FOUNDATIONS OF INFINITESIMAL CALCULUS

H. JEROME KEISLER
Department of Mathematics
University of Wisconsin, Madison, Wisconsin, USA
keisler@math.wisc.edu

February 20, 2009

CONTENTS

PREFACE vii

CHAPTER 1. THE HYPERREAL NUMBERS 1
- 1A. Structure of the Hyperreal Numbers (§1.4, §1.5) 1
- 1B. Standard Parts (§1.6) 5
- 1C. Axioms for the Hyperreal Numbers (§Epilogue) 7
- 1D. Consequences of the Transfer Axiom 9
- 1E. Natural Extensions of Sets 14
- 1F. Appendix. Algebra of the Real Numbers 19
- 1G. Building the Hyperreal Numbers 23

CHAPTER 2. DIFFERENTIATION 31
- 2A. Derivatives (§2.1, §2.2) 31
- 2B. Infinitesimal Microscopes and Infinite Telescopes 33
- 2C. Properties of Derivatives (§2.3, §2.4) 36
- 2D. Chain Rule (§2.6, §2.7) 39

CHAPTER 3. CONTINUOUS FUNCTIONS 41
- 3A. Limits and Continuity (§3.3, §3.4) 41
- 3B. Hyperintegers (§3.8) 45
- 3C. Properties of Continuous Functions (§3.5–§3.8) 47

CHAPTER 4. INTEGRATION 57
- 4A. The Definite Integral (§4.1) 57
- 4B. Fundamental Theorem of Calculus (§4.2) 62
- 4C. Second Fundamental Theorem of Calculus (§4.2) 65

CHAPTER 5. LIMITS 69
- 5A. ε, δ Conditions for Limits (§5.8, §5.1) 69
- 5B. L'Hospital's Rule (§5.2) 72

CHAPTER 6. APPLICATIONS OF THE INTEGRAL 75
- 6A. Infinite Sum Theorem (§6.1, §6.2, §6.6) 75
- 6B. Lengths of Curves (§6.3, §6.4) 80
- 6C. Improper Integrals (§6.7) 85

Chapter 7. Trigonometric Functions 89
7A. Inverse Function Theorem (§7.3) 89
7B. Derivatives of Trigonometric Functions (§7.1, §7.2) 92
7C. Area in Polar Coordinates (§7.9) 93

Chapter 8. Exponential Functions 97
8A. Extending Continuous Functions 97
8B. The Functions a^x and $\log_b x$ (§8.1, §8.2) 98
8C. Derivatives of Exponential Functions (§8.3) 100

Chapter 9. Infinite Series 103
9A. Sequences (§9.1) 103
9B. Series (§9.2 – §9.6) 106
9C. Taylor's Formula and Higher Differentials (§9.10) 108

Chapter 10. Vectors 113
10A. Hyperreal Vectors (§10.8) 113
10B. Vector Functions (§10.6) 116

Chapter 11. Partial Differentiation 119
11A. Continuity in Two Variables (§11.1, §11.2) 119
11B. Partial Derivatives (§11.3, §11.4) 120
11C. Chain Rule and Implicit Functions (§11.5, §11.6) 123
11D. Maxima and Minima (§11.7) 126
11E. Second Partial Derivatives (§11.8) 131

Chapter 12. Multiple Integration 135
12A. Double Integrals (§12.1, §12.2) 135
12B. Infinite Sum Theorem for Two Variables (§12.3) 138
12C. Change of Variables in Double Integrals (§12.5) 142

Chapter 13. Vector Calculus 149
13A. Line Integrals (§13.2) 149
13B. Green's Theorem (§13.3, §13.4) 152

Chapter 14. Differential Equations 159
14A. Existence of Solutions (§14.4) 160
14B. Uniqueness of Solutions (§14.4) 165
14C. An Example where Uniqueness Fails (§14.3) 169

Chapter 15. Logic and Superstructures 173
15A. The Elementary Extension Principle 173
15B. Superstructures 178
15C. Standard, Internal, and External Sets 182
15D. Bounded Ultrapowers 187
15E. Saturation and Uniqueness 192

References 197

Index 199

PREFACE

In 1960 Abraham Robinson (1918–1974) solved the three hundred year old problem of giving a rigorous development of the calculus based on infinitesimals. Robinson's achievement was one of the major mathematical advances of the twentieth century. This is an exposition of Robinson's infinitesimal calculus at the advanced undergraduate level. It is entirely self-contained but is keyed to the 2000 edition of my first year college text *Elementary Calculus* [Keisler 2000]. *Elementary Calculus* is available free online at www.math.wisc.edu/∼Keisler. This monograph can be used as a quick introduction to the subject for mathematicians, as background material for instructors using the book *Elementary Calculus*, or as a text for an undergraduate seminar.

This is a major revision of the first edition of *Foundations of Infinitesimal Calculus* [Keisler 1976], which was published as a companion to the first (1976) edition of *Elementary Calculus*, and has been out of print for over twenty years. A companion to the second (1986) edition of *Elementary Calculus* was never written. The biggest changes are: (1) A new chapter on differential equations, keyed to the corresponding new chapter in *Elementary Calculus*. (2) The axioms for the hyperreal number system are changed to match those in the later editions of *Elementary Calculus*. (3) An account of the discovery of Kanovei and Shelah [KS 2004] that the hyperreal number system, like the real number system, can be built as an explicitly definable mathematical structure. Earlier constructions of the hyperreal number system depended on an arbitrarily chosen parameter such as an ultrafilter.

The basic concepts of the calculus were originally developed in the seventeenth and eighteenth centuries using the intuitive notion of an infinitesimal, culminating in the work of Gottfried Leibniz (1646-1716) and Isaac Newton (1643-1727). When the calculus was put on a rigorous basis in the nineteenth century, infinitesimals were rejected in favor of the ε, δ approach, because mathematicians had not yet discovered a correct treatment of infinitesimals. Since then generations of students have been taught that infinitesimals do not exist and should be avoided.

The actual situation, as suggested by Leibniz and carried out by Robinson, is that one can form the hyperreal number system by adding infinitesimals to the real number system, and obtain a powerful new tool in analysis. The reason Robinson's discovery did not come sooner is that the axioms needed to describe the hyperreal numbers are of a kind which were unfamiliar to mathematicians until the mid-twentieth century. Robinson used methods from the branch of mathematical logic called model theory which developed in the 1950's.

Robinson called his method nonstandard analysis because it uses a nonstandard model of analysis. The older name infinitesimal analysis is perhaps more appropriate.

The method is surprisingly adaptable and has been applied to many areas of pure and applied mathematics. It is also used in such fields as economics and physics as a source of mathematical models. (See, for example, the books [AFHL 1986] and [ACH 1997]). However, the method is still seen as controversial, and is unfamiliar to most mathematicians.

The purpose of this monograph, and of the book *Elementary Calculus*, is to make infinitesimals more readily available to mathematicians and students. Infinitesimals provided the intuition for the original development of the calculus and should help students as they repeat this development. The book *Elementary Calculus* treats infinitesimal calculus at the simplest possible level, and gives plausibility arguments instead of proofs of theorems whenever it is appropriate. This monograph presents the subject from a more advanced viewpoint and includes proofs of almost all of the theorems stated in *Elementary Calculus*.

