functions which look like they should have derivatives or integrals somehow simply do not.

My response to this is to draw an analogy to the algebraic numbers. When working in the rational numbers often there are polynomials which give every impression of having roots, but where the roots are not any rational number. This situation can be dealt with either by appealing to the set of real numbers, with its attendant abilities and difficulties, or by constructing a new set which contains the rationals as well as new elements for all the missing solutions. This set is more difficult to work with than the rationals, and its elements are harder to characterize, but it is still countable, and its elements can be defined strictly in terms of their algebraic properties. The tendency for reasonable looking operations to be undefined is a general property of incomplete spaces, but for any specific set of operations an incomplete set can be found which suffices for them.

The temptation here is to simply assert that there is some countable set of numbers which is closed under differentiation and integration, and perhaps to make a few conjectures about its structure. Unfortunately, example 3.2 had already shown that this is not true. There are functions defined over the rational numbers which take any one element of the reals as a derivative somewhere. What has gone wrong?

Integration and differentiation are not operations on numbers. They are operators on functions. The correct space to try to enlarge in order to make these processes closed is a space of functions. If we are to consider all possible functions from the rational numbers to themselves, there are already an uncountable number of them and it is no surprise that in order to make the set closed under some arbitrary operation another uncountable number must be added. Since some of these functions bring with them new points in their ranges, the underlying space of numbers can be expected to become large.

If the idea of closure of a countable set under derivatives and integrals is to be saved, the set to first consider is the set of functions on which they act. We can begin by choosing an essentially arbitrary set of numbers on which these functions will be provisionally defined. It will probably need to be expanded as functions are added, but as long as it is clear how the existing functions should be extended to their new domains this does not necessarily present a problem. If we want this set to be countable when we are finished it should be countable when we start, of course, and it should have the sort of ordering properties and arithmetic operations that allow useful calculations. If these conditions are satisfied for the initial set of numbers, they will still be true when all extensions are made, since closing the set under algebraic operations should be straightforward and as long as only a countable number of functions are added to the space of functions, even if each one required the addition of a countable number of points to the set of numbers, that set will still be countable.

My inclination is to use the rational numbers as an initial set, occasionally restricting the domains of functions to intervals in the rationals to avoid division by zero and keep algebraic operations from making functions unbounded. This done, an initial set of functions should be chosen. A very large set could plausibly be used on the assumption that it would be more likely to contain every useful function when the process was finished, but it is also possible to begin with a very small set based on the philosophy that this will allow easier analysis of the final result by algebraic methods. If both the set of operations and the set of initial objects which they act on to generate the final set are small and well understood, this leaves more opportunities to develop powerful tools to examine the structure of the final result. With this in mind, I propose using the set $\{1, x\}$. (In fact, x is a redundant member, since $\int 1 = x$, but I am including it because it allows me to do things in an order I fine easier.)

This set of functions can now be closed under the operations of pointwise addition, additive inverse, multiplication and multiplicative inverse which produces the set of all rational functions. It can be closed as well under differentiation (which adds no new members, since the derivative of a rational function is another rational function) and integration, expressed as the unary operator

$$I(f(x)) = \int_0^x f(t)\,dt$$

with the usual convention that the negative of the integral is taken and the bounds reversed when $x < 0$. This last step has added new functions to the set, and with them come new numbers. The function

$$\int \frac{1}{x^2+1} dx$$

for instance, is not a rational function and most of its values are not rational numbers.

Once some initial analysis has been done to demonstrate that these operations can actually be expected to be well defined, these sets, both the numbers and the functions can be manipulated by the purely formal methods characteristic of algebra. No questions need necessarily be asked about questions such as convergence of limits, and the sets do not need to be treated as topological spaces. They can be understood strictly in terms of the interactions between the operations given; addition commutes with integration, while multiplication does not. Differentiation and multiplication are not commutative, but there is a more complex symbolic relationship, and so forth.

In fact, this is not the only interesting countable set built by these operations. Just as the rational numbers are closed under multiplication and addition, but not all polynomials constructed from them by these operations have solutions, this set does not necessarily contain the solutions to all equations created from the given operations. It is reasonable to consider the set of all functions which can solve an equation of the form

$$\mathcal{O}(f) = 0$$

where $\mathcal{O}$ is any operator constructed by finite combination of addition, subtraction, multiplication, division and members of the set of functions itself which can act meaningfully on a single other function. These operators are the logical generalization of the polynomials on the rational numbers to this space of functions, and there can only be countably many of them. If each one is only required to have a single, or even a countable number of solutions then the set of functions produced and the set of numbers they are defined over will both remain countable.

There are some difficulties with this notion. If any possible functions which satisfy these operator equations pointwise are allowed, there are not only a countable number of solutions to each one. $\mathcal{O}(f) = f^2 - 1$, for instance, has solutions $f = 1$ and $f = -1$, but also an uncountable number of hybrid choices which take on one value some places and the other elsewhere. Also, $f' = 0$ has as its solutions every possible constant function. Some method for selecting from these possibilities must be found before these sets can be constructed.

If these difficulties are avoided, this procedure will produce a pair of sets, one of numbers and one of functions, whose properties can be described algebraically, and which are countable. The set of functions will be a neglidgable subset of the analytic functions with coefficients from the set of numbers. By the standards of analysis, such a set would be startlingly small. None the less, it contains a huge number of the functions which are actually used, including the solutions to all one dimensional differential equations, and essentially every function used outside of pure mathematics is at least constructed from members of it piecewise.