Marshall University
Marshall Digital Scholar

Theses, Dissertations and Capstones

1-1-2011

Infinitesimal Time Scale Calculus

Tom Cuchta
cuchta@marshall.edu

Follow this and additional works at: http://mds.marshall.edu/etd

Part of the Number Theory Commons

Recommended Citation

INFINITESIMAL TIME SCALE CALCULUS

A thesis submitted to

the Graduate College of

Marshall University

In partial fulfillment of

the requirements for the degree of

Master of Arts in Mathematics

by
Tom Cuchta

Approved by

Dr. Bonita A. Lawrence, Ph.D., Committee Chairperson
Dr. Carl Mummert, Ph.D.
Dr. Suman Sanyal, Ph.D.

Marshall University
May 2011

ACKNOWLEDGMENTS

I thank all the wonderful friends who have been there through the good and bad times these past two years, in particular I'd like to acknowledge Cody Baker, Brittany Balcom, Paul "Zahck" Blazer, Kaci Evans, Shannon Gobel, Madison LaScola, Lisa Mathis, Jessica McClure, Aaron Payne, Ben Plunkert, Aaron "F." Statler, Adam Stephenson, and Brittany Whited.

I thank Dr. Clayton Brooks for his helpful advice over the years.

I thank Dr. Suman Sanyal for being on my committee.

I thank Dr. Carl Mummert for being on my committee and helpful advice.

I especially thank Dr. Bonita Lawrence for being on my committee and introducing me to the world of mathematical research and differential equations with her differential analyzer project.

Last but not least, I want to thank my family for unbounded support and love!

"No way of thinking, however ancient, can be trusted without proof."
- Henry David Thoreau, *Walden*

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

ABSTRACT

Calculus has historically been fragmented into multiple distinct theories such as differential calculus, difference calculus, quantum calculus, and many others. These theories are all about the concept of what it means to "change", but in various contexts. Time scales calculus (introduced by Stefan Hilger in 1988) is a synthesis and extension of all the various calculi into a single theory. Calculus was originally approached with "infinitely small numbers" which fell out of use because the use of these numbers could not be justified. In 1960, Abraham Robinson introduced hyperreal numbers, a justification for their use, and therefore the original approach to calculus was indeed logically valid. In this thesis, we combine Abraham Robinson's hyperreal numbers with Stefan Hilger's time scale calculus to create infinitesimal time scale calculus.

1. INTRODUCTION

Mathematics is the detailed study of the logical consequences of assumption. It is as ancient as human society, and the *method* by which mathematics is done was collected into a book called "The Elements" by the ancient Greek mathematician Euclid [8]. Euclid's book is primarily about the mathematics of geometry and his assumptions (called *axioms*) were mostly obvious properties that geometrical figures have such as "between any two points there is a straight line" [3]. Mathematicians have since studied much more than geometry, but the tradition of stating axioms and then investigating those axioms logically has continued ever since Euclid. We will follow that tradition. This thesis is designed to be as self-contained as possible; we have tried to justify every axiom we assume with a reason that attempts to make every assumption obvious.

This thesis is organized as follows. Section 2 begins the discussion with an introduction to logic and sets. To do mathematics requires us to study assumptions logically, so we have written a short section on how this is done. We treat mathematics as if it were a language like English, and the axioms of sets in Section 2 will be justified by appeal to properties of the English language.

We then move to defining familiar mathematical objects in Section 3. We first define numbers that are used to count and then use those to successively define negative numbers, fractions, real numbers, and the hyperreals (include "infinitely large" and "infinitely small" numbers). We define numbers from scratch using sets rather than positing their properties and taking on faith that they exist. This approach

allows us to be entirely certain that what we are talking about is indeed realistic in the framework of the language of mathematics, a problem which has historically been an issue with using infinitely large or small numbers in practice. For example, in his 1734 work *The Analyst*, the philosopher George Berkeley criticized infinitely small numbers by describing them as "ghosts of departed quantities" [4], effectively dismissing such numbers as pseudoscience. Berkeley's criticism has since been shown to be inaccurate with a construction of the hyperreals by the mathematician Abraham Robinson in 1960 [7]. We also develop the theory of slope, which is a way to assign numerical values to steepness. The numerical method of slope will then be used to study the steepness of general curves via the calculus.

Section 4 focuses on two different ways slope is generalized as is done in the mathematical fields of differential calculus and difference calculus. Our approach to differential calculus is not the standard "limit" approach which developed in response to the criticisms of the infinitely small, but rather it is an approach using "infinitesimal" numbers in the original spirit of the inventors of calculus, Issac Newton (1687) and Gottfried Leibniz (1684). Both Newton and Leibniz battled critics on the legitimacy of infinitesimals for much of their careers, but Newton inevitably gave in to critics while Leibniz defended their use [9]. The mathematics of Robinson allows us to rigorously use infinitesimals in calculus, which validates Leibniz's defense of infinitesimals.

The reader will notice that the two theories developed in Section 4 have multiple similarities which we investigate in Section 5. Historically these similarities were not investigated in depth, so differential calculus (continuous), difference calculus (discrete), and many other calculi were developed separately with roughly the same theoretical framework but all with different fundamental definitions. In 1988, Stefan

Hilger unified these calculi into a single theory in his dissertation on abstract structures called measure chains [5]. In 2001, Martin Bohner and Allan Peterson released a textbook which considers not abstract measure chains but simpler structures called time scales [2]. Thus the time scales calculus is a unification and extension of many various calculi that eliminates the multiplicity of calculi theories by developing them all simultaneously. Our particular approach to time scale calculus is new and based on infinitesimal numbers. We believe the infinitesimal approach is intuitive, easy to understand, and falls in line with the tenants of time scales calculus much better than the limit approach in [2].

2. THE LANGUAGE OF MATHEMATICS

2.1. LOGIC

In the English language, to write the sentence you are currently reading requires choosing a carefully chosen sequence of words and grammatical symbols arranged in a particular order according to the "laws of English grammar". These laws are patterns found by linguists that summarizes the "proper" form of the language in order to standardize and study it. Although, historically, standardization occurred *after* the existence of the English language, that is precisely where we begin with our theory of mathematics. Such an approach allows our language to be precise (unlike English) while still retaining many structural properties of language. When writing sentences in the basic language of mathematics, we call our mathematical nouns *sets* which must follow laws of mathematical grammar: *logic.*

Specifications of the alphabet and technical grammar of this language are beyond the scope of this thesis. We will describe an intuitive form of logic sufficient to comprehend proofs. We will agree upon an alphabet and grammar for the language of mathematics: the English language equipped with the ability to rename objects at will. A *sentence* is defined to be an arrangement of some of the letters of the alphabet according to the rules of grammar which is either true or false, called its *truth value.*

Sentences may be complicated, but simple sentences exist such as "11 is an even number", so all one must do to find its truth value is compare the definition of "even number" to the specific object denoted by the symbols "11". For that sentence,

recall that a number is an *even number* if it is twice another number and then ask the question "Is 11 twice another number?", whose answer yields the truth value of the sentence in question. The easiest way to proceed from here is to compute a list of all the even numbers which are less than or equal to 11, which we do in Table 2.1.

Number	Doubled
1	2
2	4
3	6
4	8
5	10
6	12

Table 2.1. Even numbers up to 12.

This list must contain all of such even numbers, but 11 is not on the list. This is sufficient evidence to conclude that the sentence "11 is an even number" is false. Sentences that are easily verifiable like the one above are called *simple sentences*; as we investigate sentences in more detail, we will begin using place holder symbols to stand in for the actual sentence themselves. Sometimes we would like to take a sentence and consider the sentence with opposite truth value. This is formalized with the negation operator, $\neg$. For example, if p is a sentence and p is true, then $\neg p$ is also a sentence, but $\neg p$ is false.

Some sentences are created from two simpler parts, such as "11 is an even number and the Sun is a star". Sentences that behave this way are called *compound sentences*. They are combined by the conjunction "and", which in English implies that for the sentence "11 is an even number and the Sun is a star" to be true, constituent parts "11 is an even number" and "the Sun is a star" must both be true. We call words that combine simple sentences into compound ones *connectives*. The

properties of the "and" connective is summarized in Table 2.2, and is called a *truth table*.

p	q	p and q
T	T	T
T	F	F
F	T	F
F	F	F

Table 2.2. Truth table for "and".

In Table 2.2, p and q are abbreviations for two sentences and the entries down the columns are arranged so that each possible combination of True and False between Column p and Column q appears down the rows. It is important to realize that we are not fixing specific sentences p and q, but exploring the effects of combining arbitrary sentences of any possible truth values with the word "and". For example, in row 3, the first column tells you to consider the case when p is false, the second column tells you to consider the case when q is true, and the third column tells you the truth value of "p and q" as a sentence itself when the conditions are met.

p	q	p or q	if p, then q	p if and only if q
T	T	T	T	T
T	F	T	F	F
F	T	T	T	F
F	F	F	F	T

Table 2.3. Truth table for the connectives "or", "if...then", and "if and only if".

The "and" connective is one such way to create compound sentences, but it is not the only way. Three examples of common connectives follow: the "or" connective results in a compound sentence that is true when at least one of its constituent parts is true, the "if p, then q" connective describes conditional sentences, and the "if and

only if" connective behaves as a logical "equals" showing when two different sentences express the same truth values. The behavior of these connectives is described in Table 2.3.

A combination of sentences by connectives remains a sentence, so it has a particular truth value itself. This is a powerful observation because it allows us to circumvent potentially complicated language restrictions by simply switching our perspective on the logical structure of the statement. For example, Table 2.4 shows the *logical equivalence* of the sentence "if p, then q" to the sentence "if $\neg q$, then $\neg p$" which means that the values in each of the respective columns of these statements in a truth table match up exactly. Logical equivalence is useful because it allows substitution of one statement for another that behaves the same, but may be easier to understand. So, for any sentence whatsoever in our language wherein "if p, then q" occurs, we could substitute "if $\neg q$, then $\neg p$ " and not change the truth value of the original statement. We could write the logical equivalence of p and q efficiently using the "if and only if"connective: " 'if p, then q' if and only if 'if $\neg q$, then $\neg p$' ", a true sentence. This type of analysis on sentences demonstrates that although p and q could be any two sentences whatsoever, we can still reason about how they combine and deduce properties of their combinations even though we don't know what they actually say.

p	q	$\neg p$	$\neg q$	if p, then q	if $\neg q$, then $\neg p$
T	T	F	F	T	T
T	F	F	T	F	F
F	T	T	F	T	T
F	F	T	T	T	T

Table 2.4. "if p, then q" is logically equivalent to "if $\neg q$, then $\neg p$ ".

The logic of mathematics forms a solid method of interpreting the sentences we use. It has been designed to mimic how people actually speak but, unlike casual conversation, the truth values of sentences in mathematics are always strictly monitored, since if they were not, we may say something false. Although this logic is inherent in every declarative sentence we write, we do not explicitly dwell on the logical underpinnings of our theorems very often.

2.2. SETS

Sets are the words of the mathematical language which are used by the logic to create meaningful statements. The primary purpose of a set is to encapsulate the objects it contains as an object itself, which is in fact loosely mimicked in the English language.

Imagine the objects constituting an egg: the shell, yolk, and albumen (white of the egg). They exist as separate entities "shell", "yolk", and "albumen", but when considered together, somehow these items are "egg" to an English-speaker. If we wanted to express this relationship in the language of sets, we would write $egg = \{shell, yolk, albumen\}$ i.e., egg is the set "containing" only $shell$, $yolk$, and $albumen$. This analogy should give a clear indication of what "contains" means in terms of sets: the objects a set "contains" are simply those that are "inside" of it.

In the prior example of $egg = \{shell, yolk, albumen\}$ we let the string of symbols $yolk \in egg$ to represent the sentence "$yolk$ is an object in the set egg", which can be shortened to "$yolk$ is in egg". In general, given a set X and an element y in the set X, we will write $y \in X$.

If X is a set containing only a, b, and c, we may define X by listing its objects as $X = \{a, b, c\}$. This approach is problematic if the list we wish to write is infinite.

We fix that problem by allowing sets to be defined by a sentence descriptor. For example, if we want to talk about the set of even numbers, we could try to list them and get something like *even numbers* $= \{0, 2, 4, \ldots\}$ or we could go a more language-oriented route and write *even numbers* $= \{x \colon x \text{ is a multiple of } 2\}$. The string of symbols $\{x \colon x \text{ is a multiple of } 2\}$ describes the set of even numbers by first defining a variable of consideration x and then stating x has a property that is logically equivalent to x being an even number. To separate the variable description part from the condition being enforced part, we use the symbol " $:$ ". In this case, the entire string of symbols is saying "*even numbers* is the set of all x such that x is a multiple of 2", which by its description must be the set of even numbers.

We now begin an investigation of sets and, to do so, we need to have one to work with. So we simply state the existence of a set as an *axiom* (also called an assumption). To not state this axiom leaves us without a concrete logical reason for any set to exist in the first place. The set we say exists is called the *empty set* and we will construct the entire language of sets and numbers themselves from it.

Axiom 2.1. *There exists at least one set, called the empty set, denoted by $\emptyset$. Moreover, $\emptyset$ contains no elements.*

By stipulating that at least one set exists, we have given ourselves a place to start. However, once we start we have nowhere else to go, so we seek inspiration. A frequently used process in mathematics is the process of considering a collection of distinct objects that already exist as an object itself. That process is quite natural, and it is evident (albeit rather informally) in naming almost every noun in English: *bag of barbeque chips* $= \{bag, \text{ } barbeque \text{ } chips\}$ where *barbeque chips* $= \{fried \text{ } potato,$ *seasoning*$\}$ and *seasoning* $= \{salt, \text{ } barbeque\text{-}flavored \text{ } powder\}$ and *salt* $= \{sodium \text{ } ion,$ *chlorine ion*$\}$ and *sodium ion* $= \{10 \text{ } electrons, \text{ } 11 \text{ } protons, \text{ } 11 \text{ } neutrons\}$ and so forth.

So to create the noun "*bag of chips*", first form the molecules that constitute the object, then arrange the molecules spatially into a pattern that will be recognized according to the definition of *bag of chips*. Likewise, to form a new set, we take existing sets and form them into new ones. We formalize this process with our next axiom.

Axiom 2.2. *If A and B are sets, then there exists a set C such that $A \in C$ and $B \in C$.*

This axiom says that we may combine sets that already exist in any way we wish. Since we started with only $\emptyset$, notice that we can immediately say that $\{\emptyset, \emptyset\}$ is a set by Axiom 2.2 when $A = \emptyset$ and $B = \emptyset$. Let us compare $\{\emptyset, \emptyset\}$ with a similar looking situation from English, $\{yolk, yolk\}$. Such a set would be relevant in the case of a special egg with two yolks: $special\ egg = \{shell, yolk, yolk, albumen\}$. In English, an egg with this property is unusual but is still considered an egg, which reinforces the notion that sets should be containers of not just items, but *distinct* items. That means we want $special\ egg = egg$, and with our definitions, this means $\{shell, yolk, yolk, albumen\} = \{shell, yolk, albumen\}$. In short, listing $yolk$ twice does not change *which* objects are inside of egg. This type of reasoning leads us to the conclusion $\{\emptyset, \emptyset\} = \{\emptyset\}$. To ensure that all of our sets have this property, we claim that two sets are "the same" if they contain exactly the same distinct objects, and if they have repeats, the repeated ones are ignored.

Axiom 2.3. *Two sets are equal if they contain the same objects.*

We now accept the convention that when referring to any set, we use the equivalent one with distinct elements in its list. This convention simplifies the language and does not change the logical outcome of our system. Moving on, we still have

$\{\emptyset, \emptyset\}$ which we now see by Axiom 2.3 equals $\{\emptyset\}$. By a similar procedure, given any set A we can form the set containing A (written as $\{A\}$) by first applying Axiom 2.2 to the set A to form $\{A, A\}$ and then using Axiom 2.3 to enforce our convention $\{A, A\} = \{A\}$.

Extending this reasoning, we can now form $\{\{\emptyset\}\}$ and $\{\emptyset, \{\emptyset\}\}$ by using Axiom 2.2 on the existing sets $\emptyset$ and $\{\emptyset\}$. Now with any of these sets that exist, we can create a multitude of new sets by Axiom 2.2. The next batches of sets are summarized in Table 2.5.

Time	New Sets
Axiom 2.1	$\emptyset$
Generation 1	$\{\emptyset\}$
Generation 2	$\{\{\emptyset\}\}$; $\{\emptyset, \{\emptyset\}\}$
Generation 3	$\{\{\{\emptyset\}\}\}$; $\{\{\emptyset\}, \{\emptyset\}\}$; $\{\emptyset, \{\{\emptyset\}\}\}$; $\{\emptyset, \{\emptyset, \{\emptyset\}\}\}$; $\{\{\emptyset\}, \{\{\emptyset\}\}\}$; $\{\{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$; $\{\{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}\}$; $\{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}\}$; $\{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}\}$; $\{\emptyset, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}\}$; $\{\{\emptyset\}, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}\}$; $\{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}, \{\emptyset, \{\emptyset\}\}\}$

Table 2.5. The generations of sets.

Notice in Table 2.5 that both the sets $\emptyset$ and $\{\emptyset\}$ are in the set $\{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}\}$. Sometimes it will be useful to make statements of this sort, and we will do it frequently enough that it is convenient to develop a shorthand to say it. In this case we would say that $\{\emptyset, \{\emptyset\}\}$ is a *subset* of $\{\emptyset, \{\emptyset\}, \{\{\emptyset\}\}\}$. If A is a set and B is a set such that every element of A is in B, we say that A is a subset of B (in shorthand, $A \subset B$). At this juncture, we should remind ourselves why we began our investigation of sets: to create a language of mathematics. We could continue on with this process of set generation, but it will not take us anywhere interesting. Instead, we plan to describe numbers with sets. To do so, we need a set with infinitely many elements. The axioms so far don't permit this to happen at any particular generation of sets because we are bound by our axioms to take only single steps to create new

sets. Since there are infinitely many numbers, if our language of sets is to be powerful enough to talk about them, we need to admit sets with infinitely many elements. We will conveniently state that all sets created as above together form a set.

Axiom 2.4. *All sets created by Axiom 2.2 at some generation (starting with Axiom 2.1) form an infinite set.*

Given any two sets A and B, we say that B is the *power set* of A if and only if B only contains all subsets of A. So far, we can only declare the finite sets to have power sets, because we can explicitly construct the elements of the power set from the elements of any finite set in a finite number of steps. For example, the subsets of $\{\emptyset\}$ are $\emptyset$ and $\{\emptyset\}$, so $\mathcal{P}(\{\emptyset\}) = \{\emptyset, \{\emptyset\}\}$ which is indeed a set by Table 2.5. With finite steps, we cannot construct the power set of the infinite set guaranteed to exist by Axiom 2.4. Therefore we compensate with the following axiom.

Axiom 2.5. *Given a set A, the power set $\mathcal{P}(A)$ exists.*

We can now create as many infinite sets as we wish by taking power sets of the set guaranteed to exist by Axiom 2.4. However, let X be some infinite set and pick some element $x \in X$. We have no logical reason to believe that removing the element x from the set X will result in a well defined set. Following our intuition that a set is a container, we recognize that removing an item from a container does not destroy the container, it only makes the amount of its contents smaller. Thus, removing something from a set should generate a "smaller" set. As of now, our axioms do not allow us to claim that these smaller subsets always exist, but our intuition says that they should, so we clarify this with the following axiom.