Chapters 1–14 in this monograph match the chapters in *Elementary Calculus*, and after each section heading the corresponding sections of *Elementary Calculus* are indicated in parentheses.

In Chapter 1 the hyperreal numbers are first introduced with a set of axioms and their algebraic structure is studied. Then in Section 1G the hyperreal numbers are built from the real numbers. This is an optional section which is more advanced than the rest of the chapter and is not used later. It is included for the reader who wants to see where the hyperreal numbers come from.

Chapters 2 through 14 contain a rigorous development of infinitesimal calculus based on the axioms in Chapter 1. The only prerequisites are the traditional three semesters of calculus and a certain amount of mathematical maturity. In particular, the material is presented without using notions from mathematical logic. We will use some elementary set-theoretic notation familiar to all mathematicians, for example the function concept and the symbols $\emptyset, A \cup B, \{x \in A\colon P(x)\}$.

Frequently, standard results are given alternate proofs using infinitesimals. In some cases a standard result which is beyond the scope of beginning calculus is rephrased as a simpler infinitesimal result and used effectively in *Elementary Calculus*; some examples are the Infinite Sum Theorem, and the two-variable criterion for a global maximum.

The last chapter of this monograph, Chapter 15, is a bridge between the simple treatment of infinitesimal calculus given here and the more advanced subject of infinitesimal analysis found in the research literature. To go beyond infinitesimal calculus one should at least be familiar with some basic notions from logic and model theory. Chapter 15 introduces the concept of a nonstandard universe, explains the use of mathematical logic, superstructures, and internal and external sets, uses ultrapowers to build a nonstandard universe, and presents uniqueness theorems for the hyperreal number systems and nonstandard universes.

The simple set of axioms for the hyperreal number system given here (and in *Elementary Calculus*) make it possible to present infinitesimal calculus at the college freshman level, avoiding concepts from mathematical logic. It is shown in Chapter 15 that these axioms are equivalent to Robinson's approach.

For additional background in logic and model theory, the reader can consult the book [CK 1990]. Section 4.4 of that book gives further results on nonstandard universes. Additional background in infinitesimal analysis can be found in the book [Goldblatt 1991].

I thank my late colleague Jon Barwise, and Keith Stroyan of the University of Iowa, for valuable advice in preparing the First Edition of this monograph. In the thirty years between the first and the present edition, I have benefited from equally valuable and much appreciated advice from friends and colleagues too numerous to recount here,

CHAPTER 1

THE HYPERREAL NUMBERS

We will assume that the reader is familiar with the real number system and develop a new object, called a hyperreal number system. The definition of the real numbers and the basic existence and uniqueness theorems are briefly outlined in Section 1F, near the end of this chapter. That section also explains some useful notions from modern algebra, such as a ring, a complete ordered field, an ideal, and a homomorphism. If any of these terms are unfamiliar, you should read through Section 1F. We do not require any knowledge of modern algebra except for a modest vocabulary.

In Sections 1A–1E we introduce axioms for the hyperreal numbers and obtain some first consequences of the axioms. In the optional Section 1G at the end of this chapter we build a hyperreal number system as an ultrapower of the real number system. This proves that there exists a structure which satisfies the axioms. We conclude the chapter with the construction of Kanovei and Shelah [KS 2004] of a hyperreal number system which is definable in set theory. This shows that the hyperreal number system exists in the same sense that the real number system exists.

1A. Structure of the Hyperreal Numbers (§1.4, §1.5)

In this and the next section we assume only Axioms A, B, and C below.

Axiom A
$\mathbb{R}$ is a complete ordered field.

Axiom B
$\mathbb{R}^*$ is an ordered field extension of $\mathbb{R}$.

Axiom C
$\mathbb{R}^*$ has a positive infinitesimal, that is, an element ε such that $0 < \varepsilon$ and $\varepsilon < r$ for every positive $r \in \mathbb{R}$.

In the next section we will introduce two powerful additional axioms which are needed for our treatment of the calculus. However, the algebraic facts about infinitesimals which underlie the intuitive picture of the hyperreal line follow from Axioms A–C alone.

We call $\mathbb{R}$ the field of **real numbers** and $\mathbb{R}^*$ the field of **hyperreal numbers**.

Definition 1.1. *An element $x \in \mathbb{R}^*$ is*
infinitesimal *if $|x| < r$ for all positive real r;*
finite *if $|x| < r$ for some real r;*
infinite *if $|x| > r$ for all real r.*

Two elements $x, y \in \mathbb{R}^*$ are said to be **infinitely close**, *$x \approx y$, if $x - y$ is infinitesimal. (Thus x is infinitesimal if and only if $x \approx 0$).*

Notice that a positive infinitesimal is hyperreal but not real, and that the only real infinitesimal is 0.

Definition 1.2. *Given a hyperreal number $x \in \mathbb{R}^*$, the* **monad** *of x is the set*

$$\text{monad}(x) = \{y \in \mathbb{R}^* \colon x \approx y\}.$$

The **galaxy** *of x is the set*

$$\text{galaxy}(x) = \{y \in \mathbb{R}^* \colon x - y \textit{ is finite}\}.$$

Thus monad(0) is the set of infinitesimals and galaxy(0) is the set of finite hyperreal numbers.

In *Elementary Calculus*, the pictorial device of an infinitesimal microscope is used to illustrate part of a monad, and an infinite telescope is used to illustrate part of an infinite galaxy. Figure 1 shows how the hyperreal line is drawn. In Section 2B we will give a rigorous treatment of infinitesimal microscopes and telescopes so the instructor can use them in new situations.

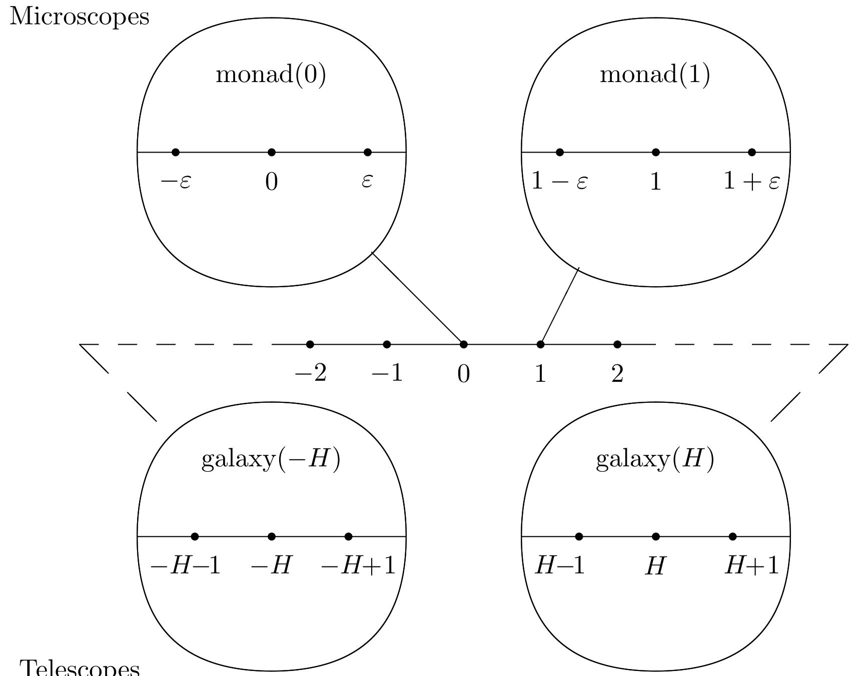

Figure 1

We now describe the algebraic structure of $\mathbb{R}^*$.

Theorem 1.3. *The set* galaxy(0) *of finite elements is a subring of* $\mathbb{R}^*$*, that is, sums, differences, and products of finite elements are finite.*

Proof. Suppose x and y are finite, say

$$|x| < r, \quad |y| < s$$

where r and s are real. Then

$$|x + y| < r + s, \quad |x - y| < r + s, \quad |xy| < rs,$$

so $x + y, x - y$, and xy are finite. ⊣

Corollary 1.4. *Any two galaxies are either equal or disjoint.*

Proof. For each $x \in \mathbb{R}^*$, the galaxy of x is the coset of x modulo galaxy(0),

$$\text{galaxy}(x) = \{x + a \colon a \in \text{galaxy}(0)\}.$$

⊣

Theorem 1.5. *The set* monad(0) *of infinitesimal elements is a subring of* $\mathbb{R}^*$ *and an ideal in* galaxy(0)*. That is:*

(i) Sums, differences, and products of infinitesimals are infinitesimal.

(ii) The product of an infinitesimal and a finite element is infinitesimal.

Proof. Let $\varepsilon, \delta \approx 0$. For each positive real r, $|\varepsilon| < r/2$, $|\delta| < r/2$, whence $|\varepsilon + \delta| < r, |\varepsilon - \delta| < r$. Hence $\varepsilon + \delta$ and $\varepsilon - \delta$ are infinitesimal.

Let b be finite, say $|b| < t, 1 \leq t \in \mathbb{R}$. Then for any positive real r we have $|\varepsilon| < r/t, |\varepsilon b| < (r/t)t = r$. Therefore εb is infinitesimal. $\dashv$

Corollary 1.6. *Any two monads are equal or disjoint. The relation $x \approx y$ is an equivalence relation on $\mathbb{R}^*$.*

Proof. For each $x \in \mathbb{R}^*$, monad(x) is the coset of x modulo monad(0),

$$\text{monad}(x) = \{x + \varepsilon \colon \varepsilon \in \text{monad}(0)\}.$$

From the definition of monad and $x \approx y$ we see that

$$x \approx y \text{ if and only if } \text{monad}(x) = \text{monad}(y),$$

so $x \approx y$ is an equivalence relation. $\dashv$

Theorem 1.7. *(i) x is infinite if and only if x^{-1} is infinitesimal.*

(ii) monad(0) *is a maximal ideal in* galaxy(0). *That is, there is no ideal I in* galaxy(0) *such that* monad$(0) \subsetneqq I \subsetneqq$ galaxy(0).

Proof. (i) The following are equivalent:

$$|x| \geq r \text{ for all positive real } r.$$
$$|x^{-1}| \leq r^{-1} \text{ for all positive real } r.$$
$$x^{-1} \text{ is infinitesimal.}$$

(ii) Let I be an ideal containing monad(0) and let $b \in I \setminus$ monad(0). By (i), b^{-1} is not infinite since $b = (b^{-1})^{-1}$ is not infinitesimal. Then $b^{-1} \in$ galaxy(0), so $1 = b \cdot (b^{-1}) \in I$. Then for any $c \in$ galaxy(0), $1 \cdot c = c \in I$, so $I =$ galaxy(0). $\dashv$

Corollary 1.8. *(i) There exist negative infinitesimals in $\mathbb{R}^*$.*

(ii) $\mathbb{R}^*$ has positive and negative infinite elements.

Proof. Axiom C says that there exists a positive infinitesimal ε. The elements $-\varepsilon, 1/\varepsilon$, and $-1/\varepsilon$ are respectively negative infinitesimal, positive infinite, and negative infinite. $\dashv$

Actually, one can go on to show that there are infinitely many infinitesimals and infinitely many galaxies in $\mathbb{R}^*$. Moreover, each galaxy is partitioned into infinitely many monads. Each monad has a complicated structure; for example, the mapping $H \mapsto x + (H^{-1})$ maps the infinite elements of $\mathbb{R}^*$ one-one onto monad$(x) \setminus \{x\}$.

We caution the reader that monad(0) is an ideal in galaxy(0) but is only a subring, not an ideal, in $\mathbb{R}^*$. In other words, the product of an infinitesimal and an infinite element need not be infinitesimal. In fact, given a positive infinitesimal ε, we see that

$\varepsilon^2 \cdot 1/\varepsilon$ is infinitesimal,
$\varepsilon \cdot 1/\varepsilon$ is finite but not infinitesimal,
$\varepsilon \cdot 1/\varepsilon^2$ is infinite.

This corresponds to the intuitive principle that $0 \cdot \infty$ is an "indeterminate form".

1B. Standard Parts (§1.6)

We now state and prove the Standard Part Principle. This is the first place where we use the fact that the real numbers are complete. This principle was stated without proof in §1.6 of *Elementary Calculus*, but a proof was given in the Epilogue. It was often used in *Elementary Calculus* instead of the Completeness Property, which made some concepts more accessible to beginning students.

THEOREM 1.9. *(Standard Part Principle) Every finite $x \in \mathbb{R}^*$ is infinitely close to a unique real number $r \in \mathbb{R}$. That is, every finite monad contains a unique real number.*

PROOF. Let $x \in \mathbb{R}^*$ be finite.

Uniqueness: Suppose r and s are real and $r \approx x, s \approx x$. Since $\approx$ is an equivalence relation we have $r \approx s$, whence $r - s \approx 0$. But $r - s$ is real, so $r - s = 0$ and $r = s$.

Existence: Let $X = \{s \in \mathbb{R} \colon s < x\}$. X is nonempty and has an upper bound because there is a positive real number r such that $|x| < r$, whence $-r < x < r$, so $-r \in X$ and r is an upper bound of X. By Axiom A, $\mathbb{R}$ is a complete ordered field, so the set X has a least upper bound t. For every positive real r we have

$$x \le t + r, \quad x - t \le r$$

and

$$t - r \le x, \quad -(x - t) \le r.$$

It follows that $x - t \approx 0$, so $x \approx t$. ⊣

DEFINITION 1.10. *Given a finite $x \in \mathbb{R}^*$, the unique real $r \approx x$ is called the* **standard part** *of x, in symbols $r = \mathrm{st}(x)$. If x is infinite, $\mathrm{st}(x)$ is undefined.*

COROLLARY 1.11. *Let x and y be finite.*
(i) $x \approx y$ if and only if $\mathrm{st}(x) = \mathrm{st}(y)$.
(ii) $x \approx \mathrm{st}(x)$.
(iii) If $r \in \mathbb{R}$ then $\mathrm{st}(r) = r$.
(iv) If $x \le y$ then $\mathrm{st}(x) \le \mathrm{st}(y)$.

PROOF. (iv) We have

$$x = \mathrm{st}(x) + \varepsilon, \quad y = \mathrm{st}(y) + \delta$$

for some infinitesimal ε and δ. Assume $x \le y$. Then

$$\begin{aligned} \mathrm{st}(x) + \varepsilon &\le \mathrm{st}(y) + \delta, \\ \mathrm{st}(x) &\le \mathrm{st}(y) + (\delta - \varepsilon). \end{aligned}$$

For any positive real r,

$$\begin{aligned} \delta - \varepsilon &< r, \\ \mathrm{st}(x) &< \mathrm{st}(y) + r, \end{aligned}$$

and therefore

$$\mathrm{st}(x) \le \mathrm{st}(y).$$

⊣

The next theorem gives the basic algebraic rules for standard parts.