Axiom 2.6. *Every possible subset of any set exists as a set.*

In English, we have words like "first" and "second" which denote a definite order, and we would like to mimic this behavior with sets. So, if the set a is to be "first" and the set b is to be "second", then we encapsulate this order as a new set called an *ordered pair*. We will define the ordered pair of a first and b second to be the set $\{a, \{a, b\}\}$. This construction may appear strange, but it contains all the information we need under the proper interpretation: we are looking at a set where a is first and b is second, where we denote b being second by including a in the set containing b. This definition is somewhat tricky to use in practice, so we invent notation that appeals to our intuition and represent $\{a, \{a, b\}\}$ as the string of symbols (a, b) which clearly shows a first and b second. We could not initially define the ordered pair in this way, because we had no interpretation in previously defined terms for what it would mean.

2.3. RELATIONS AND FUNCTIONS

You are related to your parents. The growth of bacteria is related to how much food it consumes. The strength of an explosion is related to the amount of TNT used to create it. The noun (therefore set) *relation(ship)* is an abstract property that two distinct objects may have. Your parents have with you a parent-child relation(ship), which we will shorten to $\sim_p$. Suppose that your name is $Suzy$ and your parents' names are Ed and $Jane$. We could then write $Suzy \sim_p Ed$ to express the sentence "the (ordered) pair $(Suzy, Ed)$ has the $\sim_p$ property of being in a parent-child relation" and $Jane \sim_p Suzy$ to express "the pair $(Jane, Suzy)$ has the $\sim_p$ property." Both of those relationships are true, but consider $Jane \sim_p Ed$, which is false, in which case we would write $Jane \nsim_p Ed$, which is true.

A relation itself is a set, so we must decide which set it is. If we want to say a is related to b by some relationship (denoted by $\sim$), then we can let $\sim$ be a set containing elements (a, b) and define in general an element of the set $(x, y) \in \sim$ if and only if x is related to y.

Set equality is a relation, but we conveniently write it as $A = A$. More generally, we can use the properties of $=$ to define relations that behave similar to the $=$ relationship. We say a relation $\equiv$ is an *equivalence relation* if it is symmetric (that is, $x \equiv y$ if and only if $y \equiv x$), reflexive (for all x, $x \equiv x$), and transitive (if $x \equiv y$ and $y \equiv z$, then $x \equiv z$). Let X be a set and suppose that $y \in X$. If $\equiv$ is an equivalence relation, we sometimes will consider all the objects equivalent to y as a set itself. So we define the *equivalence class of y* to be the set $\{x \in X : x \equiv y\}$. The relation $\sim_p$ is not an equivalence relation: it is neither transitive since $Jane \sim_p Suzy$ and $Suzy \sim_p Ed$ but $Jane \not\sim_p Ed$, nor reflexive since $Suzy \not\sim_p Suzy$.

Consider the sentence $2 + 1 = 3$ of arithmetic. When we read it, we sense a "process" that somehow transforms the string of symbols $2 + 1$ into the number 3. This process is in fact the addition relation. If we wanted to think of the symbol $+$ as a relation, we could say that the pair $(2, 1)$ has the $+$ relationship of addition to the number 3 i.e., $((2, 1), 3) \in +$. If that is our definition, then "$2 + 1 = 3$" is a language convention that we adopt to make the addition relation easier to read.

We can think of the $+$ relation as having the input $(2, 1)$ and the output 3. This one input-one output behavior is extra property that we frequently use: as a "process", $+$ takes a pair of numbers to exactly one other number, called their sum. The important fact about this process is that it does not take any pair of numbers to more than one defined sum, so each pair of numbers has exactly one sum. This extra

property is that of being a *function*, and we will use relations with this property very frequently.

Definition 2.7. Let f be a relation and suppose $(x, y) \in f$. The relation f is a function if for all $z \neq y$ it is true that $(x, z) \notin f$.

By convention, we do not write $(x, y) \in f$ every time we want to express that x and y are related by the function f. We could write xfy, but it would be difficult to read. We appeal to our intuition of a function as an input-output machine, so when we describe the function, we would like to say a phrase such as "with input x, the function f outputs y" and we translate this language into the symbols as $f(x) = y$.

Functions take unique inputs to single outputs. Sometimes we would like to talk about the inputs alone or the outputs alone as sets themselves. These are specifically the sets

$$\mathrm{dom}(f) = \{x \colon f(x) \text{ exists}\}$$

called the *domain of f* and

$$\mathrm{rng}(f) = \{y \colon \text{ there is an } x \in \mathrm{dom}(f) \text{ such that } f(x) = y\}$$

called the *range of f*. All functions have both a domain and a range, so they will be useful concepts in describing how functions behave and interact.

3. THE STRUCTURE OF MATHEMATICS

3.1. NATURAL NUMBERS

Number is the abstract concept of quantity. We will define numbers and their operations as sets via the axioms of arithmetic, called *Peano arithmetic.* To start building numbers we want to aim for a set that contains all of our numbers; we give the set we are constructing a name, $\mathbb{N}$, the natural numbers (or, counting numbers). Mimicking the first axiom of sets, we stipulate that $\emptyset$ serve as the number zero.

Axiom 3.1. $\emptyset \in \mathbb{N}$ *and as a number, we refer to $\emptyset$ as 0.*

It is important to realize that there is a difference between the actual set that is the number zero ($\emptyset$) and the name of the number zero (0). This distinction is purely for aesthetics; we could very well use the sets representing numbers as the numbers themselves, but we would eventually get very frustrated with the notation. We need a way to construct the next natural number, 1, and those that follow; we take care of constructing all natural numbers and giving them a friendly name in a single axiom.

Axiom 3.2. *There is a function $S : \mathbb{N} \to \mathbb{N}$ called the* successor function *such that* $S(n) = n \cup \{n\}$ *consecutively defines elements of $\mathbb{N}$. We say that* $S(0) = 0 \cup \{0\} = \{0\} = 1$ *and* $S(1) = 1 \cup \{1\} = \{0, 1\} = 2$ *and in general* $S(n) = \{0, 1, \ldots, n-1, n\} = n + 1.$

This axiom simultaneously defines new alphabet symbols (the numerals) and gives them an associated set by repeatedly invoking Axiom 2.2 on the element in $\mathbb{N}$

that is guaranteed by Axiom 3.1. We are creating a language for the number system that we are naturally comfortable with inside of the existing language of sets. There is no end to number succession, so we can say that $\mathbb{N}$ is an infinite set, which we know exists by applying Axiom 2.6 to Axiom 2.4.

The natural numbers have a very clear *order*: 1 is the first number generated, 2 is the second, 3 is the third, and so forth. So we recognize that, in a very real sense, 2 is in *less than* 3, since $2 \in 3$ as sets. We want to capture this property of number in the language of sets so that we can use it in mathematics. We do this by defining the order on $\mathbb{N}$ to be the following relation: if $a, b \in \mathbb{N}$, then $a < b$ if and only if $a \in b$ as sets. So $<$ is a set and $(a, b) \in <$ if and only if $a \in b$.

For any set X of natural numbers, you can imagine there is always a smallest natural number in X. We believe this because, given any $n \in X$, either n is the smallest element of X or some number less than n is the smallest. Since there are only finitely many numbers smaller than any given n, we simply have to check those numbers and see if there are any smaller than n in X. This is what we call the *well-ordering principle* which we state as an axiom.

Axiom 3.3. *Given any nonempty $X \subset \mathbb{N}$, there exists an $n \in X$ such that for all $m \in X$, $n < m$ or $n = m$.*

Now we want to define addition and multiplication on natural numbers. Consider the sum $3 + 2$. We think of this sum as literally combining three objects with two other objects which should result in five objects. This is a sort of "high-level" approach, since it assumes familiarity with the combinational properties of the numbers 2 and 3. But to interpret all sums in this way would require a lot of work. For example, not many people can rattle off the solution of $4212354235234 + 12121253748865$ instantly. To deal with such sums we need a general process with which we can

compute. So let us consider $3 + 2$ again. What we do is replace the left term with its equivalent set theoretical definition of $S(S(S(0)))$ and then use Axiom 3.2. Consider the computation

$$\begin{aligned}
S(S(S(0))) + 2 &= S(S(0)) + 1 + 2 \\
&= S(S(0)) + 3 \\
&= S(0) + 1 + 3 \\
&= S(0) + 4 \\
&= 0 + 1 + 4 \\
&= 5.
\end{aligned}$$

To generalize this process and compute any sum $m + n$, all we must do is realize that $m = S(m - 1)$. According to Axiom 3.2 we see that $m = (m - 1) + 1$. Now,

$$\begin{aligned}
m + n &= (m - 1) + 1 + n \\
&= (m - 1) + 1 + n \\
&= (m - 1) + S(n),
\end{aligned}$$

which when repeated $m - 1$ times yields a zero for the left term and the value of the sum as the right term. This definition does not aid in actual hand computation because traditional methods are much quicker, but it does serve to define addition in terms of prior results. Multiplication is defined simply as repeated addition as below.

Definition 3.4. Let $m, n \in \mathbb{N}$. The sum $m + n$ is defined recursively as

$$m + n = (m - 1) + S(n)$$
$$= (m - 2) + S(S(n))$$
$$= \ldots,$$

until the process stops, which it must do. Similarly, the product $m \cdot n$ is defined as

$$m \cdot n = \sum_{i=1}^{m} n = n + n + n + \ldots + n,$$

where there are m copies of n in the sum.

So, to compute the product $3 \cdot 2$ we write $3 \cdot 2 = \sum_{i=1}^{3} 2 = 2 + 2 + 2 = 6$. As with addition, this process is more complicated to use than the traditional hand calculation methods, but it is sufficiently robust to serve as a definition.

3.2. INTEGERS

Integers make sense of the concepts of *positive numbers*, numbers greater than 0, and *negative numbers*, numbers less than 0. We do not have negative numbers in $\mathbb{N}$ so we give the new set, the integers, the name $\mathbb{Z}$ ("zahlen" is German for "number"). Integers add a new component to number: its *sign*, which is the choice of "negative" or "positive"for a given integer. To interpret a natural number as an integer, we pick a number and then give it a sign, a choice to distinguish whether you have the positive or negative "copy" of it.

Starting with the natural numbers, we would like to extend our set to include another copy of the natural numbers along with a sign. Fortunately, this process

requires no new axioms. Just as we interpreted natural numbers as sets, we will do the same for integers. If z is an integer, then we represent z as an ordered pair $z = (z_1, z_2)$ for $z_1, z_2 \in \mathbb{N}$ where z_1 is called the *positive part of z* and z_2 is called the *negative part of z*. For example, we let the number 0 be represented by the ordered pair $(0, 0)$ which has positive component 0 and negative component 0, we let the number $1 \in \mathbb{Z}$ be represented by the ordered pair $(1, 0)$ which has positive component 1 and negative component 0, and we let the number $-1 \in \mathbb{Z}$ be represented as $(0, 1)$ which has positive component 0 and negative component 1.

Now, using the "obvious" way to add these numbers, we have

$$(1, 0) + (0, 1) = (1 + 0, 0 + 1) = (1, 1),$$

which is somewhat unexpected, because we expected the result $(0, 0)$. Interpreting $(1, 1)$ we see that it has positive part 1 and negative part 1. Since each of its parts are the same, we imagine that its positive and negative parts wipe each other out to have the net result of zero. So, in fact, there are many ways to write the integer 0: for all natural numbers $n \in \mathbb{N}$, we have $0 \in \mathbb{Z}$ represented by the ordered pair (n, n).

We would like to represent $0 \in \mathbb{Z}$ as a single symbol, not infinitely many, which we do with an equivalence class. So consider the equivalence relation $(a, b) \sim_{\mathbb{Z}} (c, d)$ if and only if $a + d = b + c$. Intuitively, we are saying that two integers are "the same" if and only if adding the positive part of one to the negative part of the other yields the same value. So, for example, we have $(0, 0) \sim_{\mathbb{Z}} (n, n)$ because $0 + n = n + 0$. We would like a way to think about all the possible ways to write each integer as a single symbol instead of a pair of symbols, so we define the integer $z \in \mathbb{Z}$ with

representation (z_1, z_2) to be $z \equiv \{(a, b) \in \mathbb{N} \times \mathbb{N} \colon (a, b) \sim_{\mathbb{Z}} (z_1, z_2)\}$, the set of integers equivalent to z. Thus an integer is really the set of integers equivalent to it.

3.3. RATIONAL NUMBERS

Rational numbers make sense of fractional parts. As it stands, we do not have symbols in our language to represent fractional parts, but we appeal to the construction of the set of integers for inspiration. We will represent the fraction $\frac{a}{b}$ as the ordered pair (a, b) such that $a, b \in \mathbb{Z}$ and $b \neq 0$. We call the first component of (a, b) to be its *numerator* and the second component as its *denominator*. This appears to be similar to the definition of integers with one key difference. The elements of our ordered pairs are integers, whereas they were previously natural numbers. Thus, as sets, we think of a single rational number as an ordered pair of equivalence classes of ordered pairs of natural numbers.

With this definition, we see that $\frac{1}{2} = (1, 2)$. We know how addition "should work": $(1, 2) + (1, 2) = (2, 2)$ which must be equivalent to 1. Intuitively, the rational number 1 has representation $(1, 1)$ which implies to us that there is an equivalence relation. Specifically we want to capture the behavior showing that $(2, 2)$ is the same number as $(1, 1)$. To do so we define the equivalence relation $\sim_{\mathbb{Q}}$ such that $(a, b) \sim_{\mathbb{Q}} (c, d)$ if and only if $a \cdot d = b \cdot c$. Intuitively, we are saying two rational numbers are "the same" if multiplying the numerator of one representation by the denominator of the other representation always yields the same value. For example, this definition encodes the equivalence $\frac{4}{2} = \frac{2}{1}$ because $4 \cdot 1 = 2 \cdot 2$.

Just as with integers, we would like a singular element of $\mathbb{Q}$ to actually "be" the fraction $\frac{q_1}{q_2}$, while also retaining the ability to think about, for example, $\frac{2 \cdot q_1}{2 \cdot q_2}$ as the same fraction. To accomplish this, we define the fraction $\frac{q_1}{q_2} \in \mathbb{Q}$ as

$\frac{q_1}{q_2} \equiv \{(a,b) \in \mathbb{Z} \times \mathbb{Z} \colon (a,b) \sim_{\mathbb{Q}} (q_1, q_2)\}$. Thus a fraction is really the set of fractions equivalent to it.

3.4. REAL NUMBERS

3.4.1. Construction. Trying to solve the equation $x^2 = 2$ with the set $\mathbb{Q}$ is impossible, since its solution is $\{\sqrt{2}, -\sqrt{2}\}$, and neither of those solutions is in $\mathbb{Q}$. We believe such a simple equation should be solvable, so we must extend our numbers to allow possibilities like $\sqrt{2}$. We do so with the set $\mathbb{R}$ of *real numbers* which contain both the rationals with finite or repeating decimal expansions and a set called the *irrational numbers* which have infinitely long nonrepeating expansions. The real numbers have two important properties: they are a *field* and have the *least upper bound* property.

Example 3.5. We can express the number $\sqrt{2}$ with the infinite decimal representation

$$1.41421356237309504880168872420097\ldots$$

We can represent any finite initial portion of this infinite nonrepeating decimal with a rational number. For example, we let $q_1 = 1$, $q_2 = 1.4$, $q_3 = 1.41$, $q_4 = 1.414$, and so forth to define better and better rational approximations to the number $\sqrt{2}$. We can interpret these numbers as a function: define $q \colon \mathbb{N} \to \mathbb{Q}$ so that $q(i) = q_i$. This function makes sense of incremental approximations to the real number $\sqrt{2}$.

The function q is an example of a *sequence*, which is simply a function whose domain is $\mathbb{N}$. To simplify notation, we may write q_i for $q(i)$ when q is a sequence. The sequence q has a special property: if we pick any number $\epsilon > 0$, say, $\epsilon = 0.0001$, we can always find a natural number N so that for all $m, n > N$, $|q_m - q_n| < \epsilon$. In

our example of $\sqrt{2}$, for $\epsilon = 0.0001$, if we pick $N = 10$, then if $m, n \in \mathbb{N}$ such that $m > 10$ and $n > 10$, we can easily see that $|q_m - q_n| < \epsilon$, since both numbers match each other in the first 10 decimal places. This discussion motivates the following definition by the mathematician Augustin Louis Cauchy [8].

Definition 3.6. Let x be a sequence. Then x is a *Cauchy sequence* if and only if for every $\epsilon > 0$ there exists some $N \in \mathbb{N}$ so that for all natural numbers $m, n > N$, $|x_n - x_m| < \epsilon$.

Cauchy sequences approximate real numbers, just as our sequence q approximated $\sqrt{2}$. When a sequence has the property that it approximates a number, we say that the sequence *converges*. A sequence that converges is one that is approaching some fixed value, although we may not know what that value is. There is no number in the set $\mathbb{Q}$ that q converges to. So we are constructing the set $\mathbb{R}$ in part because it has the property that all Cauchy sequences inside of it converge inside of it.

We know that the Cauchy sequence q constructed from $\sqrt{2}$ does not converge in $\mathbb{Q}$ (if it did, then $\sqrt{2}$ would be a rational number), but it seems that it gets closer and closer to $\sqrt{2}$ and indeed would converge to $\sqrt{2}$ if $\sqrt{2}$ was a rational number. This makes us believe that we could represent the number $\sqrt{2}$ with the sequence q itself. However, this approach does not exactly work, because there are multiple ways to converge to $\sqrt{2}$. For example, if $p_1 = q_2$, $p_2 = q_4$, $p_3 = q_6$, ... (that is, p is a sequence of every other term of q), then p also represents $\sqrt{2}$, but with a different sequence.

We will now begin to construct the real numbers, $\mathbb{R}$. First, begin with the set $R = \{x \colon x$ is a Cauchy sequence of rational numbers$\}$. We will turn this set R into the real numbers, but first we need to define addition and multiplication. Let $x, y \in R$, then $x = x_1, x_2, \ldots$ and $y = y_1, y_2, \ldots$ are Cauchy sequences. Define

addition as

$$x + y = (x_1, x_2, \ldots) + (y_1, y_2, \ldots) = (x_1 + y_1, x_2 + y_2, \ldots)$$

and multiplication as

$$x \cdot y = (x_1, x_2, \ldots) \cdot (y_1, y_2, \ldots) = (x_1 \cdot y_1, x_2 \cdot y_2, \ldots).$$

We now proceed with equivalence classes as we did in Section 3.2 and Section 3.3. Let $x, y \in R$ be Cauchy sequences of rational numbers. Then define the relation $x \sim_{\mathbb{R}} y$ if and only if the sequence $z_n = x_n - y_n$ converges to zero. This definition says that we consider two elements of R to be the same if they try to converge to the same place (whether or not the place they are trying to converge exists in $\mathbb{Q}$ or not). Now, for any $\hat{z} \in R$, we define the real number $z \in \mathbb{R}$ as $z = \{x \in R \colon x \sim_{\mathbb{R}} \hat{z}\}$ so that $\hat{z}$ converges to z. Thus, the real number z is the set of Cauchy sequences equivalent to the given Cauchy sequence $\hat{z}$.