THEOREM 1.12. *The standard part function is a homomorphism of the ring* galaxy(0) *onto the field of real numbers. That is, for finite x and y,*
(i) $\mathrm{st}(x+y) = \mathrm{st}(x) + \mathrm{st}(y)$,
(ii) $\mathrm{st}(x-y) = \mathrm{st}(x) - \mathrm{st}(y)$,
(iii) $\mathrm{st}(xy) = \mathrm{st}(x)\mathrm{st}(y)$.

PROOF. Let $x = r + \varepsilon, y = s + \delta$ where $r = \mathrm{st}(x), s = \mathrm{st}(y)$. Then ε and δ are infinitesimal.
(i)

$$\begin{aligned} \mathrm{st}(x+y) &= \mathrm{st}((r+\varepsilon) + (s+\delta)) \\ &= \mathrm{st}((r+s) + (\varepsilon+\delta)) \\ &= r + s. \end{aligned}$$

(ii) Similar to (i).
(iii)

$$\begin{aligned} \mathrm{st}(xy) &= \mathrm{st}((r+\varepsilon)(s+\delta)) \\ &= \mathrm{st}(rs + r\delta + s\varepsilon + \varepsilon\delta) \\ &= rs, \end{aligned}$$

since $r\delta + s\varepsilon + \varepsilon\delta$ is infinitesimal.
Thus $\mathrm{st}(\cdot)$ is a homomorphism of galaxy(0) into $\mathbb{R}$. It is obviously onto $\mathbb{R}$, because for $r \in \mathbb{R}, \mathrm{st}(r) = r$. ⊣

COROLLARY 1.13. *Let x and y be finite.*
(i) If $\mathrm{st}(y) \ne 0$ *then* $\mathrm{st}(x/y) = \mathrm{st}(x)/\mathrm{st}(y)$.
(ii) If $x \ge 0$ and $y = \sqrt[n]{x}$ then $\mathrm{st}(y) = \sqrt[n]{\mathrm{st}(x)}$.

PROOF. (i) This follows from the computation

$$\mathrm{st}(x) = \mathrm{st}((x/y)\cdot y) = \mathrm{st}(x/y)\cdot \mathrm{st}(y).$$

(ii) We have $y^n = x$ and $y \ge 0$. Taking standard parts,

$$\mathrm{st}(x) = \mathrm{st}(y^n) = (\mathrm{st}(y))^n,$$

and $\mathrm{st}(y) \ge 0$, so $\mathrm{st}(y) = \sqrt[n]{\mathrm{st}(x)}$. ⊣

1C. Axioms for the Hyperreal Numbers (§Epilogue)

The properties of a hyperreal number system are given by five axioms. The first three of these axioms were stated in Section 1A. Before giving a precise statement of the remaining two axioms, we describe the intuitive idea. The real and hyperreal numbers are related by a $*$ mapping such that:

(1) With each relation X on $\mathbb{R}$ there is a corresponding relation X^* on $\mathbb{R}^*$, called the **natural extension** of X.

(2) The relations on $\mathbb{R}$ have the same "elementary properties" as their natural extensions on $\mathbb{R}^*$.

The difficulty in making (2) precise is that we must explain exactly what an "elementary property" is. The properties "$X \subseteq Y$", "X is a function", and "X is a symmetric binary relation" are elementary. On the other hand, the Archimedean Property and the Completeness Property must not be elementary, because no proper extension of $\mathbb{R}$ is an Archimedean or complete ordered field. In most expositions of the subject an "elementary property" is taken to be any sentence in first order logic. However, it is not appropriate to begin a calculus course by explaining first order logic to the students because they have not yet been exposed to the right sort of examples. It is better to learn calculus first, and at some later time use the ε, δ conditions from calculus as meaningful examples of sentences in first order logic. Fortunately, the notion of a sentence of first order logic is not necessary at all in stating the axioms. It turns out that a simpler concept which is within the experience of beginning students is sufficient. This is the concept of a (finite) system of equations and inequalities. We shall see in Chapter 15 at the end of this monograph that we get equivalent sets of axioms using either the simple concept of a system of equations and inequalities or the more sophisticated concept of a sentence of first order logic.

The main objects of study in elementary calculus are partial functions of n real variables. Our plan is to have an axiom corresponding to (1) above but for partial functions instead of relations, and then an axiom corresponding to (2) above where an "elementary property" means a system of equations and inequalities in these functions.

In the next few paragraphs we will explain exactly what is meant by a system of equations and inequalities.

$\mathbb{R}^n$ denotes the set of all n-tuples of elements of $\mathbb{R}$. A **real function** of n variables is a mapping f from a subset of $\mathbb{R}^n$ into $\mathbb{R}$.

The letters $x, y, z, \ldots$ are called **variables**. We think of them as varying over the set $\mathbb{R}$ of real numbers. A real number $c \in \mathbb{R}$ is also called a real **constant**. Expressions built up from variables and constants by applying real functions are called terms. For example,

$$x, \quad c, \quad x + c, \quad f(x), \quad g(c, x, f(y))$$

are terms. Here is a precise definition of a term.

A **term** is an expression which can be built up using the following rules:

- Every variable is a term.
- Every real constant is a term.
- If $\tau_1, \dots, \tau_n$ are terms and f is a real function of n variables, then $f(\tau_1, \dots, \tau_n)$ is a term.

A term which contains no variables is called a **constant term**. A constant term is either undefined or has value equal to some real number. In particular, $f(c)$ is undefined if c is not in the domain of f, and has value d if $f(c) = d$. The value of a constant term is computed step by step; thus the value of

$$g(f(5) + 3, f(f(2)))$$

is computed by first computing $f(5)$, then $f(5) + 3$, then $f(2)$, then $f(f(2))$, and then $g(f(5) + 3, f(f(2)))$. If at any stage in the computation we reach an undefined part, the whole constant term is considered to be undefined. For example, since $\sqrt{-6}$ is undefined in the reals, the constant terms $\sqrt{-6} - \sqrt{-6}$ and $\sqrt{-6} \cdot \sqrt{-6}$ are also undefined in the reals (if we were working over the complex numbers, these terms would be defined and have real values 0 and -6 respectively, but in this monograph we will always be working over the reals.)

By an **equation** we mean an expression $\sigma = \tau$ where σ and τ are terms. By an **inequality** we mean an expression of one of the forms

$$\sigma \leq \tau, \quad \sigma < \tau, \quad \sigma \neq \tau$$

where σ and τ are terms. By a **formula** we mean an equation or inequality between two terms, or a statement or the form *"τ is defined"* or of the form *"τ is undefined"*. A **system of formulas** is a nonempty finite set of formulas.

The last notion we need is that of a solution of a system of formulas. Consider a system S of formulas whose variables are $x_1, \dots, x_n$. By a **real solution** of S we mean an n-tuple $(c_1, \dots, c_n)$ of real constants such that when the x_i are replaced by c_i in S, each term within an equation or inequality in S is defined, and each formula in S is true.

The notion of a system of formulas is easily motivated in an elementary calculus course, because the first step in solving a "story problem" is to set the problem up as a system of formulas.

The statements of the forms *"τ is defined"* and *"τ is undefined"* are allowed for convenience, but they are not really needed because one can always use an equation to say that a term is defined or is undefined. To see this, let $\tau(x_1, \dots, x_n)$ be a term with the variables $x_1, \dots, x_n$. Now let $g(x_1, \dots, x_n)$ be the real function such that for each n-tuple of real constants $(c_1, \dots, c_n)$,

$$g(c_1, \dots, c_n) = \begin{cases} 1, & \text{if } \tau(c_1, \dots, c_n) \text{ is defined,} \\ 0, & \text{if } \tau(c_1, \dots, c_n) \text{ is undefined.} \end{cases}$$

Then $\tau(c_1,\dots,c_n)$ is defined if and only if $(c_1,\dots,c_n)$ is a solution of the equation $g(x_1,\dots,x_n)=1$, and $\tau(c_1,\dots,c_n)$ is undefined if and only if $(c_1,\dots,c_n)$ is a solution of the equation $g(x_1,\dots,x_n)=0$.

The following axioms describe a hyperreal number system as a triple $(*,\mathbb{R},\mathbb{R}^*)$, where $\mathbb{R}$ is called the field of **real numbers**, $\mathbb{R}^*$ the field of **hyperreal numbers**, and $*$ the **natural extension mapping**.

Axiom A

$\mathbb{R}$ *is a complete ordered field.*

Axiom B

$\mathbb{R}^*$ *is an ordered field extension of* $\mathbb{R}$.

Axiom C

$\mathbb{R}^*$ *has a positive infinitesimal.*

Axiom D (Function Axiom)

For each real function f of n variables there is a corresponding hyperreal function f^* of n variables, called the **natural extension** *of f. The field operations of $\mathbb{R}^*$ are the natural extensions of the field operations of $\mathbb{R}$.*

By a **hyperreal solution** of a system of formulas S with the variables $x_1,\dots,x_n$ we mean an n-tuple $(c_1,\dots,c_n)$ of hyperreal numbers such that all the formulas in S are true when each function is replaced by its natural extension and each x_i is replaced by c_i.

Axiom E (Transfer Axiom)

Given two systems of formulas S,T with the same variables, if every real solution of S is a solution of T, then every hyperreal solution of S is a solution of T.

1D. Consequences of the Transfer Axiom

The Transfer Axiom E is much more powerful than it looks. We will postpone a full explanation of its scope to Chapter 15 at the end of this monograph, because this explanation requires notions from logic. In this section, will give some easy but general consequences of the Transfer Axiom which will facilitate our development of the calculus.

Our first corollary shows that the Transfer Axiom still holds if T has fewer variables than S. If T has variables $x_1,\dots,x_k$, we say that a tuple $(c_1,\dots,c_k,\dots,c_n)$ **contains a solution** of T if $(c_1,\dots,c_k)$ is a solution of T.

Corollary 1.14. *Given two systems of formulas S, T such that the variables in T form a subset of the variables in S. If every real solution of S contains a solution of T, then every hyperreal solution of S contains a solution of T.*

Proof. Let the variables of S be $x_1, \dots, x_n$. Form the system of formulas T' by adding to T the trivial equations $x_1 = x_1, \dots, x_n = x_n$. T' has the same meaning as T but has the same variables as S. Every real solution of S is a solution of T' because it contains a solution of T. By Transfer, every hyperreal solution of S is a hyperreal solution of T', and thus contains a hyperreal solution of T. ⊣

Corollary 1.15. *Any two systems of formulas with the same variables which have the same real solutions have the same hyperreal solutions. (This was called the Solution Axiom in the 1976 edition).*

Proof. Suppose S and T are systems of formulas with the same real solutions. Then every real solution of S is a solution of T. By Transfer, every hyperreal solution of S is a solution of T. Similarly, every hyperreal solution of T is a solution of S. ⊣

Often a real function f is defined by a rule of the form

$$f(x_1, \dots, x_n) = y \text{ if and only if } S$$

where S is a system of formulas with the same variables. If a real function f is defined by a rule of this kind, then by the above theorem, the natural extension f^* is defined by the same rule applied to the hyperreal numbers.

For example, the square root function on the reals is defined by the rule

$$\sqrt{x} = y \text{ if and only if } \{y^2 = x,\ 0 \le y\}.$$

By Transfer, the natural extension of the square root function is defined by the same rule where x and y vary over the hyperreal numbers.

Corollary 1.16. *(i) If a system S of formulas is true for all real numbers, it is also true for all hyperreal numbers.*

(ii) If a system of formulas has no real solutions, it has no hyperreal solutions.

Proof. (i) Let S have the variables $x_1, \dots, x_n$ and let T be the system of equation $x_1 = x_1, \dots, x_n = x_n$. Since S is true for all real numbers, S has the same real solutions as T. By Corollary 1.15, S has the same hyperreal solutions as T, and therefore S is true for all hyperreal numbers.

(ii) The proof is similar but uses the inequality $x_i \neq x_i$ instead of the equation $x_i = x_i$. ⊣

Corollary 1.17. *Let f be a real function of n variables and let $c_1, \dots, c_n$ be real constants. If $f(c_1, \dots, c_n)$ is defined then $f^*(c_1, \dots, c_n) = f(c_1, \dots, c_n)$.*

Proof. Let $f(c_1, \dots, c_n) = c$. The system of formulas

$$f(c_1, \dots, c_n) = c, \quad x = x$$

is true for all real numbers x. By Corollary 1.16 it is true for all hyperreal numbers. Therefore $f^*(c_1, \dots, c_n) = c$. $\dashv$

From now on, we will usually drop the asterisk on the natural extension f^* and write $f(c_1, \dots, c_n)$ for the value of f^* at $(c_1, \dots, c_n)$, whether the c_i's are real or hyperreal. By Corollary 1.17, this will cause no trouble. Occasionally, we will still put in the asterisk if we wish to call attention to the fact that we are working with the natural extension rather than the original real function.

The axioms in *Elementary Calculus* were stated in a more leisurely way that avoided the phrase "ordered field". Axioms A, C, D, and E are the same in *Elementary Calculus* as here. But instead of Axiom B, *Elementary Calculus* had the weaker axiom that $\mathbb{R}^*$ is an extension of $\mathbb{R}$ and the relation $<^*$ satisfies the Transitive Law

$$a <^* b \text{ and } b <^* c \text{ implies } a <^* c$$

and the Trichotomy Law, which says that forall $a, b, c \in \mathbb{R}^*$, exactly one of

$$a <^* b,\ a = b,\ b <^* a$$

holds. We will now use the Transfer Axiom to show that the axioms in *Elementary Calculus* are equivalent to the present Axioms A—E.