3.4.2. Properties of $\mathbb{R}$. Now that we have the real numbers, we can properly define what it means to converge.

Definition 3.7. Let $x \colon \mathbb{N} \to \mathbb{R}$ be a sequence. Then x converges if and only if there exists a number $\bar{x} \in \mathbb{R}$ so that for every $\epsilon > 0$, there exists some $N \in \mathbb{N}$ with $N > 0$ and for all $n \in \mathbb{N}$ with $n > N$, we have $|x_n - \bar{x}| < \epsilon$.

Let $X \subset \mathbb{R}$ and let x be a convergent sequence, with $x_i \in X$ for all i, that converges to the number $\bar{x} \in X$. We call $\bar{x}$ a *limit point* of X because there exists a sequence x so that $x \to \bar{x}$. If $z \in X$, then we know z is a limit point of X because the sequence $x_1 = z, x_2 = z, x_3 = z, \ldots$ converges to z. However, not all limit points

of X necessarily lie in X. Example 3.5 shows us that if $X = \mathbb{Q}$, then $\sqrt{2}$ is a limit point of $\mathbb{Q}$ but $\sqrt{2} \notin \mathbb{Q}$.

The set of real numbers has many useful properties. Among these properties are a description of the ways we may manipulate sums and products of real numbers contained in the following theorem. We do not prove these properties here, but reference the reader to [11].

Theorem 3.8. (*Field Axioms*) *The set $\mathbb{R}$ is a* field, *that is, a set with addition and multiplication and the following properties for all $a, b, c \in \mathbb{R}$:*

(*i*) *Commutativity:* $a + b = b + a$ *and* $a \cdot b = b \cdot a$,

(*ii*) *Associativity:* $a + (b + c) = (a + b) + c$ *and* $a \cdot (b \cdot c) = (a \cdot b) \cdot c$,

(*iii*) *Zero: there is a unique number 0 such that $0 \cdot r = 0$ and $0 + r = r$ for all $r \in \mathbb{R}$,*

(*iv*) *One: there is a unique number $1 \neq 0$ such that $1 \cdot r = r$ for all $r \in \mathbb{R}$ and $r \cdot \frac{1}{r} = 1$ for all $r \neq 0$,*

(*v*) *Inverses:* $a + (-a) = 0$ *and* $a \cdot \frac{1}{a} = 1$ *except for* $a = 0$,

(*vi*) *Distribution:* $a \cdot (b + c) = a \cdot b + a \cdot c$.

Sets of the form $A = \{x \colon x < 1\}$ seem to have a "maximum" element 1, but it is really not a maximum element of this set, since 1 is not inside A. We say instead that 1 is the *least upper bound* (or *supremum*) of A (i.e., $\sup A = 1$) which means that 1 is the least of all numbers bigger than those in the set. A nonempty set $A \subset \mathbb{R}$ ordered by $<$ is called *bounded above* when there exists some $y \in \mathbb{R}$ (called an *upper bound*) such that for all $x \in A$, $x < y$. We do not prove here that $\mathbb{R}$ has the least upper bound property, but reference the reader to [11].

Theorem 3.9. (*Completeness Axiom*) *The set $\mathbb{R}$ has the* least upper bound *property, that is, for every nonempty bounded above set $A \subset \mathbb{R}$, there is a number, $\sup A$, such that for all upper bounds b of A, $\sup A \leq b$.*

Similarly, a nonempty set A is said to be *bounded below* when there exists some $y \in \mathbb{R}$ (called a *lower bound*) such that for all $x \in A$, $y < x$. So we define the *greatest lower bound* (or *infimum*) to be the number $\inf A = \sup\{x \colon \text{for all } y \in A, x < y\}$. We require that A be nonempty because there is no greatest lower bound of $A = \emptyset$ since all numbers by definition are lower bounds of $\emptyset$. However, when applying the theory of real numbers to calculus, we sometimes (see Section 5.2) interpret $\inf \emptyset$ to be a particular value but such a choice is done purely for convenience of argument or notation.

3.5. HYPERREAL NUMBERS

3.5.1. Construction. The set of *hyperreal numbers* is an extension of the real numbers which makes the idea of an infinitely small number logically rigorous. The early developments of calculus were approached with free use of infinitely small numbers, but their use was criticized on grounds of logical justification until using them fell out of academic style. This changed in 1960 when Abraham Robinson developed nonstandard analysis and made rigorous the original approach to calculus by Newton and Leibniz [7]. In 1976, Jerome Keisler published a book on elementary calculus on the infinitesimal numbers [6] which motivated some developments in Section 5.

Calculus naturally leads one to desire division by zero which we cannot do without logical inconsistencies in arithmetic (such as $0 = 1$), so we will instead divide by a non-zero element called an *infinitesimal* which has enough zero-like properties to satisfy our situation. Let us describe the properties an infinitesimal should have. A positive infinitesimal should be less than any positive real number (like zero is), but is greater than zero (which allows us to divide by it). We can also speak of infinite

numbers: if $\epsilon > 0$ is a positive infinitesimal, then for all real numbers $r > 0$ we have $\epsilon < r$; therefore $\frac{1}{\epsilon} > r$ and we define numbers of the form $\frac{1}{\epsilon}$ as *infinite numbers* (similarly, a negative infinitesimal $\delta < 0$ admits the negative infinite number $\frac{1}{\delta}$).

We will define the set of hyperreals by construction. A very rigorous construction of the hyperreal numbers can be found on [4, p.23]. We let $\mathbb{R}^{\mathbb{N}}$ denote the set of infinite sequences of real numbers. If $x \in \mathbb{R}^{\mathbb{N}}$, then $x = (x_0, x_1, \ldots)$ for real numbers $x_i \in \mathbb{R}$. We represent the real number x inside of $\mathbb{R}^{\mathbb{N}}$ as $(x, x, x, \ldots)$. We do the arithmetic componentwise and in each component we inherit the algebra of $\mathbb{R}$: if $x, y, z \in \mathbb{R}^{\mathbb{N}}$ such that $x = (x_1, x_2, \ldots), y = (y_1, y_2, \ldots)$, and $z = (z_1, z_2, \ldots)$, then

$$z \cdot (x + y) = (z_0 \cdot (x_0 + y_0), z_1 \cdot (x_1 + y_1), \ldots)$$
$$= (z_0 x_0 + z_0 y_0, z_1 x_1 + z_1 y_1, \ldots).$$

Consider $r \in \mathbb{R}^{\mathbb{N}}$ such that $r = (0, 1, 1, \ldots)$. If these sequences are to be a useful extension of the real numbers, all nonzero elements should have a multiplicative inverse. However, that would require there to exist a number $r^{-1} \in \mathbb{R}^{\mathbb{N}}$ so that

$$r \cdot r^{-1} = (0 \cdot r_0^{-1}, 1 \cdot r_1^{-1}, \ldots)$$
$$= (1, 1, 1, \ldots)$$

which is impossible because by Axiom 3.8 there does not exist a real number r_0^{-1} such that $0 \cdot r_0^{-1} = 1$. Any element of $\mathbb{R}^{\mathbb{N}}$ containing a zero will not have a multiplicative inverse for the same reason. We want to treat r as if it were "similar to" 1 so we need a way to explain how this number has the property that it is "similar to" $(1, 1, 1, \ldots)$. Specifically, we say that $(0, 1, 1, \ldots)$ is similar to $(1, 1, 1, \ldots)$ because they are equal

in all except the first component, that is "almost all" of its components are equal to 1.

We will describe what "almost all" means in terms of a set called a *filter*. Let I be a set. A *filter $\mathcal{F}$ over I* is a subset of $\mathcal{P}(I)$ such that $\emptyset \notin \mathcal{F}$, for all $A \in \mathcal{F}$ and $B \in \mathcal{F}$ the intersection $A \cap B$ is in $\mathcal{F}$ and when for any $A \in \mathcal{F}$ all sets C such that $A \subset C \subset I$ we have $C \in \mathcal{F}$.

Example 3.10. For $I = \mathbb{N}$, a simple example of a filter is the *co-finite subsets of $\mathbb{N}$*, which is the set $\mathcal{F}^{co} = \{A \subset \mathbb{N} \colon \mathbb{N} - A \text{ is finite}\}$. The set $\mathcal{F}^{co}$ is a filter because if $A, B \in \mathcal{F}^{co}$, then A and B are both infinite sets with the property that $\mathbb{N} - A$ and $\mathbb{N} - B$ are finite. Therefore, the intersection $A \cap B$ will have infinitely many elements but be missing those in the finite set $(\mathbb{N} - A) \cup (\mathbb{N} - B)$, so $A \cap B$ is a subset of $\mathbb{N}$ such that $\mathbb{N} - (A \cap B)$ is finite. Also, if $A \in \mathcal{F}^{co}$ and $A \subset C$, then C is a set bigger than A, meaning $\mathbb{N} - C$ is smaller than $\mathbb{N} - A$, and therefore $C \in \mathcal{F}^{co}$.

A filter $\mathcal{U}$ with the additional property that for all $A \subset I$, either $A \in \mathcal{U}$ or $I - A \in \mathcal{U}$ is called an *ultrafilter*. Our previous filter $\mathcal{F}^{co}$ is not an ultrafilter, for the set $\{0, 2, 4, 6, \ldots\} \notin \mathcal{F}^{co}$ and the set $\mathbb{N} - \{0, 2, 4, 6, \ldots\} = \{1, 3, 5, 7, \ldots\} \notin \mathcal{F}^{co}$ are both infinite.

Example 3.11. Let $i \in I$. The set $\mathcal{F}^i = \{A \subset I \colon i \in A\}$ is an ultrafilter because $\emptyset \notin \mathcal{F}^i$, if $A, B \in \mathcal{F}^i$ then $i \in A \cap B$, if $A \subset C$ then $i \in C$, and if $A \in \mathcal{F}^i$ then $i \notin I - A$. The ultrafilter $\mathcal{F}^i$ is called the *principal ultrafilter generated by i* because it focuses all attention on the single element i of I.

Principal ultrafilters are not useful for hyperreal numbers, because using them would force our intuition of "almost all" to always include the element i. Ultrafilters not of the type $\mathcal{F}^i$ are called *non-principal ultrafilters*. Non-principal ultrafilters cannot be described directly in standard set theory, so their existence is usually

demonstrated with a logical contradiction. We refer to Corollary 2.6.2 on [4, p.21] for proof that a non-principal ultrafilter exists over the set $\mathbb{N}$ and that $\mathcal{F}^{co}$ is a subset of any such non-principal ultrafilter. Since $\mathcal{F}^{co} \subset \mathcal{F}$, we see that no finite set must be in $\mathcal{F}$ since the complement of any set in $\mathcal{F}$ is not in $\mathcal{F}$.

Let $\mathcal{F}$ be a nonprincipal ultrafilter over $\mathbb{N}$. We will say $x \in \mathbb{R}^{\mathbb{N}}$ has a property P *almost everywhere* if $\{n \in \mathbb{N} : x_n \text{ has property } P\} \in \mathcal{F}$, which means that the property is true in infinitely many components of x. For example, we would say that $r = (0, 1, 1, \ldots)$ equals 1 almost everywhere because

$$\{n \in \mathbb{N} : r_n = 1\} = \{1, 2, 3, \ldots\} \in \mathcal{F}$$

since $\mathbb{N} - \{1, 2, 3, \ldots\} = \{0\}$ is a finite set showing that $\{1, 2, 3, 4, \ldots\} \in \mathcal{F}^{co} \subset \mathcal{F}$.

From now on we will abbreviate the set $\{n \in \mathbb{N} : f_n \text{ has property } P\}$ as $\|P(f)\|$. We use this notation in the following example.

Example 3.12. We will describe a particular infinitesimal number. Let $\epsilon \in \mathbb{R}^{\mathbb{N}}$ so that $\epsilon = (1, \frac{1}{2}, \frac{1}{3}, \ldots)$. Then for all positive $r \in \mathbb{R}, r = (r, r, r, \ldots)$ is a hyperreal number and

$$\|\epsilon < r\| = \{n \in \mathbb{N} : \epsilon_n < r_n\}$$
$$= \{m, m+1, \ldots\} \in \mathcal{F},$$

where $\epsilon_m = \frac{1}{m}$ is the first fraction such that $\frac{1}{m} < r_n = r$. Since r was an arbitrary real number and $\|x < r\| \in \mathcal{F}$, we see that $x < r$ almost everywhere, which we interpret as meaning x is a positive infinitesimal. It then follows that if $\omega = \{1, 2, 3, \ldots\}$, then $\|\omega > r\| \in \mathcal{F}$ (ω is an infinite number) and $\|\omega = \frac{1}{\epsilon}\| \in \mathcal{F}$ (ω is the reciprocal of ϵ). This is the reasoning that for an infinitesimal ϵ, why $\frac{1}{\epsilon}$ is an infinite number.

Unfortunately our construction has the same problem we have seen many times before. We have multiple representations of the same number. Consider the elements $a, b \in \mathbb{R}^{\mathbb{N}}$ such that $a = (1, 0, 0, \ldots)$ and $b = (0, 1, 0, 0, \ldots)$. Then,

$$\|a = b\| = \{n \in \mathbb{N} \colon a_n = b_n\}$$
$$= \{3, 4, 5, \ldots\} \in \mathcal{F}.$$

This fact means that two hyperreal numbers that are equal almost everywhere may have different representations, just as the fraction $\frac{2}{4}$ is a different representation of the fraction $\frac{1}{2}$. So we will create an equivalence relation of equality and then collect them into equivalence classes which we will call the hyperreal numbers. We define our equivalence relation, $\equiv$, on elements of $\mathbb{R}^{\mathbb{N}}$ such that if $x, y \in \mathbb{R}^{\mathbb{N}}$, $x \equiv y$ if and only if $\|x = y\| \in \mathcal{F}$. Now we consider the sets of our equivalent numbers as a single set; we represent all the elements of $\mathbb{R}^{\mathbb{N}}$ that are equivalent to x as $[x] = \{z \in \mathbb{R}^{\mathbb{N}} \colon z \equiv x\}$. The hyperreal numbers are simply the set $^*\mathbb{R} = \{[z] \colon z \in \mathbb{R}^{\mathbb{N}}\}$, the set of sets of equivalent elements of $\mathbb{R}^{\mathbb{N}}$.

3.5.2. Properties of $^*\mathbb{R}$. The set $^*\mathbb{R}$ is algebraically exactly as nice as $\mathbb{R}$. It turns out that $^*\mathbb{R}$ with operations inherited from $\mathbb{R}^{\mathbb{N}}$ is a field, just like $\mathbb{R}$. Proof can be found on [4, p.25].

Theorem 3.13. *The set $^*\mathbb{R}$ is a field.*

An important property of the hyperreal numbers is that they provide a framework to *extend* a subset of $\mathbb{R}$ to a subset of $^*\mathbb{R}$. Let $A \subset \mathbb{R}$. We define the *extension of A* as $^*A$ where $[r] \in {}^*A$ if and only if $\|r \in A\| \in \mathcal{F}$. Informally this means that if $r \in A$ almost everywhere, then $[r] \in {}^*A$. For example if $A = \mathbb{N} = \{0, 1, \ldots\}$, then

$[r] \in {}^*\mathbb{N}$ if and only if $\|r \in \mathbb{N}\|$. So, in particular, every element of $\mathbb{N}$ is in ${}^*\mathbb{N}$ and also infinite hyperreal numbers such as $\omega \in {}^*\mathbb{N}$ because $\|\omega \in \mathbb{N}\| \in \mathcal{F}$.

Let $f\colon \mathbb{R} \to \mathbb{R}$ be a function. We would like to know that f has a corresponding hyperreal-valued function, but do not yet know if f extends, because $f \subset \mathbb{R} \times \mathbb{R}$, not $f \subset \mathbb{R}$. It does extend to ${}^*f\colon {}^*\mathbb{R} \to {}^*\mathbb{R}$ if we define ${}^*f([r]) = [s]$ if and only if $\|(f(r_1), f(r_2), \ldots) = s\| \in \mathcal{F}$. The $<$ relation on real numbers extends to the ${}^*{<}$ relation on hyperreals by way of $x \, {}^*{<} \, y$ if and only if $\|x < y\| \in \mathcal{F}$.

The discussion above demonstrates that the hyperreal numbers have a lot of nice properties. They contain the real numbers, sets of reals extend to sets of hyperreals, real-valued functions extend to real-valued functions, and the order relation on $\mathbb{R}$ is preserved. In the following theorem we summarize these properties similarly to [6, p.28].

Theorem 3.14. *(Extension Principle) At least one infinitesimal exists in ${}^*\mathbb{R}$. The real numbers are a subset of the hyperreal numbers $(\mathbb{R} \subset \mathbb{R}^*)$, every set $A \subset \mathbb{R}$ has a unique extension to a set ${}^*A \subset {}^*\mathbb{R}$, all functions $f\colon \mathbb{R} \to \mathbb{R}$ extend to a unique function ${}^*f\colon {}^*\mathbb{R} \to {}^*\mathbb{R}$, and the order relation $<$ extends to the order relation ${}^*{<}$.*

We adopt the convention to drop the star and write ${}^*{<}$ as $<$. By Theorem 3.13 we can form sums such as

$$5 + \epsilon = (5 + 1, 5 + \frac{1}{2}, 5 + \frac{1}{3}, \ldots)$$
$$= (6, \frac{11}{2}, \frac{16}{3}, \ldots),$$

which is a well defined hyperreal number. This number is not an infinitesimal; it is a finite number plus an infinitesimal. We think of $5 + \epsilon$ as being "infinitely close" to the number 5 because $(5 + \epsilon) - 5 = \epsilon$, in other words, 5 is different from $5 + \epsilon$

by an infinitesimal amount, an amount less than any positive real number. This suggests that to determine when two hyperreal numbers are "infinitely close", we should inspect their difference. So if $x, y \in {}^*\mathbb{R}$, we say that x is *infinitely close* to y if and only if $x - y$ is an infinitesimal. So define the equivalence relation $\sim$ such that for hyperreal numbers x and y, $x \sim y$ if and only if $x - y$ is an infinitesimal. If $x \in \mathbb{R}$ and $x \neq 0$, then we say that $x + \epsilon$ is a *finite hyperreal*. We use these concepts in the following axiom.

Axiom 3.15. *If $x \in {}^*\mathbb{R}$, then there is exactly one real number that is infinitely close to x. So there exists a function* st *(called the standard part function) from the finite hyperreals to the reals so that* $\text{st}(x)$ *is that unique real number which x is infinitely close to.*

We began to develop the hyperreal numbers to aid us in developing calculus. Axiom 3.15 allows us to use the hyperreal numbers in the calculation part of calculus and after that, we go back to the real numbers with the standard part function. This way of doing calculus greatly simplifies it, at the cost of using a different number system. The standard part function has nice properties that allow us to use it in a variety of situations. It turns out that standard part of a sum of the standard parts.

Lemma 3.16. *If $x, y \in \mathbb{R}^*$, then $st(x + y) = \text{st}(x) + st(y)$.*

Proof. By definition, $st(x+y) = x+y+\delta$ for some infinitesimal δ. So for appropriate infinitesimals δ_x and δ_y, we have

$$\text{st}(x + y) = \text{st}(x) + \delta_x + \text{st}(y) + \delta_y + \delta,$$
$$\text{st}(\text{st}(x + y)) = \text{st}(\text{st}(x) + \delta_x + \text{st}(y) + \delta_y + \delta),$$
$$\text{st}(x + y) = \text{st}(x) + \text{st}(y). \qquad \square$$

Similarly, the standard part of a product is the product of the standard parts.