Proposition 1.18. *Assume Axioms A, C, D, E, and also that $\mathbb{R}^*$ with the relation $<^*$ and the functions $+, -, \cdot, ^{-1}$ is an extension of $\mathbb{R}$ which satisfies the Trichotomy Law. Then $\mathbb{R}^*$ is an ordered field, so Axiom B holds.*

Proof. First, observe that the proof of Corollary 1.16 did not use Axiom B, so it follows from the remaining axioms. By definition, an ordered field is a structure with a relation $\le$ and operations $+, -, \cdot, ^{-1}$ that satisfies the laws stated Section 1E. Each of the laws for a field is an equation, except for the inequality $0 \ne 1$. Since $\mathbb{R}$ is a field by Axiom A, each of these laws is true for $\mathbb{R}$. By Corollary 1.16, each of these laws is true for $\mathbb{R}^*$, so $\mathbb{R}^*$ is a field. Except for the Trichotomy Law, each of the order laws is of the form "if S then T" where S, T are systems of formulas. For example, the Sum Law says that if $a < b$ and $c = c$ then $a + c < b + c$. Since $\mathbb{R}$ is an ordered field, these laws are true for $\mathbb{R}$. By Transfer, these laws are also true for $\mathbb{R}^*$. By hypothesis, the Trichotomy Law is also true for $\mathbb{R}^*$. Therefore $\mathbb{R}^*$ is an ordered field. $\dashv$

Hereafter, we will usually leave out the stars on the hyperreal order relations $<^*, \le^*$ in a system of formulas.

The next theorem extends the Transfer Axiom to the case where the system of formulas T has more variables than S. It will be used frequently in this monograph, and gives a way to show that a hyperreal number with a certain property exists.

Definition 1.19. *Let T be a system of formulas with variables $x_1, \dots, x_k, \dots, x_n$. A* **partial real solution** *of T is a k-tuple $(c_1, \dots, c_k)$ of real constants which can be extended to a real solution $(c_1, \dots, c_k, \dots, c_n)$ of T. A* **partial hyperreal solution** *is defined similarly.*

Theorem 1.20. *(Partial Solution Theorem) Let S be a system of formulas with the variables $x_1, \dots, x_k$ and T a system of formulas with the variables $x_1, \dots, x_k, \dots, x_n$. The following are equivalent.*
(i) Every real solution of S is a partial real solution of T.
(ii) Every real solution of S is a partial hyperreal solution of T.
(iii) Every hyperreal solution of S is a partial hyperreal solution of T.

Proof. We prove (i) $\Rightarrow$ (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i). To simplify notation let S have the single variable x and T have the two variables x, y.

Assume (i). For each real solution x_0 of S choose a real number $y_0 = f(x_0)$ such that (x_0, y_0) is a real solution of T. Then

(1) Every real solution of S is a solution of "$f(x)$ is defined".

(2) Every real solution of $S \cup \{y = f(x)\}$ is a solution of T.

By Transfer, (1) and (2) also hold for the hyperreal numbers. Let x_1 be a hyperreal solution of S. By (1), $f(x_1)$ is defined; let $y_1 = f(x_1)$. By (2), (x_1, y_1) is a hyperreal solution of T. Thus x_1 is a partial hyperreal solution of T. This shows that (i) implies (iii).

(iii) trivially implies (ii). Assume (ii), and let x_0 be a real solution of S. Suppose x_0 is not a partial real solution of T. Let $T(x_0, y)$ be the system of formulas obtained by replacing the variable x by the constant x_0 in T. Then $T(x_0, y)$ has no real solutions. By Corollary 1.16, $T(x_0, y)$ has no hyperreal solutions. But then x_0 is not a partial hyperreal solution of T, contradicting our assumption (ii). We conclude that x_0 is a partial real solution if T, so (ii) implies (i). $\dashv$

In the nonstandard approach to calculus, the Completeness Property of the real numbers is seldom used directly. It is always possible, and usually easier, to use the Standard Part Principle (Theorem 1.9) instead. This is explained by the following theorem, which shows that in the presence of the other axioms, the Completeness Property can be replaced by the Standard Part Principle and the Archimedean Property.

The set of natural numbers (non-negative integers) is denoted by $\mathbb{N}$. An ordered field F is said to have the **Archimedean Property** if every element of F is less than some natural number. Equivalently, the set $\mathbb{N}$ of natural numbers has no upper bound in F.

Lemma 1.21. *In any ordered field F, the set $\mathbb{N}$ of natural numbers does not have a least upper bound.*

Proof. Suppose x is an upper bound of $\mathbb{N}$. Then for any $y \in \mathbb{N}$ we have $y+1 \in \mathbb{N}$, so $y+1 \le x$ and hence $y \le x-1$. Therefore $x-1$ is also an upper bound of $\mathbb{N}$. By the order laws we have $x-1 < x$, so x cannot be a least upper bound of $\mathbb{N}$. ⊣

Corollary 1.22. *Every complete ordered field has the Archimedean Property.*

Proof. In a complete ordered field, $\mathbb{N}$ cannot have an upper bound because, by the preceding lemma, $\mathbb{N}$ does not have a least upper bound. ⊣

Corollary 1.23. *The ordered field $\mathbb{R}^*$ of hyperreal numbers does not have the Archimedean Property.*

Proof. By Corollary 1.8, $\mathbb{R}^*$ has a positive infinite element H. By definition, H is an upper bound of $\mathbb{R}$ and $\mathbb{N} \subseteq \mathbb{R}$, so H is an upper bound of $\mathbb{N}$. ⊣

Theorem 1.24. *Assume Axioms B, C, D, E, the Standard Part Principle 1.9, and that $\mathbb{R}$ is an ordered field with the Archimedean Property. Then $\mathbb{R}$ has the Completeness Property, so Axiom A holds.*

Proof. First observe that Axiom A was not used in the proof of the Partial Solution Theorem, so this theorem follows from the remaining Axioms B, C, D, and E. Let X be a nonempty subset of $\mathbb{R}$ with an upper bound. Let f be the function

$$f(y) = \begin{cases} 1, & \text{if } y \text{ is an upper bound of } X, \\ 0, & \text{otherwise.} \end{cases}$$

We will find a real number b such that $f(x) = 0$ for all real $x < b$, and $f(y) = 1$ for all real $y > b$. Since X is nonempty and has an upper bound, there are real numbers a, c with

$$a < c, \quad f(a) = 0, \quad f(c) = 1.$$

Let t be any positive real number and consider the points

$$a, a+t, a+2t, \dots, a+nt, \dots.$$

By the Archimedean Principle there is an n such that $(c-a)/t < n$. Then $c < a+nt$. Since $f(c) = 1$, c is an upper bound of X. Then $a+nt$ is an upper bound of X, so $f(a+nt) = 1$. Therefore there is a least positive integer n_0 such that $f(a+n_0 t) = 1$. Hence

$$f(a+n_0 t - t) = 0, \quad f(a+n_0 t) = 1.$$

Then c is an upper bound of X but $a+n_0 t - t$ is not, so

$$a - t \le a + n_0 t - t < c$$

and thus

$$a \le a + n_0 t < c + t.$$

Taking $u = a + n_0 t$, we see that every real solution of

$$(3) \qquad 0 < t$$

is a partial real solution of

$$(4) \qquad f(u - t) = 0, \quad f(u) = 1, \quad a \le u < c + t.$$

Let t_1 be positive infinitesimal. Then t_1 is a hyperreal solution of (3). By the Partial Solution Theorem, t_1 is a partial hyperreal solution of (4). So there is a hyperreal number u_1 with

$$f(u_1 - t_1) = 0, \quad f(u_1) = 1, \quad a \le u_1 \le c + t_1.$$

Then u_1 is finite. By the Standard Part Principle, u_1 has a standard part $b = \mathrm{st}(u_1)$.