Lemma 3.17. *If $x, y \in {}^*\mathbb{R}$, then $st(x \cdot y) = \text{st}(x) \cdot \text{st}(y)$.*

Proof. By definition, $st(x \cdot y) = x \cdot y + \delta$ for some infinitesimal δ. So, for appropriate infinitesimals δ_x and δ_y, we have

$$\text{st}(x \cdot y) = xy + \delta,$$
$$\text{st}(x \cdot y) = (\text{st}(x) + \delta_x) \cdot (\text{st}(y) + \delta_y) + \delta,$$
$$\text{st}(x \cdot y) = \text{st}(x) \cdot \text{st}(y) + \text{st}(x) \cdot \delta_y + \text{st}(y) \cdot \delta_x + \delta_y \cdot \delta_x + \delta,$$
$$\text{st}(st(x \cdot y)) = st(\text{st}(x) \cdot st(y) + \text{st}(x) \cdot \delta_y + \text{st}(y) \cdot \delta_x + \delta_y \cdot \delta_x + \delta),$$
$$\text{st}(x \cdot y) = \text{st}(x) \cdot \text{st}(y). \qquad \square$$

It will be very useful later to refer to the set of hyperreal numbers that are equivalent to a given hyperreal number. To do this requires us to pick an element of the set $\mathcal{P}({}^*\mathbb{R})$, which we know exists because of Axiom 2.5. So define the function $\text{halo} \colon {}^*\mathbb{R} \to \mathcal{P}({}^*\mathbb{R})$ by $\text{halo}(x) = \{y \colon y \sim x\}$. By definition, the halo of any finite hyperreal number will contain exactly one real number by Axiom 3.15.

The following lemma is useful when we want to switch perspective from an inequality of real numbers to a relationship of hyperreals. Suppose $x, y \in {}^*\mathbb{R}$ and $x - y$ is finite and noninfinitesimal, then we expect that $st(x) - st(y)$ is nonzero, because if it were zero, we would expect $x - y$ to be infinitesimal. This is intuitively reasonable, because a real-numbered distance is greater than any infinitesimal distance. We prove this in the following lemma.

Lemma 3.18. *Let $x, y \in {}^*\mathbb{R}$. If $\operatorname{st}(x) > \operatorname{st}(y)$, then $x \not\sim y$.*

Proof. Suppose that $\operatorname{st}(x) > \operatorname{st}(y)$ and $x \sim y$. Then, we can write $x = \operatorname{st}(y) + \delta$ for infinitesimal $\delta = x - y$. But then, $\operatorname{st}(x) = \operatorname{st}(\operatorname{st}(y) + \delta)$ and by Lemma 3.16, we see that $\operatorname{st}(x) = \operatorname{st}(y)$, a contradiction. $\square$

Many facts about reals can be easily rewritten as facts about the hyperreals. So once a sentence about real numbers is in the "correct form" we can "transfer" that sentence into one about hyperreals. Axiom 3.14 guarantees that the sets involved in some sentence about sets of real numbers themselves extend properly but says nothing about whether the sentences themselves are true.

Example 3.19. Let $A = [0, 1]$. Consider the sentence "there exists $u \in \mathbb{R}$ such that for all $z \in \mathbb{R}$ with the property 'for all $x \in A, z \geq x$', we know $u \leq z$", which is a definition of u being a least upper bound, so $u = \sup A = 1$. Thus the sentence is a particular case of Axiom 3.9 (Completeness Axiom) for the set A. The set ${}^*A$ has the same property, that is "there exists $u \in {}^*\mathbb{R}$ such that for all $z \in {}^*\mathbb{R}$ with the property 'for all $x \in {}^*A, z \geq x$', we know $u \leq z$". In this case, it is still true that $u = \sup {}^*A = 1$.

Now consider the sentence "for all $A \subset \mathbb{R}$, there exists $u \in \mathbb{R}$ such that for all $z \in \mathbb{R}$ with the property 'for all $x \in A, z \geq x$', we know $u \leq z$". The difference between this sentence and the sentence in Example 3.19 is that this one is prefaced by the words "for all $A \subset \mathbb{R}$" which means this sentence is about all subsets of $\mathbb{R}$. This sentence is the actual Completeness Axiom of the reals, but it does not transfer. This is because the sentence "for all $A \subset {}^*\mathbb{R}$, there exists $u \in {}^*\mathbb{R}$ such that for all $z \in {}^*\mathbb{R}$ with the property 'for all $x \in A, z \geq x$', we know $u \leq z$" is demonstratably false.

Proposition 3.20. *The set of infinitesimals,* halo(0)*, does not have a least upper bound.*

Proof. Suppose that there is a number $z = \sup \text{halo}(0)$ and we will see that this leads to a logical contradiction. Since there are infinitesimals greater than zero, we know that in particular, $z > 0$. Either $z \in \text{halo}(0)$ or $z \notin \text{halo}(0)$, which would mean that $z \in {}^*\mathbb{R} - \text{halo}(0)$. If $z \in \text{halo}(0)$, then z is not an upper bound because $2z$ is an element of halo(0) bigger than z, so z is not even an upper bound of halo(0). If $z \in {}^*\mathbb{R} - \text{halo}(0)$, then $\frac{z}{2}$ is a smaller upper bound than z is, a contradiction since $z = \sup \text{halo}(0)$. This contradiction will occur with any $z > 0$ that is not an infinitesimal. Therefore we have proved $z \notin \text{halo}(0)$ and $z \notin {}^*\mathbb{R} - \text{halo}(0)$, which means that there is no least upper bound for halo(0). □

This indicates that not all sentences about real numbers will transfer, only some of them. To rigorously define the transferable sentences requires defining a brand new language of formulas and then a proof that all formulas constructable in the language will transfer properly. Such constructions are beyond the scope of this thesis, but can be found in [4, p.47]. In short, if the sentence involves a particular $A \subset \mathbb{R}$ and some number of phrases of the form "for all $x \in A$" or "there exists $x \in A$" and some properties these variables have, then it is transferable. Phrases of the form "for all $A \subset \mathbb{R}$" or "there exists $A \subset \mathbb{R}$" will not necessarily transfer, so we avoid them. Once a sentence S is written out entirely in a proper form, we transfer it with the *-transform* of S which is the sentence formed by replacing each set and function appearing in S with their unique hyperreal extensions.

Axiom 3.21. (*Transfer Principle*) *Sentences involving allowable quantifiers about real numbers can be extended to sentences about hyperreal numbers.*

Now we know when some facts about $\mathbb{R}$ correspond to facts about $^*\mathbb{R}$. In particular, the theory of calculus will require us to use the suprema of sets of real numbers and Axiom 3.21 guarantees that those suprema remain suprema in the hyperreals. This fact will be especially useful in Section 5.

3.6. SETS OF NUMBERS

An arbitrary set of numbers has little structure beyond being simply that. It will be difficult to do mathematics on such a set, so we offer various criteria sets of numbers can have so that they are better behaved. Some other examples of sets of numbers are $\{2, \pi, 4, 9, 11\}$, $\{\frac{1}{n} : n \in \mathbb{N}\}$ and $\mathbb{R}$ itself. Let $a, b \in \mathbb{R}$. Another common set of numbers is $\{x \in \mathbb{R} : a \leq x \leq b\}$, and it is so common that we create a notational convention $[a, b]$ to represent it. It turns out that $[a, b]$ is especially nice compared to a set like $\mathbb{R}$ because $[a, b]$ is *bounded*, i.e., for every $x \in [a, b]$, there exists $M \in \mathbb{R}$ such that $[a, b] \subset [x - M, x + M]$.

We will also need to specify some topological properties of sets of numbers for later use. Specifications on how numbers in the set converge are topological properties. Let $A \subset \mathbb{R}$. We say that A is *closed* if and only if it contains all of its limit points. The property of being closed tells us that when sequences of elements of X converge, they converge to a point inside of X.

Theorem 3.22. *If $K \subset \mathbb{R}$ is closed and bounded, then $\inf K \in K$.*

Proof. Let $t_1 \in K$. Since t_1 and $\inf K$ are real numbers, the midpoint is the number $\frac{t_1 + \inf K}{2}$. So choose $t_2 \in [\inf K, \frac{t_1 + \inf K}{2}] \cap K$. If it turns out that the set is empty, let $t_2 = t_1$. In general, given t_k pick $t_{k+1} \in [\inf K, \frac{t_k + \inf K}{2}] \cap K$ (and if it turns out that the set is empty, let $t_{k+1} = t_k$). The sequence $\{t_k\}$ converges to $\inf K$. Therefore, since K is closed, $\lim t_n = \inf K \in K$. $\qquad\square$

Let $A \subset \mathbb{R}$ and let $x \in A$. Eventually, in Section 5, we will be seeking a way to pick a "next" element of A after x. To understand this notion, we will need the following lemma.

Lemma 3.23. *Suppose that $A \subset \mathbb{R}$, $x \in A$, and that $\{t \in A : t > x\}$ is nonempty. Then, $\inf\{t \in A : t > x\} = x$ if and only if there exists some sequence $a \colon \mathbb{N} \to A$ such that $a_n \to x$.*

Proof. ($\Rightarrow$) Suppose no such (a_n) exists. Then there is some $\epsilon \in \mathbb{R}$ such that $\epsilon > 0$ and $(x-\epsilon, x+\epsilon) \cap A = \emptyset$. But then, it must be true that $\inf\{t \in A : t > x\} \geq x+\epsilon > x$, which contradicts our hypothesis.

($\Leftarrow$) Suppose that $z = \inf\{t \in A : t > x\} > x$. Let $\epsilon = z - x$. But, by definition of (a_n) converging to x, we have that there exists some $N \in \mathbb{N}$ such that for all $n > N$, $|a_n - x| < \epsilon$. Therefore, $\inf\{t \in A : t > x\} \neq z$ since there are numbers smaller than z which are also lower bounds, which is a contradiction. $\square$

The following lemma is proven in [4, p.114]. It will also be useful in Section 5.

Lemma 3.24. *Let $A \subset \mathbb{R}$ and let $x \in \mathbb{R}$. Then, x is limit point of A if and only if $\mathrm{halo}(x) \cap {}^*A - \{x\}$ is nonempty.*

These lemmas present structural components of the sets $\mathbb{R}$ and ${}^*\mathbb{R}$. Both lemmas present properties that limit points of sets of numbers have. We will use Lemma 3.23 in conjunction with Lemma 3.24 while proving Lemma 5.5, a structural component of the theory of time scales calculus.

3.7. EQUATIONS

Arithmetic is elegant, but it cannot answer a question such as "what number, when added to 2, equals 5?" without further analysis. The question can be restated

in the language of algebra as "Solve the equation $x + 2 = 5$". We must now define what we mean by equation, what we mean by "x", and what it means to solve one.

An *equation* is a sentence which expresses the equality of two expressions. The algebraic equation $x + 2 = 5$ has a *variable* x which is a letter designed to range over all elements in a set of numbers, while the differential equation $y' = -y$ has y as its variable which ranges over all elements over a set of functions. To *solve* an equation is to discover the set $\{x : x \text{ makes the equation a true sentence}\}$ where x is the variable of the equation.

For example, to solve the equation $3x + 2 = 5$ we can think of the variable as ranging over the domain of the function $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) = 3x+2$. Consider a horizontal line on the plane where the line is made of points representing the real numbers. Perpendicular to this line is a vertical copy of this original line. Every point on the first line is designated $(r, 0)$ for some real number r and every point on the second line is designated $(0, s)$ for some real number s. Now for all $r, s \in \mathbb{R}$ we name the rest of the points in the plane $p = (r, s)$ where p is the point vertical to $(r, 0)$ and horizontal to $(0, s)$. To *graph* the function $f(x)$, all we must do is plot the points $(x, f(x))$ such that $x \in \text{dom}(f)$ as in Figure 3.1.

The result is a straight line. This motivates the definition of functions of the form $f(x) = ax + b$ for $a, b \in \mathbb{R}$ as *linear functions*. We will see that linear functions behave very nicely, and that understanding more complicated curves as we understand the straight line requires some new ideas.

3.8. SLOPE

Steepness is a quality that describes many objects and processes in nature. A person can ascertain that one hill is steeper than the other by walking the same
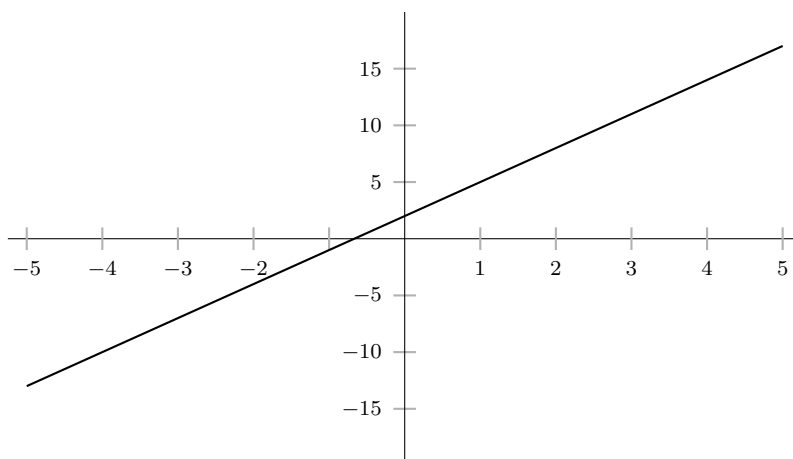
Figure 3.1. The graph of the function $f(x) = 3x + 2$.

distance up both hills from rest state and then comparing how tired (s)he is after both walks. If one examines the growth of bacteria in two Petri dishes, the one that grows more bacteria in the same timespan is said to be growing more quickly. In short, these situations of steepness can be described with the simple idea of the slope of a function.

Slope is a numerical interpretation of "steepness". We think of slope as measuring the change between two points of a function. This interpretation leads us to consider a difference $f(t_{new}) - f(t_{old})$. If the same change of a function f occurs in less time, we would say that the function is changing more quickly. This motivates us to divide by the time it takes for the change to occur to get the formula $\frac{f(t_{new}) - f(t_{old})}{t_{new} - t_{old}}$. This is the formula for slope.

Definition 3.25. If $\mu > 0$, then the slope a function f from x to $x + \mu$ is equal to the number $\frac{f(x + \mu) - f(x)}{\mu}$.

We now compute the slope of every point of a linear function and discover that it is constant, powerful evidence that linear functions are simple to understand.

Theorem 3.26. *If $f(x) = ax + b$, then the slope of f is a.*

Proof. Let $f(x) = ax + b$ be a linear function. Now compute

$$
\begin{aligned}
f'(x) &= \frac{f(x+h) - f(x)}{h} \\
&= \frac{(a(x+h) + b) - (ax + b)}{h} \\
&= \frac{ah}{h} \\
&= a,
\end{aligned}
$$

which is a constant. $\square$

This theorem shows us that the well understood set of linear functions could be a simple case of a potentially more general theory: do functions exists without constant slope? We will see that they do by taking the formula from Definition 3.25 for slope and then modifying it for nonlinear functions, and then computing their slope. Section 4 first does this in the classical way via differential calculus and then again via difference equations and Section 5 does it in a general sense using time scales calculus.

4. CLASSICAL CALCULI

4.1. DIFFERENTIAL CALCULUS

4.1.1. Derivatives. Nature appears continuous. Leaves *gradually* turn from green to orange. Water *flows* downhill. Differential calculus gives us a language in which Nature's processes can be studied as continuous.

We learned in the last section that the slope of a linear function is constant. What would a function without a constant slope look like? We consider a specific example.

Example 4.1. If the slope of an unknown function f is $2t$, then at the point $t = 0$ we get a slope of 0, at the point $t = -1$ we get a slope of -2, and at the point $t = 1$ we get a slope of 2. It turns out that the slope of $f(t)$ is $2t$, then $f(t) = t^2 + c$ for some real number c. In Figure 4.1, we illustrate these slopes as tangent lines to the particular function when $c = 0$, $f(t) = t^2$.

This means that we want to find the slope of the tangent line of f at t. To do so in the set of real numbers requires $\mu = 0$, which by the definition of slope leads us to the algebraically undefined statement $\frac{f(t+0)-f(t)}{0} = \frac{0}{0}$. So we cannot let μ truly equal zero, but we may get as near to zero as we wish. That means that a smaller μ offers a better approximation to the actual slope. The hyperreal numbers are well suited for this task, since an infinitesimal ϵ is nearer to zero than any real number.
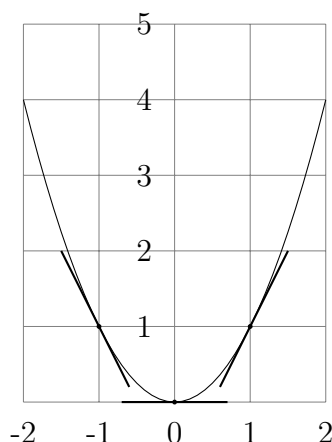
Figure 4.1. Graph of the slopes given by the function $f'(t) = 2t$.

Definition 4.2. Let ϵ be an infinitesimal. The *derivative* of a function $f : \mathbb{R} \to \mathbb{R}$ is defined to be the function $f'(t) = \text{st}(\frac{^*f(t+\epsilon) - ^*f(t)}{\epsilon})$. For the derivative to be well defined, the formula must yield the same value for all infinitesimals.

The derivative of f is denoted as f' and is also called the *first derivative*. The *second derivative* f'' is, if it exists, the derivative of f'. The *third derivative* f''' is the derivative of f''. In general, the nth derivative of f, if it exists, is denoted $f^{(n)}$, is the derivative of $f^{(n-1)}$.

Many functions have derivatives that are as simple to compute as the slope of the linear function.

Example 4.3. Let $f(t) = t^2$. Then we have, for any infinitesimal ϵ,

$$f'(t) = \text{st}\left(\frac{(t+\epsilon)^2 - t^2}{\epsilon}\right)$$
$$= \text{st}\left(\frac{2t\epsilon + \epsilon^2}{\epsilon}\right)$$

$$= \operatorname{st}(2t + \epsilon)$$

$$= 2t.$$

Therefore the nonlinear function $f(t) = t^2$ has slope $2x$ at a point $x \in \mathbb{R}$.

Being differentiable itself indicates something useful about the function: the function doesn't change "too much" as we vary its input slightly. The definition of the derivative does not seem to reflect this directly, so we must first define what "too much" change is. In short, if we vary the input to a continuous function f by an infinitesimal, then we expect the output of f to also vary by an infinitesimal. We borrow the definition from [6, p.125] to describe this relationship in terms of the relation $\sim$ on hyperreal numbers.

Definition 4.4. Let $A \subset \mathbb{R}$. A function $f : A \to \mathbb{R}$ is called *continuous* if and only if it has the following property: if $s, t \in \operatorname{dom}(^*f)$ and $s \sim t$, then $^*f(s) \sim {}^*f(t)$.

So to determine whether a real-valued function is continuous requires a computation using the hyperreal function $^*f$ guaranteed to exist by Axiom 3.14. Continuous functions have a nice property: they commute with the standard part function. This fact will allow us to simplify many calculations.

Lemma 4.5. *Let ϵ be an infinitesimal. If $f \colon \mathbb{R} \to \mathbb{R}$ is a continuous function then* $\mathrm{st}({}^{*}f(t + \epsilon)) \sim {}^{*}f(\mathrm{st}(t))$.

Proof. Let $x \in \mathbb{R}$, so $x = \mathrm{st}(x)$. Since f is continuous,

$$
{}^{*}f(t + \epsilon) \sim {}^{*}f(t) = {}^{*}f(\mathrm{st}(t)).
$$

Therefore,

$$
\mathrm{st}\left({}^{*}f(t + \epsilon)\right) \sim \mathrm{st}({}^{*}f(\mathrm{st}(t))) \sim {}^{*}f(\mathrm{st}(t)). \qquad \square
$$

Now we can properly state what we mean by "differentiable functions don't change too much". The proof boils down to the fact that when the derivative exists, ${}^{*}f(t + \epsilon) - {}^{*}f(t)$ is an infinitesimal. Our proof is inspired by the proof on [6, p.127].

Theorem 4.6. *If a function is differentiable, then it is continuous.*

Proof. Let f be a differentiable function and let $t \in \mathrm{dom}(f)$. Let ϵ be an infinitesimal. Then $t \sim t + \epsilon$, so consider Lemma 3.17 and the following:

$$
\begin{aligned}
{}^{*}f(t + \epsilon) - {}^{*}f(t) &\sim \mathrm{st}({}^{*}f(t + \epsilon) - {}^{*}f(t)) \\
&= \mathrm{st}\left(\epsilon \cdot \frac{{}^{*}f(t + \epsilon) - {}^{*}f(t)}{\epsilon}\right) \\
&= \mathrm{st}(\epsilon)f'(t) \\
&= 0.
\end{aligned}
$$

Thus ${}^{*}f(t + \epsilon) - {}^{*}f(t)$ is an infinitesimal and by definition, ${}^{*}f(t + \epsilon) \sim {}^{*}f(t)$, so f is continuous. $\qquad \square$

Unfortunately, some continuous functions are not differentiable everywhere on their domain. Continuity is therefore a weaker property than differentiability: there are more continuous functions than there are differentiable ones. We now give an example of a continuous function which is not differentiable because it is not differentiable at $t = 0$.

Example 4.7. We define the *absolute value* of $t \in \mathbb{R}$ to be $|t| = t$ when $t \geq 0$ and $|t| = -t$ when $t < 0$. This definition forces $|t| \geq 0$ for all $t \in \mathbb{R}$. Consider $f(t) = |t|$. If ϵ is a positive infinitesimal and $t = 0$, then $\text{st}(\frac{|0+\epsilon|-|0|}{\epsilon}) = 1$, but for negative infinitesimal δ and $t = 0$, we have $\text{st}(\frac{|0+\delta|-|0|}{\delta}) = \text{st}(\frac{|\delta|}{\delta}) = -1$. This is unfortunate because we cannot proceed to say that all continuous functions are differentiable. Continuity is nonetheless a useful property.

The derivative f' can tell us more than simply "f is a continuous function".

Theorem 4.8. *Let f be a differentiable function and $x \in \mathbb{R}$. Then, for all t, $f'(t) = 0$ if and only if f is a constant function, i.e., there is a $C \in \mathbb{R}$ so that for all $t \in \text{dom}(f)$, $f(t) = C$.*

Just as we created compound sentences out of simpler sentences in Section 2, we can form more complicated functions from simpler ones using various algebraic operators. By studying how this happens, we will be able to make many conclusions about how the derivatives of these new functions behave. We focus now on how a function can be defined as the product of two simpler functions. The functions in our example are all differentiable because their derivatives are defined for all $x \in \mathbb{R}$.

Example 4.9. Let $g(t) = 3t$ and $h(t) = 4t^2 + 1$. Then if $f(t) = 12t^3 + 3t$, we can write $f(t)$ as the product of functions $f(t) = g(t) \cdot h(t)$ because

$$g(t) \cdot h(t) = (3t) \cdot (4t^2 + 1) = 12t^3 + 3t$$

using Axiom 3.8 (*vi*).

In Example 4.9, $g(t)$ and $h(t)$ are differentiable functions and their product $f(t)$ is also a differentiable function. It turns out that if we knew $g'(t)$ and $h'(t)$, we can write $f'(t)$ in terms of $g(t), g'(t), h(t)$, and $h'(t)$ using a property of derivatives called the *product rule*. We will prove the product rule along with other convenient rules of differentiation in Theorem 5.11 in Section 5. For now, we state it so we can use it to solve a differential equation in the next section.

Theorem 4.10. (*Product Rule*) *Let $g(t)$ and $h(t)$ be differentiable functions so that $f(t) = g(t) \cdot h(t)$. Then $f'(t) = g(t) \cdot h'(t) + h(t) \cdot g'(t)$.*

We now focus our attention on a particular very well known and very well behaved differentiable function. There exists a number $e = 2.71828183\ldots \in \mathbb{R}$ such that if $f(t) = e^t$ (called an *exponential function*), then $f'(t) = e^t$. The function and number are proved to exist and have these properties in [6, pp.441-445]. The inverse function to the exponential is the Log function, so $\text{Log}(e^t) = e^{\text{Log}(t)} = t$. We sometimes write $\exp(t)$ for e^t to improve readability. For our purposes, this function will be used to investigate differential equations, and the property it has, $f'(t) = f(t)$, will be studied. Equations of this type are called differential equations, and have application in both pure mathematics and science.

4.1.2. Differential Equations. Differential calculus is used to investigate continuous processes of Nature. Suppose we would like to study how widgets change. In this section, we develop a language in which a description of how widgets change is used to find a function $y(t)$ that describes the widgets at time t. Understanding the nature of $y(t)$ may yield insight into how widgets work. Any possible observation made in the universe can be recorded with independent variable of time and can be studied in this way.

Example 4.11. If we are told, over a series of weeks, how the balance in our bank account has changed, but not told the account balance at any point, we could easily deduce how much money we had any given week. Consider the situation described in Table 4.1. If we knew only the information in the middle column and that we

Week	Change in account	Balance
1	-	$75
2	$-\$25$	$50
3	$+\$10$	$60
4	$+\$75$	$135
5	$-\$30$	$105

Table 4.1. A bank account.

had, say, $75 in the account initially, we could deduce the third column with simple arithmetic. In this case, the object of study is dollars in the bank, and the change we are studying is the sum of changes $-\$25 + \$10$ to the account.

The process of creating Column 3 in Example 4.1 is precisely that which we will use to solve differential equations in Section 5. For now, it demonstrates how we interpret derivatives when working with differential equations. With information of how a system is changing (how much my account changed, Column 2), we can make specific conclusions about the system (my balance, Column 3).

Definition 4.12. A *differential equation* is an equation involving a function y, its independent variable, and any number of its derivatives $y^{(1)}, y^{(2)}, \ldots, y^{(n)}$.

A differential equation is a description of a property that an unknown and desired differentiable function $y(t)$ may have. The equation $y' = -y$ is a differential

equation. It describes a function $y(t)$ with the property that $y'(t) = -y(t)$ for all $t \in \text{dom}(y)$. To find a function that meets this description is to *solve* the differential equation; the function found is called a *solution*. For example, the function $y_1(t) = 5e^{-t}$ is a solution to this differential equation, since $y_1'(t) = -5e^{-t} = -y_1(t)$. However, so is $y_2(t) = \pi e^{-t}$ and also $y_3(t) = -40000e^{-t}$. In fact any constant multiple of e^{-t} will be a solution. This behavior is not too surprising: we are very familiar with algebraic equations such as $x^2 - 1 = 0$ that have more than one solution.

When studying algebraic equations, we are not satisfied until all solutions are found, and this pathos is adopted for differential equations as well. So, we say a family of functions, $\mathcal{F}$, is the *general solution* of a differential equation if

$$\mathcal{F} = \{y \colon y \text{ is a solution to the differential equation }\}.$$

That is, all solutions of the differential equation are represented by the general solution. In the case of $y' = -y$, the general solution is

$$\mathcal{F} = \{y \colon \mathbb{R} \to \mathbb{R} \colon y(t) = Ce^{-t} \text{ for } C \in \mathbb{R}\}.$$

The arbitrary constant C gives this family the generality to include all solutions. From this point on, we will use the convention that a function defined with arbitrary parameters describes a family of functions $\mathcal{F}$ by including all valid parameter variations. So, we will say the general solution of $y' = -y$ is $y(t) = Ce^{-t}$, where C is an arbitrary real number.

It is nice to be told that this is the general solution, but how do we *know* that all the solutions are of the form $y(t) = Ce^{-t}$?

Theorem 4.13. *If $\phi(t)$ is an arbitrary solution to $y' = -y$, then $\phi(t)e^t = C$ for some constant C.*

Proof. We use Theorem 4.10 to differentiate and the fact that $\phi(t)$ is a solution to $y' = -y$ to say that $\phi'(t) = -\phi(t)$:

$$\begin{aligned}
\left(\phi(t)e^t\right)' &= \phi'(t)e^t + \phi(t)e^t \\
&= -\phi(t)e^t + \phi(t)e^t \\
&= 0.
\end{aligned}$$

Since the derivative of $\phi(t)e^t$ is zero, Theorem 4.8 guarantees that $\phi(t)e^t$ is a constant value C for all t. Thus it follows that $\phi(t) = Ce^{-t}$, i.e., the two functions are equal up to a constant factor C. $\qquad\square$

This theorem tells us that any function that is a solution to $y' = -y$ is equal to Ce^{-t} for a particular C. This is sufficient evidence to claim that $y(t) = Ce^{-t}$, where C is an arbitrary real number, is the general solution of $y' = -y$.

4.2. DIFFERENCE CALCULUS

Light can be thought of as traveling as particles. This means that the information received by our eyes is inherently not continuous, but a (very rapid) stream of light particles. Our brain interprets each of these particles as a piece of data and creates the images we see. When we use calculus to model real world information, the data we generate is inherently noncontinuous, so the idea that our mathematics of change absolutely must be continuous seems suspect. For this reason, we need a calculus which is suited for noncontinuous settings.
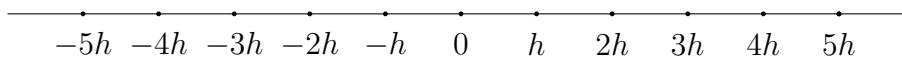
Figure 4.2. The set $h\mathbb{Z}$ for $h > 0$.

This noncontinuous calculus will need to be based on some sort of noncontinuous derivative. Luckily we have strong intuition that leads us to use the slope formula yet again, but since we are not doing calculus on the set $\mathbb{R}$, the denominator of our derivative will be a non-infinitesimal value. So, for $h > 0$ we define the set $h\mathbb{Z} = \{h \cdot z \colon h > 0 \text{ and } z \in \mathbb{Z}\}$, which is the set of points including zero equally spaced by the amount h along the real number line as in Figure 4.2.

Definition 4.14. Let h be a positive real number. The *difference operator* Δ of a function $y \colon h\mathbb{Z} \to \mathbb{R}$ at the point $t \in h\mathbb{Z}$ is denoted and defined by

$$\Delta y(t) = \frac{y(t+h) - y(t)}{h}.$$

We learned that the derivative of a function f is zero everywhere if and only if it is constant everywhere from Theorem 4.8. An analogous rule applies for the difference operator, and its proof will be given formally in Section 5.

Theorem 4.15. *Let $f \colon h\mathbb{Z} \to \mathbb{R}$. Then $\Delta f(t) = 0$ for all $t \in h\mathbb{Z}$ if and only if f is a constant function, i.e., there is some $C \in \mathbb{R}$ so that for all $t \in h\mathbb{Z}(f)$, $f(t) = C$.*

In Example 4.3 we computed the derivative of t^2. We now perform the difference operator on the same function and see what happens.

Example 4.16. Let $f(t) = t^2$. Then we have for any $h > 0$,

$$
\begin{aligned}
\Delta f(t) &= \frac{(t+h)^2 - t^2}{h} \\
&= \frac{t^2 + 2th + h^2 - h^2}{h} \\
&= 2t + h.
\end{aligned}
$$

This is a different result from that which we found in Example 4.3, but that is expected, since the difference operator is not the same as a derivative. This result reveals to us that calculus on $h\mathbb{Z}$ is different than calculus on $\mathbb{R}$.

In Example 4.9, we saw that a product of functions is still differentiable. After Example 4.9, we stated Theorem 4.10, the product rule for differentiable functions. Another rule we could have used is called the *quotient rule*, which applies to $f(t) = \frac{g(t)}{h(t)}$. We present the quotient rule for differences below, but wait until Theorem 5.11 to prove it.

Theorem 4.17. (*Quotient Rule*) *Let* $g, h \colon h\mathbb{Z} \to \mathbb{R}$ *so that* $f(t) = \frac{g(t)}{h(t)}$. *Then*

$$
\Delta f(t) = \frac{h(t) \cdot \Delta g(t) - g(t)\Delta h(t)}{h(t+1)h(t)}.
$$

A difference equation is a description of a property that an unknown and desired function $y(t)$ may have. The equation $\Delta y = -y$ is a difference equation. It describes a function $y(t)$ with the property that $\Delta y(t) = -y(t)$ for all $t \in h\mathbb{Z}$. To find a function that meets this description is to *solve* the difference equation, and the function found is called a *solution*. For example, the function $y_1(t) = (1-h)^{\frac{t}{h}}$ is a

solution to this difference equation, since

$$\Delta y(t) = \frac{(1-h)^{\frac{t+h}{h}} - (1-h)^{\frac{t}{h}}}{h}$$
$$= (1-h)^{\frac{t}{h}} \frac{(1-h) - 1}{h}$$
$$= -(1-h)^{\frac{t}{h}}$$
$$= -y(t).$$

However, so is $y_2(t) = (1-h)^{\frac{t-5}{h}}$ and also $y_3(t) = (1-h)^{\frac{t-\pi}{h}}$. In fact any function of the form $y(t) = (1-h)^{\frac{t-t_0}{h}}$ will work. The number t_0 is an important parameter of this solution called the *initial condition* which we will see is a crucial piece of information to get the particular solution that we want from the family of solutions of a differential equation.

Let us investigate whether this is truly the general solution of the difference equation like we did in Theorem 4.13.

Theorem 4.18. *If $\phi(t)$ is an arbitrary solution to $\Delta y = -y$, then $\frac{\phi(t)}{(1-h)^{\frac{t}{h}}} = C$ for some constant C.*

Proof. Following the definition of the difference operator and the fact that ϕ is a solution to $\Delta y = -y$, we use Theorem 4.17,

$$\Delta \left(\frac{\phi}{(1-h)^{\frac{t-t_0}{h}}} \right) = \frac{(1-h)^{\frac{t-t_0}{h}} \Delta\phi - \left(-\phi(1-h)^{\frac{t-t_0}{h}} \right)}{(1-h)^{\frac{t+h-t_0}{h}} (1-h)^{\frac{t-t_0}{h}}}$$
$$= \frac{-(1-h)^{\frac{t-t_0}{h}}\phi + \phi(1-h)^{\frac{t-t_0}{h}}}{(1-h)^{\frac{t+h-t_0}{h}} (1-h)^{\frac{t-t_0}{h}}}$$
$$= 0.$$

Since the difference operator applied to $\dfrac{\phi}{(1-h)^{\frac{t-t_0}{h}}}$ equals zero, we conclude by Theorem 4.15 that $\dfrac{\phi}{(1-h)^{\frac{t-t_0}{h}}} = C$ for some real number C. $\qquad\square$

This theorem tells us that any function that is a solution to $\Delta y = -y$ is equal to $C(1-h)^{\frac{t-t_0}{h}}$ for a particular C. This is sufficient evidence to claim that $y(t) = C(1-h)^{\frac{t-t_0}{h}}$, where C is an arbitrary real number, is the general solution of $\Delta y = -y$.

5. TIME SCALES CALCULUS

A time scale is a domain on which to perform calculus. The literature for time scales calculus is primarily based on real numbers [2], but our approach will be based on hyperreal numbers. Many theorems that normally require complicated limit arguments for so-called "dense" points (points that behave as if they were part of a real line interval) are simplified by using the methods of our infinitesimal calculus. Changing scope to the hyperreals does not alter the fundamental theorems of the subject, but offers a more elegant representation.

5.1. TIME SCALES

Section 4 offers evidence that there is a great deal of similarity between the theory of differential equations and the theory of difference equations. The *time scales calculus* unifies and extends the study of these equations and many more in a general language. The heart of these two theories is in the definition of the derivative-like operator in use and to sufficiently generalize such an operator, we make a few observations. Derivative-like operators look like fractions: in the numerator we have the difference of two terms, and the denominator is a diffrence that is related to the stepsize.

It is convenient to perform a thought experiment to better understand stepsize. Suppose we throw a ball and record its height for every 0.01 seconds that the ball remains in the air, in which case the position of the ball at times in the set $\{0, 0.01, 0.02, 0.03, \ldots, 5\}$ are recorded (this ball hit the ground after 5 seconds). If

we now choose to compare the laws of physics to the scores of the ball, we may use a difference equation on the set $0.01\mathbb{Z}$. In this case, we say that $0.01\mathbb{Z}$ is the *time scale* for the experiment. Time scales do not have to necessarily correspond to time, but the notion of recording data at particular times leads one to easily speculate on how complicated such sets can be.

In this sense a time scale is the domain of the function on which you are performing calculus. It will be often useful to talk about the largest or smallest element of a time scale of finite length. This property is sensible when the set is closed, by Theorem 3.22. Common time scales are $\mathbb{R}$, $h\mathbb{Z}$, and any finite set S.

Definition 5.1. A *time scale* $\mathbb{T}$ is a closed subset of $\mathbb{R}$.

Although Theorem 3.22 requires that the set in question be bounded and closed, we do not require that the time scale be bounded. This is because when the time comes to use Theorem 3.22 in a proof, we will be focusing on bounded subsets of $\mathbb{T}$, which will be necessarily closed. Also, requiring $\mathbb{T}$ to be bounded would mean that the set $\mathbb{R}$ is not a time scale, and then we could no longer say that time scales calculus generalizes differential calculus.

5.2. DELTA DERIVATIVES

To complete our definition of the derivative, we still need to define some sort of stepsize operation. For differential equations, we let an infinitesimal be our stepsize for $\mathbb{R}$ and for difference equations we take this stepsize to be the distance from a domain element to the next largest domain element.

For example, we can define $\mathbb{T} = \{\frac{1}{n} \colon n \in \mathbb{N}\} \cup \{0\}$. Then the "next" element of $\mathbb{T}$ after $\frac{1}{3}$ is $\frac{1}{2}$ and the "next" element of $\mathbb{T}$ after $\frac{1}{2}$ is 1. All points of $\mathbb{T}$ have an obvious
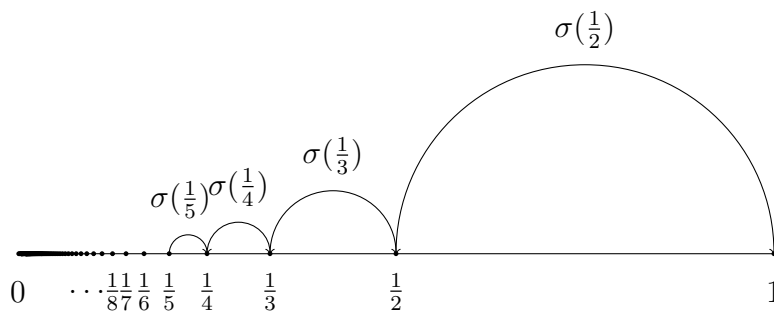
Figure 5.1. $\mathbb{T} = \left\{ \frac{1}{n} \colon n \in \mathbb{N} \right\} \cup \{0\}$.