We show that for any real $x < b$ and $y > b$, $f(x) = 0$ and $f(y) = 1$. Every real solution of

$$(5) \qquad x < z, \quad f(z) = 0$$

contains a solution of

$$(6) \qquad f(x) = 0.$$

By Corollary 1.14, every hyperreal solution of (5) contains a solution of (6). Since $f(u_1 - t_1) = 0$, $f(x_1) = 0$ for all hyperreal $x_1 < u_1 - t_1$. In particular, if x is real and $x < b = \mathrm{st}(u_1)$, then $x < u_1 - t_1$ so $f(x) = 0$. A similar argument shows that $f(y) = 1$ for all real $y > b$. It follows that b is the least upper bound of X. $\dashv$

1E. Natural Extensions of Sets

Our axioms provide a natural extension f^* of each real function f. We now define the natural extension of a set of reals. Later on we will extend our discussion to relations on the reals.

DEFINITION 1.25. *Let Y be a set of reals, $Y \subseteq \mathbb{R}$, and let C_Y be the characteristic function of Y, defined by*

$$C_Y(x) = \begin{cases} 1 & \text{if } x \in Y, \\ 0 & \text{if } x \notin Y. \end{cases}$$

The **natural extension** *of Y is the set*

$$Y^* = \{x \in \mathbb{R}^* \colon C_Y(x) = 1\}.$$

PROPOSITION 1.26. *For any set Y of reals, the natural extension Y^* is the unique set of hyperreal numbers such that every system of formulas which has Y as its set of real solutions has Y^* as its set of hyperreal solutions.*

PROOF. It is clear that there is at most one such set. Suppose a system of formulas S has Y as its set of real solutions. Then S has the same set of real solutions as the equation $C_Y(x) = 1$. By Corollary 1.15, S has the same set of hyperreal solutions as $C_Y(x) = 1$, which by definition is the set Y^*. ⊣

Given a term τ and a set $Y \subseteq \mathbb{R}$, we will sometimes write $\tau \in Y$ for the formula $C_Y(\tau) = 1$, and $\tau \notin Y$ for the formula $C_Y(\tau) = 0$.

EXAMPLES Since $[a, b] = \{x \in \mathbb{R} \colon a \leq x,\, x \leq b\}$, we have

$$[a, b]^* = \{x \in \mathbb{R}^* \colon a \leq x,\, x \leq b\}.$$

Since $(a, b) = \{x \in \mathbb{R} \colon a < x,\, x < b\}$, we have

$$(a, b)^* = \{x \in \mathbb{R}^* \colon a < x,\, x < b\}.$$

PROPOSITION 1.27. *Let $X, Y \subseteq \mathbb{R}$.*
(i) $X \subseteq X^*$ and $X^* \cap \mathbb{R} = X$.
(ii) The natural extension mapping preserves Boolean operations, i.e.,

$$(X \cap Y)^* = X^* \cap Y^*,$$

$$(X \cup Y)^* = X^* \cup Y^*,$$

$$(X \setminus Y)^* = X^* \setminus Y^*,$$

$$X \subseteq Y \textit{ if and only if } X^* \subseteq Y^*.$$

(iii) For any real function f of one variable,

$$(\operatorname{domain}(f))^* = \operatorname{domain}(f^*),$$

$$(\operatorname{range}(f))^* = \operatorname{range}(f^*).$$

PROOF. We prove (iii). Let X be the domain of f. Then X is the set of all real solutions of the formula

$$f(x) \text{ is defined.}$$

By Proposition 1.26, X^* is the set of all hyperreal solutions of this formula, which is the domain of f^*.

Let Y be the range of f. For each $y \in Y$ choose a real number $x = g(y)$ such that $f(x) = y$. Then Y is the set of all real solutions of the equation

$$f(g(y)) = y, \tag{7}$$

so by Proposition 1.26, Y^* is the set of all hyperreal solutions of (7). It follows that

$$Y^* \subseteq \operatorname{range}(f^*).$$

Moreover, every real solution of the equation $f(x) = y$ contains a solution of (7). By Corollary 1.14, every hyperreal solution of the equation $f(x) = y$ contains a solution of (7). Therefore

$$\operatorname{range}(f^*) \subseteq Y^*.$$

⊣

We will now explain the connection between hyperreal numbers and topological properties of sets of reals. By a **real neighborhood** of a real point $x \in \mathbb{R}$ we mean an open interval of the form

$$(x - r, x + r)$$

where r is a positive real number. x belongs to the **interior** of Y if some neighborhood of x is included in Y. An **open set** is a set which is equal to its interior. x belongs to the **closure** of Y if every neighborhood of x meets Y. A **closed set** is a set Y which is equal to its closure.

Theorem 1.28. *Let $c \in \mathbb{R}$ and $Y \subseteq \mathbb{R}$. Y includes a neighborhood of c (i.e., c is in the interior of Y) if and only if Y^* includes the monad of c.*

Proof. Suppose Y contains a neighborhood $(c - r, c + r)$ of c. Then every real solution of

$$|c - x| < r$$

belongs to Y. By Transfer, every hyperreal solution belongs to Y^*, and thus the monad of c is included in Y^*.

Now suppose Y does not contain a neighborhood of c. Then each real solution of

$$x > 0 \tag{8}$$

is a partial real solution of

$$y \notin Y, \quad |c - y| < x. \tag{9}$$

Let x_1 be a positive infinitesimal. Then x_1 is a hyperreal solution of (8). By the Partial Solution Theorem 1.20, x_1 is a partial hyperreal solution of (9). Thus there is a hyperreal y_1 with

$$y_1 \notin Y^*, \quad |c - y_1| < x_1.$$

Then y_1 belongs to the monad of c but not to Y^*. ⊣

Corollary 1.29. *The closure of a set $Y \subseteq \mathbb{R}$ of reals is equal to the set*

$$\{\mathrm{st}(x) \colon x \text{ is finite and } x \in Y^*\}.$$

Thus Y is closed if and only if $\mathrm{st}(x) \in Y$ for all finite $x \in Y^*$.

Proof. Let $Z = \mathbb{R} \setminus Y$, so $Z^* = \mathbb{R}^* \setminus Y^*$. The following are equivalent:

$c \in$ closure of Y.
$c \notin$ interior of Z.
The monad of c is not included in Z^*.
There is an $x \in Y^*$ such that $x \approx c$.
$c = \mathrm{st}(x)$ for some finite $x \in Y^*$. ⊣

Corollary 1.30. *A real function f is defined at every point of some neighborhood of c if and only if f^* is defined at every point of the monad of c.*

Proof. By Theorems 1.27 (iii) and 1.28. ⊣

A set $Y \subseteq \mathbb{R}$ is **bounded** if Y is included in some closed real interval $[a, b]$.

THEOREM 1.31. *Let $Y \subseteq \mathbb{R}$. Then Y is bounded if and only if every element of Y^* is finite.*

PROOF. If Y is bounded, $Y \subseteq [a, b]$, then every element $y \in Y$ is a solution of

$$a \leq y, \quad y \leq b. \tag{10}$$

By Transfer, every $y_1 \in Y^*$ is a solution of (10), and hence is finite.