"next" element of $\mathbb{T}$ except for 0. In this case, we recall that in Section 4, the stepsize in differential calculus was an infinitesimal. As it stands, we cannot immediately say that the element "next" to 0 is an infinitesimal, because the time scale $\mathbb{T}$ does not contain any infinitesimals. Therefore, we have to leave $\mathbb{T}$ and perform our calculus in the set $^*\mathbb{T}$ guaranteed to exist by Theorem 3.14. The element 0 is a limit point of $\mathbb{T}$ and therefore, Lemma 3.24 shows that $\text{halo}(0) - \{0\} \cap {}^*\mathbb{T}$ is a nonempty set. This guarantees that a nonzero infinitesimal in $^*\mathbb{T}$. There may be many possible choices for our "next" element after 0, but only the one guaranteed by Lemma 3.24 will be necessary for our calculations. For this reason, we will define the "next" element to be in some halo. This offers the possibility of multiple choices for a particular "next" element, but we will see that no matter which one is chosen, the theorems do not change.

Informally, if there is no obvious "next" element of $t \in \mathbb{T}$, then we pick an arbitrary element of $\text{halo}(t)$ not equal to t as the "next" element. In general, the halo of an element may remain a subset of $\mathbb{R}$ or it may also include infinitesimals. Since we are picking the "next" to be in the halo of t, this choice will sometimes lie

in the set $^*\mathbb{T} - \mathbb{T}$, the hyperreal numbers gained by the extension of the set $\mathbb{T}$ (this extension is guaranteed to exist by Theorem 3.14). We will see that the choice of halo element we pick will not affect our calculus.

Which number is it whose halo we are picking from? For $t \in \mathbb{T}$, the "next" element of $\mathbb{T}$ after t will be the infimum of all elements of $\mathbb{T}$ greater than t. This definition forces us to first form the set $\{x \in \mathbb{T} \colon x > t\}$ and then take its infimum, which since $\mathbb{T}$ is closed, will always be an element of $\mathbb{T}$. Therefore, we first find the number $\inf\{x \in \mathbb{T} \colon x > t\}$ and then pick the "next" element of $\mathbb{T}$ after t to be an element of $\mathrm{halo}(\inf\{x \in \mathbb{T} \colon x > t\}) \subset {}^*\mathbb{T}$.

Definition 5.2. Let $\mathbb{T}$ be a time scale and suppose $t \neq \max\mathbb{T}$. The *jump operator* $\sigma \colon \mathbb{T} \to {}^*\mathbb{T}$ is defined as $\sigma(t) \in \mathrm{halo}(\inf\{x \in \mathbb{T} \colon x > t\}) \cap {}^*\mathbb{T}$. If $t = \max\mathbb{T}$, then define $\sigma(t) = t$.

If $t \neq \max\mathbb{T}$, by definition, all possible choices for $\sigma(t)$ lie in the halo around the real number $\inf\{x \in \mathbb{T} \colon x > t\}$ which is guaranteed to exist by Axiom 3.9. The jump operator is usually not a function, but it is for finite time scales.

Example 5.3. Let $\mathbb{T} = \{t_1, t_2, \ldots, t_n\}$ for some $n \in \mathbb{N}$ and assume that $t_1 < t_2 < \ldots < t_n$ (if not, relabel your points). We can describe $\mathbb{T}$ with the sentence "for all $t \in \mathbb{T}$, $t = t_1$, or $t = t_2$, or $\ldots$, or $t = t_n$" which by Axiom 3.21 will transfer to the true sentence "for all $t \in {}^*\mathbb{T}$, $t = t_1$, or $t = t_2$, or $\ldots$, or $t = t_n$". Thus we have $^*\mathbb{T} = \mathbb{T}$, so for any $t \in \mathbb{T}$, since $\sigma(t) \in \mathrm{halo}(\inf\{x \in \mathbb{T} \colon x > t\}) \cap {}^*\mathbb{T}$, we see that there is only one choice for $\sigma(t)$ in $\mathbb{T}$. Thus, for an arbitrary finite time scale, the jump operator is a function.

However, this will not always work, because we know that $\mathbb{T} = \mathbb{R}$ is a time scale and we know that $^*\mathbb{R} \neq \mathbb{R}$. Because the element in the halo that is picked may not

be unique each time it is chosen. However, this is perfectly fine, because we will see that the choice reduces to an infinitesimal error in the results of the calculus, which we remove by taking standard parts.

While the jump operator tells us a "next" element of the time scale, it is sometimes useful to talk about the length from one point to the next. This is because the formula for that length is a formula for the stepsize of the timescale from the t to $\sigma(t)$.

Definition 5.4. The graininess operator $\mu \colon \mathbb{T} \to {}^*\mathbb{R}$ is defined as $\mu(t) = \sigma(t) - t$.

The definition of μ forces that $\mu(t)$ never be negative. It is possible that $\mu(t) = 0$ for a particular choice of $\sigma(t)$, but we can always avoid this problem.

Lemma 5.5. *Let $\mathbb{T}$ be a time scale and suppose that $t \in \mathbb{T}$ such that $t \neq \max \mathbb{T}$ if it exists. Then, it is always possible to choose $\sigma(t)$ so that $\mu(t)$ is nonzero.*

Proof. If $\mu(t) > 0$, we are done. So suppose that for a particular choice of $\sigma(t)$, that $\mu(t) = 0$. By definition of μ, we see that $\sigma(t) = t$, so since $t \in \mathbb{T}$ and $\mathbb{T} \subset \mathbb{R}$, we know that t is a real number and that $t \in \mathrm{halo}(\inf\{x \in \mathbb{T} \colon x > t\}) \cap \mathbb{T}^*$. But since there is only one real number in any halo, we see that $t \in \mathrm{halo}(t) \cap \mathbb{T}^*$. By Lemma 3.23 we know there is $a \colon \mathbb{N} \to \mathbb{T}$ such that $a_n \to x$. So by Lemma 3.24 we see that the set $\mathrm{halo}(t) \cap \mathbb{T}^* - \{t\}$ is nonempty. Therefore we may pick $\sigma(t) \in \mathrm{halo}(t) \cap \mathbb{T}^* - \{t\}$ which allows $\mu(t)$ to be nonzero. $\qquad\square$

We know that all choices for $\sigma(x)$ lie in a single halo, the same is also true for graininess.

Lemma 5.6. *Let* $\mathbb{T}$ *be a time scale. Then for all* $t \in \mathbb{T}$ *and for all choices of* $\sigma(t)$, $\mu(t) = \sigma(t) - t$ *lies in a single halo.*

Proof. In the formula $\mu(t) = \sigma(t) - t$, all legitimate choices of $\sigma(t)$ are infinitesimally close. Let $\sigma_0(t)$ be any particular fixed choice of $\sigma(t)$. Varying our choice of $\sigma(t)$ alters the quantity $\sigma(t) - t$ by an infinitesimal amount. So for all choices of $\sigma(t)$ we have $\sigma(t) - t \in \text{halo}(\sigma_0(t) - t)$. □

Lemma 5.5 and Lemma 5.6 reveal why we do not need to keep track of the particular hyperreal chosen by σ (assuming there is more than one choice). Our calculus will be based on our derivative operator, and that the stepsize of the operator will be defined by μ, and the previous two lemmas allow us to verify that every infinitesimal will lead to the same real number-valued result after we take standard parts.

We are now prepared to define the derivative. The definition is very similar to the differential and difference calculi definitions, but we replace the notion of stepsize with the graininess function. Since the graininess function takes real numbers to hyperreal numbers, we will be operating in $^*\mathbb{R}$, so we will take a standard part to make the derivative real-valued. The derivative cannot be defined at the point $\max \mathbb{T}$ (if it exists), because $\mu(\max \mathbb{T}) = 0$ (Lemma 5.5 does not apply, since $t = \max \mathbb{T}$). So, we define a new set, $\mathbb{T}^\kappa = \mathbb{T} - \{\max \mathbb{T}\}$ which is the time scale without its maximum point. In addition we require a delta-differentiable function to be continuous.

Definition 5.7. Let $\mathbb{T}$ be a time scale and $f \colon \mathbb{T} \to \mathbb{R}$ a function which is continuous. If at some point $t \in \mathbb{T}^\kappa$, we have $\text{st}\left(\frac{^*f(\sigma(t)) - ^*f(t)}{\mu(t)}\right)$ with the same value for all choices of $\sigma(t)$, then we say that $f^\Delta(t) = \text{st}\left(\frac{^*f(\sigma(t)) - ^*f(t)}{\mu(t)}\right)$ is the delta-derivative of f at t.

We want to show why it is important the f be continuous.

Example 5.8. Consider the time scale $\mathbb{T} = [0,1] \cup \{2,3\}$ and the function

$$f(x) = \begin{cases} 5; & x \in [0,1) \\ 4; & x \in 1,2,3. \end{cases}$$

This function is not continuous at the point $1 \in \mathbb{T}$. However the rest of the definition of the delta-derivative of this function is well defined and we would have for all $t \in \mathbb{T}$, $f^{\Delta}(t) = 0$. The problem here is not that the formula for the derivative derivative fails, but that the function itself is not continuous. We make the restriction that f be continuous to preserve Theorem 4.6 in the general setting of time scales.

In Example 4.3 and Example 4.16 we computed the derivative of $f(x) = x^2$ first over the time scale $\mathbb{R}$ and then over the time scale $h\mathbb{Z}$. We now compute the delta-derivative of $f(x) = x^2$ and then show we can deduce both of the results in Example 4.3 and Example 4.16.

Example 5.9. Let us compute the delta-derivative of $f(t) = t^2$ for a general time scale $\mathbb{T}$. So,

$$\begin{aligned} f^{\Delta}(t) &= \text{st}\left(\frac{(\sigma(t))^2 - t^2}{\mu(t)}\right) \\ &= \text{st}\left(\frac{(t + \mu(t))^2 - t^2}{\mu(t)}\right) \\ &= \text{st}\left(\frac{t^2 + 2t\mu(t) + \mu(t)^2 - t^2}{\mu(t)}\right) \\ &= \text{st}\left(\frac{2t\mu(t) + \mu(t)^2}{\mu(t)}\right) \\ &= \text{st}\left(2t + \mu(t)\right). \end{aligned}$$

Now, if $\mathbb{T} = \mathbb{R}$, for all $t \in \mathbb{T}$, $\sigma(t) = t + \epsilon$, so $f^{\Delta}(t) = \text{st}(2t + \epsilon) = 2t$ for all ϵ, as expected from Example 4.3. However, if $\mathbb{T} = h\mathbb{Z}$, for all $t \in \mathbb{T}$, $\sigma(t) = t + h$, so $f^{\Delta}(t) = \text{st}(2t + h) = 2t + h$, as expected from Example 4.16.

The derivative is useful for characterizing other properties of a function. For example, if the derivative is zero at all $t \in \mathbb{T}$, we think of the function having no change and therefore should keep the same value. The following theorem formalizes that concept, but first we must define a couple of terms. An element $t \in \mathbb{T}$ is called *left-dense* if and only if for all $\bar{t} \in \mathbb{T}$ such that $\bar{t} < t$, $\sigma(\bar{t}) \neq t$. A point which is not left-dense is called *left-scattered*.

Theorem 5.10. *Let f be a delta-differentiable function. For all $t \in \mathbb{T}^{\kappa}$, $f^{\Delta}(t) = 0$ if and only if f is a constant function, i.e., there exists $C \in \mathbb{R}$ so that for all $t \in \mathbb{T}$, $f(t) = C$.*

Proof. ($\Rightarrow$) Let $t_1, t_2 \in \mathbb{T}$ so that $t_1 \neq t_2$ and $t_1 < t_2$. We proceed by contradiction. Suppose that $f(t_1) = C$ and $f(t_2) \neq f(t_1)$. If $\sigma(t_1) = t_2$, then since

$$
\begin{aligned}
0 =& f^{\Delta}(t_1) \\
=& \text{st}\left(\frac{{}^*f(\sigma(t_1)) - {}^*f(t_1)}{\mu(t_1)} \right) \\
=& \frac{f(t_2) - f(t_1)}{t_2 - t_1},
\end{aligned}
$$

we see that $f(t_1) = f(t_2)$, a contradiction. Therefore, $\sigma(t_1) \neq t_2$.

Define $Q = \{t \in \mathbb{T} : t_1 < t < t_2 \text{ and } f(t) \neq f(t_2)\}$ which must be nonempty and then define $\hat{t} = \sup Q$. We know that $\hat{t} \in \mathbb{T}$ because $\mathbb{T}$ is closed. Now assume that $f(\hat{t}) = f(t_2)$. Suppose that $\hat{t}$ is left-scattered, then there would be some $\bar{t} \in \mathbb{T}$

so that $\sigma(\bar{t}) = \hat{t}$. But we know that

$$
\begin{aligned}
0 &= f^{\Delta}(\bar{t}) \\
&= \mathrm{st}\left(\frac{{}^*f(\sigma(\bar{t})) - {}^*f(\bar{t})}{\mu(\bar{t})}\right) \\
&= \frac{f(\hat{t}) - f(\bar{t})}{\hat{t} - \bar{t}},
\end{aligned}
$$

which implies $f(\bar{t}) = f(\hat{t}) = f(t_2)$, a contradiction because it would mean $\bar{t}$ would be a smaller upper bound of Q than $\hat{t}$ is. Thus if $f(\hat{t}) = f(t_2)$, then $\hat{t}$ is left dense.

But since $\hat{t}$ is left dense, Lemma 3.24 guarantees that there is an element $w \in {}^*Q$ so that $w \neq \hat{t}$ and $w \sim \hat{t}$. Since

"for all $t \in Q$, $f(t) \neq f(t_2)$"

is a true sentence, it transfers to the true sentence

"for all $t \in {}^*Q$, ${}^*f(t) \neq {}^*f(t_2)$",

we see that ${}^*f(w) \neq f(t_2)$. But this is impossible since f is continuous, i.e. $w \sim \hat{t}$ implies ${}^*f(w) \sim {}^*f(\hat{t})$, which contradicts the assumption that $f(\hat{t}) = f(t_2)$. Therefore we have proved that whether $\hat{t}$ is left scattered or left dense, $f(\hat{t}) \neq f(t_2)$.

Now assume that $\sigma(\hat{t}) = t_2$, then since

$$
\begin{aligned}
0 &= f^{\Delta}(\hat{t}) \\
&= \mathrm{st}\left(\frac{{}^*f(\sigma(\hat{t})) - {}^*f(\hat{t})}{\mu(\hat{t})}\right) \\
&= \frac{f(t_2) - f(\hat{t})}{t_2 - \hat{t}},
\end{aligned}
$$

we have $f(\hat{t}) = f(t_2)$, a contradiction. Therefore $\sigma(\hat{t}) \neq t_2$. Since $\hat{t} = \sup Q$, for $\sigma(\hat{t}) > \hat{t}$ we have $f(\sigma(\hat{t})) = f(t_2)$. If $\sigma(\hat{t}) \sim \hat{t}$ we have a contradiction since f is continuous. If $\sigma(\hat{t}) \not\sim \hat{t}$ we have a contradiction since $f^\Delta(\hat{t}) = 0$ implies $f(\hat{t}) = f(\sigma(\hat{t}))$. Thus our original assumption that $f(t_1) \neq f(t_2)$ always leads to a contradiction. Therefore if for all $t \in \mathbb{T}$, $f^\Delta(t) = 0$ and $f(t_1) = C$, then we cannot allow there to be a t_2 so that $f(t_2) \neq f(t_1)$. This means for all $t \in \mathbb{T}$, $f(t) = C$.

($\Leftarrow$) If there is some $C \in \mathbb{R}$ so that for all $t \in \mathbb{T}$, $f(t) = C$, then for all $t \in {}^*\mathbb{T}$, ${}^*f(t) = C$. By definition

$$f^\Delta(t) = \text{st}\left(\frac{{}^*f(\sigma(t)) - {}^*f(t)}{\mu(t)}\right) = \text{st}\left(\frac{C - C}{\mu(t)}\right) = 0.$$

Therefore if for all $t \in \mathbb{T}$, $f(t) = C$, then for all $t \in \mathbb{T}^\kappa$, $f^\Delta(t) = 0$. $\qquad\square$

Derivatives have nice properties when it comes to sums, products, and quotients of differentiable functions. These properties allow one to differentiate complicated functions in terms of simpler ones. We summarize here the most common such properties.

Theorem 5.11. *Let $\mathbb{T}$ be a time scale, $f, g : \mathbb{T} \to \mathbb{R}$ be delta-differentiable functions, and $\alpha \in \mathbb{R}$ a constant. Then*

(i) *(Sum Rule)*

$$(f + g)^\Delta(t) = f^\Delta(t) + g^\Delta(t),$$

(ii) *(Constant Multiple Rule)*

$$(\alpha f)^\Delta(t) = \alpha f^\Delta(t),$$

(iii) (*Product Rule*)

$$(fg)^\Delta(t) = f^\Delta(t)g(t) + \text{st}(^*f(\sigma(t)))g^\Delta(t)$$

$$= f(t)g^\Delta(t) + f^\Delta(t)\text{st}(^*g(\sigma(t))),$$

(iv) (*Reciprocal Rule*) *if* $f(t)\text{st}(^*f(\sigma(t))) \neq 0$ *then*

$$\left(\frac{1}{f}\right)^\Delta(t) = -\frac{f^\Delta(t)}{f(t)\text{st}(^*f(\sigma(t)))},$$

(v) (*Quotient Rule*) *if* $g(t)\text{st}(^*g(\sigma(t))) \neq 0$ *then*

$$\left(\frac{f}{g}\right)^\Delta(t) = \frac{f^\Delta(t)g(t) - f(t)g^\Delta(t)}{g(t)\text{st}(^*g(\sigma(t)))}.$$

Proof. The proofs of these theorems are straightforward calculations and follow directly from the definition. To prove (i) compute

$$(f+g)^\Delta(t) = \text{st}\left(\frac{^*(f+g)(\sigma(t)) - {}^*(f+g)(t)}{\mu(t)}\right)$$

$$= \text{st}\left(\frac{^*f(\sigma(t)) + {}^*g(\sigma(t)) - {}^*f(t) - {}^*g(t)}{\mu(t)}\right)$$

$$= \text{st}\left(\frac{^*f(\sigma(t)) - {}^*f(t) + {}^*g(\sigma(t)) - {}^*g(t)}{\mu(t)}\right)$$

$$= \text{st}\left(\frac{^*f(\sigma(t)) - {}^*f(t)}{\mu(t)}\right) + \text{st}\left(\frac{^*g(\sigma(t)) - {}^*g(t)}{\mu(t)}\right)$$

$$= f^\Delta(t) + g^\Delta(t).$$

To prove (ii) compute

$$(\alpha f)^\Delta(t) = \mathrm{st}\left(\frac{{}^*(\alpha f)(\sigma(t)) - {}^*(\alpha f)(t)}{\mu(t)}\right)$$

$$= \mathrm{st}\left(\frac{\alpha\left({}^*f(\sigma(t)) - {}^*f(t)\right)}{\mu(t)}\right)$$

$$= \alpha\mathrm{st}\left(\frac{{}^*f(\sigma(t)) - {}^*f(t)}{\mu(t)}\right)$$

$$= \alpha f^\Delta(t).$$

For (iii), we must prove two different equalities and we will do so by a different versions of the same trick. For the first equality, we add and subtract ${}^*f(\sigma(t)){}^*g(t)$ in the numerator and for the second we add and subtract ${}^*f(t){}^*g(\sigma(t))$ in the numerator. This trick appears in the second line of each proof. The first equality follows from the computation

$$(fg)^\Delta(t) = \mathrm{st}\left(\frac{{}^*(fg)(\sigma(t)) - {}^*(fg)(t)}{\mu(t)}\right)$$

$$= \mathrm{st}\left(\frac{[{}^*f(\sigma(t)){}^*g(\sigma(t)) - {}^*f(\sigma(t)){}^*g(t)] - [{}^*f(t){}^*g(t) - {}^*f(\sigma(t)){}^*g(t)]}{\mu(t)}\right)$$

$$= \mathrm{st}\left({}^*f(\sigma(t))\frac{{}^*g(\sigma(t)) - {}^*g(t)}{\mu(t)}\right) + \mathrm{st}\left({}^*g(t)\frac{{}^*f(\sigma(t)) - {}^*f(t))}{\mu(t)}\right)$$

$$= \mathrm{st}\left({}^*f(\sigma(t))\right)\mathrm{st}\left(\frac{{}^*g(\sigma(t)) - {}^*g(t)}{\mu(t)}\right) + \mathrm{st}({}^*g(t))\mathrm{st}\left(\frac{{}^*f(\sigma(t) - {}^*f(t))}{\mu(t)}\right)$$

$$= \mathrm{st}({}^*f(\sigma(t)))g^\Delta(t) + g(t)f^\Delta(t).$$

The second equality follows from the computation

$$(fg)^{\Delta}(t) = \text{st}\left(\frac{{}^*(fg)(\sigma(t)) - {}^*(fg)(t)}{\mu(t)}\right)$$

$$= \text{st}\left(\frac{[{}^*f(\sigma(t)){}^*g(\sigma(t)) - {}^*f(t){}^*g(\sigma(t))] - [{}^*f(t){}^*g(t) - {}^*f(t){}^*g(\sigma(t))]}{\mu(t)}\right)$$

$$= \text{st}\left({}^*g(\sigma(t))\frac{{}^*f(\sigma(t)) - {}^*f(t)}{\mu(t)}\right) + \text{st}\left({}^*f(t)\frac{{}^*g(\sigma(t)) - {}^*g(t))}{\mu(t)}\right)$$

$$= \text{st}({}^*g(\sigma(t)))\text{st}\left(\frac{{}^*f(\sigma(t)) - {}^*f(t)}{\mu(t)}\right) + \text{st}({}^*f(t))\text{st}\left(\frac{{}^*g(\sigma(t) - {}^*g(t))}{\mu(t)}\right)$$

$$= \text{st}({}^*g(\sigma(t)))f^{\Delta}(t) + f(t)g^{\Delta}(t).$$

To prove (iv), compute

$$\left(\frac{1}{f}\right)^{\Delta}(t) = \text{st}\left(\frac{{}^*(\frac{1}{f})(\sigma(t)) - {}^*(\frac{1}{f})(t)}{\mu(t)}\right)$$

$$= \text{st}\left(\frac{1}{\mu(t){}^*f(\sigma(t))} - \frac{1}{\mu(t){}^*f(t)}\right)$$

$$= \text{st}\left(\frac{{}^*f(t) - {}^*f(\sigma(t))}{\mu(t){}^*f(t){}^*f(\sigma(t))}\right)$$

$$= \frac{\text{st}\left(\frac{{}^*f(t) - {}^*f(\sigma(t))}{\mu(t))}\right)}{\text{st}\left({}^*f(t){}^*f(\sigma(t))\right)}$$

$$= \frac{-f^{\Delta}(t)}{f(t)\text{st}\left({}^*f(\sigma(t))\right)}.$$

To prove (v), we use (iii) and (iv) by splitting $\frac{f}{g}$ into the product $f \cdot \frac{1}{g}$ and computing

$$\left(\frac{f}{g}\right)^{\Delta}(t) = \left(f \cdot \frac{1}{g}\right)^{\Delta}(t)$$

$$= f(t)\left(\frac{1}{g}\right)^{\Delta}(t) + f^{\Delta}(t)\text{st}\left({}^{*}\left(\frac{1}{g}\right)(\sigma(t))\right)$$

$$= \frac{-f(t)g^{\Delta}(t)}{g(t)\text{st}({}^{*}g(\sigma(t)))} + \frac{f^{\Delta}(t)}{\text{st}({}^{*}g(\sigma(t)))}$$

$$= \frac{f^{\Delta}(t)g(t) - f(t)g^{\Delta}(t)}{g(t)\text{st}({}^{*}g(\sigma(t)))}.$$

So sums, products, and quotients of delta-differentiable functions are relatively easy to compute. □

5.3. DELTA INTEGRATION

Differentiation is called such because it is a type of difference, and since it is a difference, subtraction is involved and some information about the function is lost. So, to "un-differentiate", we should compute a sum which we will call an *integral*. The difference is analogous to the sum as differentiation is to integration, so we seek a way to sum a differentiated function to "undo" differentiation just as subtraction "undoes" addition. Our construction of the integral is inspired by [4]. What can

t	f^{Δ}	$f(t) = 1 + f(t-1)$	$f(t)$ when $f(1) = 2$
1	1	?	$f(1) = 2$
2	1	1 + f(1)	$f(2) = 3$
3	1	1 + f(2)	$f(3) = 4$
4	1	1 + f(3)	$f(4) = 5$
5	-	1 + f(4)	$f(5) = 6$

Table 5.1. The generating sequence for f given $f^{\Delta} = 1$ with $\mathbb{T} = \{1, 2, 3, 4, 5\}$.

we recover about a function knowing only its derivative? The easy case here is when the time scale is finite, such as $\mathbb{T} = \{1, 2, 3, 4, 5\}$. Consider the function $f\colon \mathbb{T} \to \mathbb{R}$, $f(t) = t + 1$. Section 5.2 shows us that for all $t \in \mathbb{T}^\kappa$,

$$f^\Delta(t) = \mathrm{st}\left(\frac{{}^*f(\sigma(t)) - {}^*f(t)}{\mu(t)}\right)$$

$$= f(t+1) - f(t)$$

$$= 1.$$

Rearranging this equation yields the sequence $f(t+1) = 1 + f(t)$ called the *generating sequence of f*, but no other information. Table 5.1 illustrates the computation of the generating sequence of f and we note that a lot of information about f can be recovered from its derivative, but it all depends on the value of $f(1)$ which is unfortunately not determinable without information outside of the generating sequence. Column 3 shows that if we choose $f(1) = 2$, it turns out that the resulting function is set equal to $f(t) = t + 1$. In general, if we force $f(1) = f_1 > 0$ for some $f_1 \in \mathbb{R}$, then we will get the equation $f(t) = t + f_1$ as the result. In general, picking $f(t) \in \mathbb{R}$ for some $t \in \mathrm{dom}(f)$ is called specifying an *initial condition* for the generating sequence.

Generating sequences are really instruction manuals which explain how to recover a differentiated function from its derivative. However, since differentiation causes a loss of information, the instructions are incomplete. Any given generating sequence produces infinitely many possible functions, and only one will be the specific solution being sought. The missing information is recovered by specifying the initial condition for the generating sequence.

Let $\mathbb{T} = \{t_0, t_1, \ldots, t_n\}$ be a finite time scale labeled so that $t_0 < t_1 < \ldots < t_n$. Rearrange the equation $f^\Delta = \frac{f(\sigma(t)) - f(t)}{\mu(t)}$ to yield $f(\sigma(t)) = \mu(t)f^\Delta(t) + f(t)$, the

most general generating sequence, and specify an initial condition $f(t_0) = f_0$ for some $t_0 \in \mathbb{T}$ so we can proceed with the computation. We will then define the *integral of f^Δ over the finite time scale $\mathbb{T}$* as

$$f(t) = \sum_{i=1}^{j} f^\Delta(t_i)\mu(t_i) \text{ for } 0 \leq j \leq n.$$

While any empirical data will fit into a time scale of this type, we are not able to recover the integral over $\mathbb{T} = [0, 1] \subset \mathbb{R}$ or similar time scales. We will restrict our attention to *bounded* time scales, i.e., time scales which have a least element a and by Theorem 3.22 a greatest element b. So let $\mathbb{T}$ be a bounded time scale and $\Delta t \in \mathbb{R}$; we wish to break $\mathbb{T}$ into intervals spaced by Δt. Define $N \colon \mathbb{R} \to \mathbb{N}$, where $N(\Delta t)$ is the smallest natural number such that

$$b \in [a + N(\Delta t)\Delta t, a + (N(\Delta t) + 1)\Delta t].$$

Set $I_i = [a + i\Delta t, a + (i + 1)\Delta t] \cap \mathbb{T}$. Now, define the set

$$P(\Delta x) = \{\inf(I_i) \colon 0 \leq i \leq N(\Delta x)\}.$$

Since for all real Δt, $P(\Delta t)$ is a finite time scale, we label its elements as $P(\Delta t) = \{t_0, \ldots, t_{M(\Delta t)}\}$ for appropriate $M(\Delta t)$ such that $t_0 < \ldots < t_{M(\Delta t)}$ and we let $\mu_{P(\Delta t)}$ be its respective graininess function. It follows from Example 5.3 that for all $\Delta t \in \mathbb{R}$, $\mu_{P(\Delta t)}$ is a function. Now, we define the integral of f over the finite discrete time scale $P(\Delta t)$.

Definition 5.12. Let $f\colon \mathbb{T} \to \mathbb{R}$ be a function and $\Delta t \in \mathbb{R}$. We define the function ${}_f\mathcal{S}_a^b\colon \mathbb{R} \to \mathbb{R}$ which is the integral of f over the finite time scale $P(\Delta t)$:

$$
{}_f\mathcal{S}_a^b(\Delta t) = \sum_{i=0}^{M(\Delta t)} f(t_i)\mu_{P(\Delta t)}(t_i).
$$

This integral is no different than the one we had defined previously, but now we have definitions in place that will allow us to extend this function to ${}^*\mathbb{T}$. The definition of the function $N(\Delta t)$ transfers to the function ${}^*N(\Delta t)\colon {}^*\mathbb{R} \to {}^*\mathbb{N}$ (which may be an infinite number) and the finite time scale $P(\Delta t)$ extends to the set

$$
{}^*P(\Delta t) = \{t_0, \ldots, t_{{}^*M(\Delta t)}\}
$$

for some ${}^*M(\Delta t) \in {}^*\mathbb{N}$. Now, ${}_f\mathcal{S}_a^b$ is a function which transfers to ${}_f^*\mathcal{S}_a^b\colon {}^*\mathbb{R} \to {}^*\mathbb{R}$ where

$$
{}_f^*\mathcal{S}_a^b = \sum_{i=0}^{{}^*M(\Delta t)} {}^*f(t_i)\mu_{{}^*P(\Delta t)}(t_i),
$$

which is a well defined summation by the methods on [4, p.153]. So we may let Δt be an infinitesimal and consider ${}_f^*\mathcal{S}_a^b(\Delta t)$. The function ${}_f^*\mathcal{S}_a^b$ is difficult to analyze directly, so we consider approximations to it called *upper sums* and *lower sums.* The upper sum is an overestimate of the final integral and the lower sum is an underestimate.

Definition 5.13. Let $f\colon \mathbb{T} \to \mathbb{R}$ be a function, and $\Delta t \in \mathbb{R}$. Let $M_i = \max\limits_{t \in [t_i, t_{i+1}] \cap \mathbb{T}} f(t)$ and $m_i = \min\limits_{t \in [t_i, t_{i+1}] \cap \mathbb{T}} f(t)$, which are guaranteed to exist in the set $f(\mathbb{T})$. Then, the

upper sum is a function $_f\mathcal{U}_a^b \colon \mathbb{R} \to \mathbb{R}$ such that

$$_f\mathcal{U}_a^b(\Delta t) = \sum_{i=0}^{N(\Delta t)} M_i \mu_{P(\Delta t)}(t_i)$$

and the *lower sum* is a function $_f\mathcal{L}_a^b \colon \mathbb{R} \to \mathbb{R}$ such that

$$_f\mathcal{L}_a^b(\Delta t) = \sum_{i=0}^{N(\Delta t)} m_i \mu_{P(\Delta t)}(t_i).$$

Immediately from the definition, it is clear that

"for all $\Delta t \in \mathbb{R}$, $_f\mathcal{L}_a^b(\Delta t) \leq {}_f\mathcal{S}_a^b(\Delta t) \leq {}_f\mathcal{U}_a^b(\Delta t)$"

is a true sentence which by Axiom 3.21 transfers to the true sentence

"for all $\Delta t \in {}^*\mathbb{R}$, $_f^*\mathcal{L}_a^b(\Delta t) \leq {}_f^*\mathcal{S}_a^b(\Delta t) \leq {}_f^*\mathcal{U}_a^b(\Delta t)$".

So we will say a function is *integrable* when the standard parts of all these values are equal. It turns out that the standard parts are all equal if f is a continuous function.

Theorem 5.14. *Let $\mathbb{T}$ be a time scale. If $f \colon [a,b] \cap \mathbb{T} \to \mathbb{R}$ is a continuous function, then for all nonzero positive infinitesimals Δt, $_f^*\mathcal{L}_a^b(\Delta t) \sim {}_f^*\mathcal{U}_a^b(\Delta t)$.*

Proof. Let $\mathbb{T}$, let $\Delta t \in \mathbb{R}$, let M_i and m_i be defined as in Definition 5.13 and let

$$\nu(\Delta t) = \max\{M_1 - m_1, M_2 - m_2, \ldots, M_{N(\Delta t)} - m_{N(\Delta t)}\}.$$

So for some j, we have

$$\nu(\Delta t) = M_j - m_j = f(c(\Delta t)) - f(d(\Delta t))$$

for some $c(\Delta t), d(\Delta t) \in \mathbb{T}$. By definition, $c(\Delta t)$ and $d(\Delta t)$ are in the interval $[t_j, t_{j+1}]$ (note: it is possible that $t_j = t_{j+1}$), so

$$|c(\Delta t) - d(\Delta t)| \leq t_{j+1} - t_j.$$

Now consider the difference

$$
\begin{aligned}
{}_f\mathcal{U}_a^b(\Delta t) - {}_f\mathcal{L}_a^b(\Delta t) &= \sum_{i=0}^{N(\Delta t)} M_i \mu_{P(\Delta t)}(t_i) - \sum_{i=0}^{N(\Delta t)} m_i \mu_{P(\Delta t)}(t_i) \\
&= \sum_{i=0}^{N(\Delta t)} (M_i - m_i) \mu_{P(\Delta t)}(t_i) \\
&\leq \sum_{i=0}^{N(\Delta t)} \nu(\Delta t) \mu_{P(\Delta t)}(t_i) \\
&= \nu(\Delta t) \sum_{i=0}^{N(\Delta t)} (t_{i+1} - t_i) \\
&= (f(c(\Delta t)) - f(d(\Delta t)))(b - a).
\end{aligned}
$$

The above inequality is a property that holds for all $\Delta t \in \mathbb{R}$. Apply the transfer principle to this inequality and let Δt be a nonzero positive infinitesimal. Then, we have the formula

$$ {}_f^*\mathcal{U}_a^b(\Delta t) - {}_f^*\mathcal{L}_a^b(\Delta t) \leq (f(c(\Delta t)) - f(d(\Delta t)))(b - a) $$

for some $c(\Delta t), d(\Delta t) \in [t_j, t_{j+1}]$. Since Δt is a nonzero positive infinitesimal, we have that $t_{j+1} - t_j$ are nonnegative infinitesimal and therefore $c(\Delta t) \sim d(\Delta t)$. Since f is continuous, $c(\Delta t) \sim d(\Delta t)$ implies $f(c(\Delta t)) \sim f(d(\Delta t))$ and therefore, $f(c(\Delta t)) - f(d(\Delta t))$ is an infinitesimal. Since $b - a$ is a real-valued quantity, the

product $(f(c(\Delta t)) - f(d(\Delta t)))(b-a)$ is infinitesimal. Therefore $\,{}^*_f\mathcal{L}^b_a(\Delta t) - {}^*_f\mathcal{U}^b_a(\Delta t)$ is an infinitesimal implying $\,{}^*_f\mathcal{L}^b_a(\Delta t) \sim {}^*_f\mathcal{U}^b_a(\Delta t)$. $\hfill\square$

We now mention that in the literature [2, pp.22-27], it is assumed that the function f being integrated is *right-dense continuous* which means f is continuous at $t \in \mathbb{T}$ such that $t \sim \sigma(t)$ and limits exist at so-called left-dense points . Theorem 1.60 in [2, p.22] establishes that if f is a continuous function, then f is right-dense continuous, so we simply write our proofs for continuous f. So for a continuous f, we have for all infinitesimals $\Delta t \in {}^*\mathbb{R}$,

$$ {}^*_f\mathcal{L}^b_a(\Delta t) \le {}^*_f\mathcal{S}^b_a(\Delta t) \le {}^*_f\mathcal{U}^b_a(\Delta t) $$

and $\,{}^*_f\mathcal{L}^b_a(\Delta t) \sim {}^*_f\mathcal{U}^b_a(\Delta t)$. Therefore, $\,{}^*_f\mathcal{L}^b_a(\Delta t) \sim {}^*_f\mathcal{S}^b_a(\Delta t) \sim {}^*_f\mathcal{U}^b_a(\Delta t)$, so if f is continuous, then f is integrable. We now define the integral of f over $\mathbb{T}$.

Definition 5.15. Let $\mathbb{T}$ be a bounded time scale such that $a = \inf \mathbb{T}$ and $b = \sup \mathbb{T}$. Let $f \colon \mathbb{T} \to \mathbb{R}$ be a continuous function and Δt be an infinitesimal. Then define the integral of f from a to b to be $\int_a^b f(t)\Delta t = \mathrm{st}({}^*_f\mathcal{S}^b_a(\Delta t))$.

The integral has properties similar to the properties of the derivative shown in Theorem 5.11.

Theorem 5.16. *Let $\mathbb{T}$ be a bounded time scale and $f, g : \mathbb{T} \to \mathbb{R}$ integrable functions, and $\alpha \in \mathbb{R}$ a constant. Then,*

(*i*) (*Constant Multiple Rule*)

$$ \int_a^b \alpha f(t)\Delta t = \alpha \int_a^b f(t)\Delta t; $$

(*ii*) (*Sum Rule*)

$$\int_a^b f(t) + g(t)\Delta t = \int_a^b f(t)\Delta t + \int_a^b g(t)\Delta t;$$

(*iii*) (*Combination Rule*)

$$\int_a^b f(t)\Delta t = \int_a^c f(t)\Delta t + \int_c^b f(t)\Delta t.$$

Proof. These proofs follow from verifying this to be true in the case for real numbers and then using the transfer principle to apply it to the actual integrals. For (i), let $\Delta t \in \mathbb{R}$ and consider the sum

$$\begin{aligned}
{}_{\alpha f}\mathcal{S}_a^b(\Delta t) &= \sum_{i=0}^{M(\Delta t)} \alpha f(t_i)\mu_{P(\Delta t)}(t_i) \\
&= \alpha \sum_{i=0}^{M(\Delta t)} f(t_i)\mu_{P(\Delta t)}(t_i) \\
&= \alpha_f \mathcal{S}_a^b(\Delta t).
\end{aligned}$$

Then an application of the transfer principle yields

$$_{\alpha f}{}^{*}\mathcal{S}_a^b(\Delta t) = \alpha_f^{*}\mathcal{S}_a^b(\Delta t)$$

for all $\Delta t \in {}^{*}\mathbb{R}$. Let Δt be a nonzero positive infinitesimal and take the standard part to get

$$\int_a^b \alpha f(t)\Delta t = \alpha \int_a^b f(t)\Delta t.$$

For (ii), again let $\Delta t \in \mathbb{R}$ and consider

$$_{f+g}\mathcal{S}_a^b(\Delta t) = \sum_{i=0}^{M(\Delta t)} (f(t_i) + g(t_i))\mu_{P(\Delta t)}(t_i)$$

$$= \sum_{i=0}^{M(\Delta t)} f(t_i)\mu_{P(\Delta t)}(t_i) + \sum_{i=0}^{M(\Delta t)} g(t_i)\mu_{P(\Delta t)}(t_i)$$

$$= {}_f\mathcal{S}_a^b(\Delta t) + {}_g\mathcal{S}_a^b(\Delta t).$$

Then an application of the transfer principle yields

$$_{f+g}{}^*\mathcal{S}_a^b(\Delta t) = {}_f^*\mathcal{S}_a^b(\Delta t) + {}_g^*\mathcal{S}_a^b(\Delta t)$$

for all $\Delta t \in {}^*\mathbb{R}$. Letting Δt be a nonzero positive infinitesimal, take the standard part, and use Lemma 3.16 to get

$$\int_a^b (f(t) + g(t))\, \Delta t = \int_a^b f(t)\Delta t + \int_a^b g(t)\Delta t.$$

For (iii), again let $\Delta t \in \mathbb{R}$ and let $c \in \mathbb{T}$. Then there is some natural number $N(c) \in \mathbb{N}$ such that $c \in [a + N(c)\Delta t, a + (N(c)+1)\Delta t]$. So consider the computation

$$_f\mathcal{S}_a^b(\Delta t) = \sum_{i=0}^{M(\Delta t)} f(t_i)\mu_{P(\Delta t)}(t_i)$$

$$= \sum_{i=0}^{N(c)} f(t_i)\mu_{P(\Delta t)}(t_i) + \sum_{i=N(c)}^{M(\Delta t)} f(t_i)\mu_{P(\Delta t)}(t_i)$$

$$= {}_f\mathcal{S}_a^c(\Delta t) + {}_f\mathcal{S}_c^b(\Delta t).$$

Then an application of the transfer principle yields

$$\,^*_f\mathcal{S}^b_a(\Delta t) = \,^*_f\mathcal{S}^c_a(\Delta t) + \,^*_f\mathcal{S}^b_c(\Delta t)$$

for all $\Delta t \in {}^*\mathbb{R}$. Let Δt be a nonzero positive infinitesimal and take the standard part to get

$$\int_a^b f(t)\Delta t = \int_a^c f(t)\Delta t + \int_c^b f(t)\Delta t.$$

$\square$

When we constructed the finite integral from the formula for derivatives and got the integral over finite time scales, we knew from our construction that it "undifferentiates". Although we modelled our new integral from the finite one, we must prove that it too is an inverse to differentiation. We demonstrate in the following theorem.

Theorem 5.17. (*Fundamental Theorem of Calculus*) *Let* $\mathbb{T}$ *be a time scale,* $t, t_0 \in \mathbb{T}$, $f \colon \mathbb{T} \to \mathbb{R}$ *be a differentiable function. Then,* $\int_{t_0}^t f^\Delta(\tau)\Delta\tau = f(t) - f(t_0)$.

Proof. Let $\Delta t \in \mathbb{R}$. Then the time scale $P(\Delta t) = \{t_0, \ldots, t_{M(\Delta t)}\}$ finite, so we have that $\mu(t_i) = t_{i+1} - t_i$ is real-valued for all $t \in \mathbb{T}$ and $\mathrm{st}({}^*f(t)) = f(t)$. We now compute

$$_{f^\Delta}\mathcal{S}_a^b(\Delta t) = \sum_{i=0}^{M(\Delta t)} f^\Delta(t_i)\mu_{P(\Delta t)}(t_i)$$

$$= \sum_{i=0}^{M(\Delta t)} \mathrm{st}\left(\frac{{}^*f(\sigma(t_i)) - {}^*f(t_i)}{\mu_{P(\Delta t)}(t_i)}\right)\mu_{P(\Delta t)}(t_i)$$

$$= \sum_{i=0}^{M(\Delta t)} \mathrm{st}\left(\frac{{}^*f(t_{i+1}) - {}^*f(t_i)}{t_{i+1} - t_i}\right)(t_{i+1} - t_i)$$

$$= \sum_{i=0}^{M(\Delta t)} \frac{f(t_{i+1}) - f(t_i)}{t_{i+1} - t_i}(t_{i+1} - t_i)$$

$$= \sum_{i=0}^{M(\Delta t)} f(t_{i+1}) - f(t_i)$$

$$= f(t_{M(\Delta t)}) - f(t_0).$$

This establishes the truth of the sentence

$$\text{``for all } \Delta t \in \mathbb{R},\ _{f^\Delta}\mathcal{S}_a^b(\Delta t) = f(t_{M(\Delta t)}) - f(t_0)\text{''}$$

so the Transfer Principle yields

$$\text{``for all } \Delta t \in {}^*\mathbb{R},\ _{f^\Delta}^*\mathcal{S}_a^b(\Delta t) = {}^*f(t_{*M(\Delta t)}) - {}^*f(t_0)\text{''}.$$

Now let Δt be an infinitesimal. By definition, we have $t_{*M(\Delta t)} \sim t$ and then since f is continuous, $f(t_{*M(\Delta t)}) \sim {}^*f(t)$. These facts establish

$$_{f^\Delta}^*\mathcal{S}_a^b(\Delta t) \sim {}^*f(t) - {}^*f(t_0).$$

Now take the standard part of each side to get

$$\mathrm{st}\left(_{f^\Delta}^*\mathcal{S}_a^b(\Delta t)\right) = \mathrm{st}({}^*f(t) - {}^*f(t_0)).$$

Finally, Lemma 3.16, Definition 5.15, and the fact that $\mathrm{st}({}^*f(t)) = f(t)$ and $\mathrm{st}({}^*f(t_0))$ yields the desired result

$$\int_{t_0}^t f^\Delta(\tau)\Delta\tau = f(t) - f(t_0). \quad \square$$

The fundamental theorem allows us to simplify expressions containing derivatives with integrals. We will use this property of integration in the next section to solve a dynamic equation.

5.4. DYNAMIC EQUATIONS

Dynamic equations are a generalization of differential equations and difference equations. Using the delta-derivative, we can regard both differential and difference equations as simply two species of our general theory.

Definition 5.18. A *dynamic equation* is an equation involving a function y, its independent variable, and any number of its delta derivatives $y^{\Delta^1}, y^{\Delta^2}, \ldots, y^{\Delta^n}$.

We would like to solve the dynamic equation $y^\Delta(t) = p(t)y(t)$ and we will do so by manipulating the formula algebraically to obtain a formula for $y(t)$. Let $t \in \mathbb{T}^\kappa$, $\Delta t \in \mathbb{R}$. Therefore on the time scale $P(\Delta t)$, $\mu(t)$ is real-valued and ${}^*y(t) = y(t)$. Consider the following computation:

$$\begin{aligned}
y^\Delta(t) &= \mathrm{st}\left(\frac{{}^*y(\sigma(t)) - {}^*y(t)}{\mu_{P(\Delta t)}(t)}\right) \\
&= \mathrm{st}\left(\frac{y(\sigma(t)) - y(t)}{\mu_{P(\Delta t)}(t)}\right) \\
&= \frac{y(\sigma(t)) - y(t)}{\mu_{P(\Delta t)}(t)}.
\end{aligned}$$

So now to solve the dynamic equation $y^\Delta(t) = y(t)$ we simply use the above form of y^Δ and get the equation

$$\frac{y(\sigma(t)) - y(t)}{\mu_{P(\Delta t)}(t)} = y(t).$$

Algebraic manipulation easily yields the equation

$$\frac{y(\sigma(t))}{y(t)} = \mu_{P(\Delta t)}(t) + 1.$$

On the left side of the equation, we have a quotient. So we apply the Log function to both sides, which admits

$$\mathrm{Log}(y(\sigma(t))) - \mathrm{Log}(y(t)) = \mathrm{Log}(p(t)\mu_{P(\Delta t)}(t) + 1).$$

The left side is the numerator of a delta-derivative. However we now see that the function p requires a restriction. We cannot let $p(t)\mu(t) + 1 = 0$, so we must require $p(t)$ to be *regressive* which means it is continuous and for all $t \in \mathbb{T}$, $1 + \mu(t)p(t) \neq 0$. We now divide by $\mu_{P(\Delta t)}(t)$ to get

$$\frac{\mathrm{Log}(y(\sigma(t))) - \mathrm{Log}(y(t))}{\mu_{P(\Delta t)}(t)} = \frac{\mathrm{Log}(p(t)\mu_{P(\Delta t)}(t) + 1)}{\mu_{P(\Delta t)}(t)}.$$

Since $\Delta t \in \mathbb{R}$, the left side of the equation equals its standard part and therefore it equals the derivative $(\mathrm{Log}(y))^\Delta(t)$. This gives us

$$(\mathrm{Log}(y))^\Delta(t) = \frac{\mathrm{Log}(p(t)y(t) + 1)}{\mu_{P(\Delta t)}(t)}.$$

We simplify our notation by defining $Q(t) := \frac{\text{Log}(p(t)\mu(t)+1)}{\mu_{P(\Delta t)}(t)}$. Now plug both sides into $_z\mathcal{S}_{t_0}^t(\Delta t)$, replacing z with the left and right sides of the equation respectively to get

$$_{(\text{Log}(y))^\Delta(t))}\mathcal{S}_{t_0}^t(\Delta t) =_Q \mathcal{S}_{t_0}^t(\Delta t).$$

Now Theorem 5.17 yields

$$\text{Log}(y(t)) - \text{Log}(y(t_0)) =_Q \mathcal{S}_{t_0}^t(\Delta t).$$

Therefore we have

$$\text{Log}(y(t)) =_Q \mathcal{S}_{t_0}^t(\Delta t) + \text{Log}(y(t_0)).$$

The initial condition of $y^\Delta = py$ will give us a numerical value for $y(t_0)$ so we have now isolated the unknown function $y(t)$ on the left inside of the Log function. So invert the logarithm on both sides to get

$$y(t) = \exp\left(_Q\mathcal{S}_{t_0}^t(\Delta t) + \text{Log}(y(t_0))\right).$$

Now the Transfer Principle yields

$$^*y(t) = \exp\left(_Q^*\mathcal{S}_{t_0}^t(\Delta t) + \text{Log}(y(t_0))\right),$$

and no star is included on the term $\text{Log}(y(t_0))$ since it is real-valued. This derivation suggests the following definition.

Definition 5.19. If p is a regressive function and $s, t \in \mathbb{T}$, define the *exponential function* $e_p(t, t_0) = \text{st}\left(\exp\left(_Q^*\mathcal{S}_{t_0}^t(\Delta t) + \text{Log}(y(t_0))\right)\right)$ where $y(t_0) \neq 0$.

Before we continue, we need a lemma.

Lemma 5.20. *Let $\mathbb{T}$ be a bounded time scale, Δt be an infinitesimal, $t \in \mathbb{T}$, and $f \colon \mathbb{T} \to \mathbb{R}$ be a continuous function. Then, ${}^*_f \mathcal{S}_t^{\sigma(t)}(\Delta t) = {}^*f(t)\mu(t)$.*

Proof. First suppose $\Delta t \in \mathbb{R}$. Then the time scale $P(\Delta t) = \{t_0, \ldots, t_{M(\Delta t)}\}$ is finite, so $\mu(t_i) = t_{i+1} - t_i$ is real-valued for all $t \in \mathbb{T}$ and $\mathrm{st}({}^*f(t)) = f(t)$. Let $\sigma_{P(\Delta t)}$ be defined as the jump operator on $P(\Delta t)$. We now compute from definition:

$$
{}_f\mathcal{S}_t^{\sigma_{P(\Delta t)}(t)}(\Delta t) = \sum_{i=0}^{1} f(t_i)\mu_{P(\Delta t)}(t_i)
$$
$$
= f(t)\mu_{P(\Delta t)}(t).
$$

We have established the truth of the sentence

$$
\text{``for all } \Delta t \in \mathbb{R}, \ {}_f\mathcal{S}_t^{\sigma_{P(\Delta t)}(t)}(\Delta t) = f(t)\mu_{P(\Delta t)}(t)\text{''}.
$$

Let μ be the graininess function $\mathbb{T}$ and σ be the jump operator of $\mathbb{T}$. Finally, for infinitesimal Δt, the Transfer Principle yields

$$
{}^*_f\mathcal{S}_t^{\sigma(t)}(\Delta t) = {}^*f(t)\mu(t). \quad \square
$$

We will now prove $e_p(t, t_0)$ is a solution to $y^\Delta = p(t)y$ through a direct computation similar to [2, p.28].

Theorem 5.21. *Let $\mathbb{T}$ be a time scale, $t, t_0 \in \mathbb{T}$, and $p(t)$ a regressive function. The function $e_p(t, t_0)$ satisfies the dynamic equation $y^\Delta(t) = p(t)y(t)$.*

Proof. Now assume that Δt is an infinitesimal. This forces us to turn all the graininess functions $\mu_{P(\Delta t)}$ to μ and we get by definition and then factoring

$$y^\Delta(t) = \mathrm{st}\left(\frac{\exp\left({}^*_Q\mathcal{S}^{\sigma(t)}_{t_0}(\Delta t) + \mathrm{Log}(y(t_0))\right) - \exp\left({}^*_Q\mathcal{S}^t_{t_0}(\Delta t) + \mathrm{Log}(y(t_0))\right)}{\mu(t)}\right)$$

$$= \mathrm{st}\left(\exp\left({}^*_Q\mathcal{S}^t_{t_0}(\Delta t) + \mathrm{Log}(y(t_0))\right)\right)\mathrm{st}\left(\frac{\exp\left({}^*_Q\mathcal{S}^{\sigma(t)}_t(\Delta t)\right) - 1}{\mu(t)}\right).$$

Now using Lemma 5.20,

$$y^\Delta(t) = \mathrm{st}\,({}^*y(t))\,\mathrm{st}\left(\frac{\exp\left(Q(t)\mu(t)\right) - 1}{\mu(t)}\right)$$

$$= \mathrm{st}\,({}^*y(t))\,\mathrm{st}\left(\frac{\exp\left(\frac{\mathrm{Log}(p(t)\mu(t)+1)}{\mu(t)}\mu(t)\right) - 1}{\mu(t)}\right)$$

$$= \mathrm{st}\,({}^*y(t))\,\mathrm{st}\left(\frac{\exp\left(\mathrm{Log}(p(t)\mu(t) + 1)\right) - 1}{\mu(t)}\right).$$

If we define $\xi_h(z) := \frac{\mathrm{Log}(1+zh)}{h}$, then it has inverse $\xi_h^{-1}(z) := \frac{e^{zh}-1}{h}$ and so

$$y^\Delta(t) = \mathrm{st}\,({}^*y(t))\,\mathrm{st}\left(\xi_{\mu(t)}(\xi_{\mu(t)}^{-1}(p(t)))\right)$$

$$= \mathrm{st}\,({}^*y(t))\,\mathrm{st}\,(p(t))$$

$$= y(t)p(t). \qquad \square$$

We started off by defining derivatives on arbitrary time scales and then deduced a way to reverse engineer a derivative with integrals. So in our study of abstract

change, we have come full circle: we can take an equation derived from observing changes and then deduce specific information about what we have observed and conversely we can take specific information and then deduce how that information changes.

6. CONCLUSIONS

We have seen that the language of sets can be used to define numbers, which we use to define slope, which is the basis of calculus. We then showed that calculus itself can be greatly generalized.

Logic and set theory are the foundations of mathematical theories and are the basis of its rigor. Our language of sets was frequently justified with and compared to the English language. It was defined so that we could have the freedom to write the proceeding mathematics in a fluid, versatile way.

Numbers are popularly thought of as abstract measures of quantity, but our language of sets was used to define them concretely. This construction allows us to answer the question "What is the number 2, *really*?" with the simple answer "the set $\{0, 1\}$", which although pleasing, is not why we constructed them from sets. We did so in order to introduce the construction of $^*\mathbb{R}$, which we felt necessary due to the numerous objections to their use classically, and to introduce the methods of proof with hyperreal numbers to justify the infinitesimal time scale calculus.

Calculus is the abstract study of change. We studied differential calculus for the practice of using infinitesimals in calculus, and we studied the discrete calculus to better understand stepsize. We saw that these two theories have common properties, which inspired us to investigate a possible generalization.

Time scale calculus unites and extends differential and difference calculus. Proofs in traditional time scale calculus often require splitting proofs up into arguments for isolated points and arguments for dense points. We believe that the more the proofs resembled algebraic techniques, the easier they would be to understand, so we had to treat dense points as if they were discrete. That choice naturally led us to consider hyperreal numbers. We then defined in generality the delta-derivative operator and proceeded to study integration. After that, we used the theories of derivatives and integrals to study dynamic equations over an arbitrary time scale, and then derived the generalized exponential function.

We can now see that since the following quotation was written, the time scales calculus has addressed precisely its concern.

"A major task of mathematics today is to harmonize the continuous and the discrete, to include them in one comprehensive mathematics, and to eliminate obscurity from both."

- E.T. Bell, *Men of Mathematics* (1937) [1]

BIBLIOGRAPHY

[1] Bell, E.T. Men of Mathematics. Simon and Schuster, 1937.

[2] Bohner, Martin and Peterson, Allan. Dynamic Equations on Time Scales: An Introduction with Applications, Birkhauser, New York, 2001.

[3] Euclid. The Thirteen Books of The Elements. Trans. Sir Thomas L. Heath. 3 Volumes. Dover Publications.

[4] Goldblatt, Robert. Lectures on the Hyperreals: An Introduction to Nonstandard Analysis. Springer, 1998.

[5] Hilger, S. Ein Maβkettenkalül mit Anwendung auf Zentrumsmannigfaltigkeiten. Phd Thesis, Universität Würzburg, 1988.

[6] Keisler, Jerome H. Elementary Calculus: An Infinitesimal Approach, Second Edition. Creative Commons Attribution-NonCommercial-Share Alike License, 2007. (First published 1976)

[7] ———————. Foundations of Infinitesimal Calculus. Creative Commons Attribution-Noncommercial-Share Alike 2.5 License, 2009.

[8] Krantz, Steven G. An Episodic History of Mathematics: Mathematical Culture Through Problem Solving. Mathematical Association of America, 2010.

[9] Reyes, Mitchell G. The Rhetoric in Mathematics: Newton, Leibniz, the Calculus, and the Rhetorical Force of the Infinitesimal. Quarterly Journal of Speech, Vol. 90, No. 2, May 2004, pp. 163-188.

[10] Rudin, Walter. Principles of mathematical analysis. McGraw-Hill Inc., 1976.

[11] Spivak, Michael. Calculus, Third Edition. Publish or Perish, Inc. Houston, Texas, 1994.