Now suppose Y is not bounded. Then either Y has no upper bound or no lower bound, say Y has no upper bound. Thus each real number x is a partial real solution of

$$x < y, \quad y \in Y. \tag{11}$$

Let x_1 be a positive infinite hyperreal number. By the Partial Solution Theorem, x_1 is a partial hyperreal solution of (11), so there is a y_1 such that $x_1 < y_1, y_1 \in Y^*$. Thus y_1 is a positive infinite element of Y^*. ⊣

We now extend our discussion to relations on the reals. The **plane**, or **real plane**, is the set $\mathbb{R}^2$ of ordered pairs of real numbers, and the **hyperreal plane** is the set $(\mathbb{R}^*)^2$ of ordered pairs of hyperreal numbers. In n variables we have the real n-space $\mathbb{R}^n$ and the hyperreal n-space $(\mathbb{R}^*)^n$. By a **real relation** in n variables we mean a subset of the real n-space $\mathbb{R}^n$.

DEFINITION 1.32. *Let Y be a real relation in n variables. The characteristic function of Y is the function C_Y defined by*

$$C_Y(x_1, \dots, x_n) = \begin{cases} 1 & \text{if } (x_1, \dots, x_n) \in Y, \\ 0 & \text{if } (x_1, \dots, x_n) \notin Y. \end{cases}$$

The **natural extension** *of Y is the hyperreal relation*

$$Y^* = \{x \in (\mathbb{R}^*)^n : C_Y(x) = 1\}.$$

Proposition 1.26 has the following analogue for relations.

PROPOSITION 1.33. *Let Y be a real relation in n variables. The natural extension Y^* of Y is the unique subset $Y^* \subseteq (\mathbb{R}^*)^n$ such that every system of formulas which has Y as its set of real solutions has Y^* as its set of hyperreal solutions.*

DEFINITION 1.34. *The* **distance** *between two points*

$$x = (x_1, \dots, x_n), \quad y = (y_1, \dots, y_n)$$

of $\mathbb{R}^n$ or $(\mathbb{R}^*)^n$ is defined by

$$|x - y| = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}.$$

A **real neighborhood** *of a point $x \in \mathbb{R}^n$ is a set of the form*

$$N_r(x) = \{y \in \mathbb{R}^n : |x - y| < r\}$$

where $0 < r \in \mathbb{R}$. *A point* y *is* **infinitely close** *to* x, *in symbols* $y \approx x$, *if* $|x - y|$ *is infinitesimal. The* **monad** *of* x *is defined as the set*

$$\text{monad}(x) = \{y \in (\mathbb{R}^*)^n : x \approx y\}.$$

x *is* **finite** *if each* x_i *is finite.*

With the above definitions, all the results in this section hold for subsets of the real and hyperreal n-spaces.

1F. Appendix. Algebra of the Real Numbers

This appendix is a summary of some basic notions from algebra which leads up to the characterization of the real number system (Theorem 1.38). The details, and the proofs of the theorems stated in this appendix, can be found in most undergraduate texts on modern algebra. Following the normal mathematical practice, we work within Zermelo-Fraenkel set theory.

(Commutative) Ring: A ring is a structure $(R, 0, +, -, \cdot)$ such that $0 \in R$, $+, \cdot$ are binary functions on R, $-$ is a unary function on R, and the following laws hold for all $a, b, c \in R$.

Commutative Laws	$a + b = b + a$,	$a \cdot b = b \cdot a$
Associative Laws	$a + (b + c) = (a + b) + c$,	$a \cdot (b \cdot c) = (a \cdot b) \cdot c$
Identity Law	$a + 0 = a$	
Inverse Law	$a + (-a) = 0$	
Distributive Law	$a \cdot (b + c) = a \cdot b + a \cdot c$	

We write $a - b$ for $a + (-b)$.

Subring: A subring of a ring R is a subset S of R which contains 0 and is closed under $+, -, \cdot$. It follows that S itself is a ring.

Ideal: An ideal in a ring R is a subring I of R such that whenever $a \in I$ and $r \in R$, $a \cdot r \in I$.

Coset: Given a ring R and a subring S, the coset of an element $r \in R$ modulo S is the set

$$\operatorname{coset}(r) = \{r + s \colon s \in S\}.$$

Equivalence Relation: An equivalence relation on a set A is a binary relation $\equiv$ on A such that for all $a, b, c \in A$,

Reflexive Law	$a \equiv a$
Symmetry Law	$a \equiv b$ implies $b \equiv a$
Transitive Law	$a \equiv b$ and $b \equiv c$ implies $a \equiv c$

Proposition 1.35. *If S is a subring of a ring R, then the relation $a - b \in S$ is an equivalence relation on R. Moreover,*

If $a - b \in S$ then $\operatorname{coset}(a) = \operatorname{coset}(b)$,
If $a - b \notin S$ then $\operatorname{coset}(a) \cap \operatorname{coset}(b) = \emptyset$.

Homomorphism: A homomorphism from a ring R into a ring S is a function $h \colon R \to S$ such that for all $a, b \in R$,

$$h(0) = 0, \quad h(a + b) = h(a) + h(b), \quad h(-a) = -h(a), \quad h(a \cdot b) = h(a) \cdot h(b).$$

Isomorphism: An isomorphism from a ring R to a ring S is a homomorphism $h \colon R \to S$ such that h maps R one to one onto S.

Field: A field is a structure $(F, 0, 1, +, -, \cdot, ^{-1})$ such that $(F, 0, +, -, \cdot)$ is a ring, $1 \in F$, $^{-1}$ is a function from $F \setminus \{0\}$ into F, and the following laws hold for all $a \neq 0$ in F:

Nontriviality	$1 \neq 0$
Identity Law	$1 \cdot a = a$
Inverse Law	$a \cdot (a^{-1}) = 1$

Subfield: A subfield of a field F is a subset G of F which contains $0, 1$ and is closed under the functions $+, -, \cdot, ^{-1}$. It follows that G itself is a field.

Field Extension: Given a field G, a field extension of G is a field F such that G is a subfield of F. A proper field extension of G is a field extension F such that and $F \neq G$.

Ordered Field: An ordered field is a field F with a binary relation $<$ such that the following laws hold for all $a, b, c \in F$.

Transitive Law	If $a < b$ and $b < c$ then $a < c$.
Trichotomy Law	Exactly one of the relations $a < b, a = b, b < a$ hold.
Sum Law	If $a < b$ and $c = c$ then $a + c < b + c$.
Product Law	If $a < b$ and $0 < c$ then $a \cdot c < b \cdot c$.

Examples

$\mathbb{Z}$, the ring of integers.
$\mathbb{Q}$, the ordered field of rational numbers.
$\mathbb{R}$, the ordered field of real numbers.
$\mathbb{C}$, the field of complex numbers.
$GF(2)$, the field with exactly two elements $0, 1$, where $1 + 1 = 0$.

$\mathbb{Z}$ is a subring of $\mathbb{Q}$, $\mathbb{Q}$ is a subfield of every ordered field (including $\mathbb{R}$), and $\mathbb{R}$ is a subfield of $\mathbb{C}$. The set of even integers is an ideal in $\mathbb{Z}$. The mapping $h \colon \mathbb{Z} \to GF(2)$ given by

$$h(n) = \begin{cases} 0 & \text{if } n \text{ is even} \\ 1 & \text{if } n \text{ is odd} \end{cases}$$

is a homomorphism of $\mathbb{Z}$ onto $GF(2)$.

As usual, ab means $a \cdot b$, a/b means $a \cdot (b^{-1})$, and $a \leq b$ means that either $a < b$ or $a = b$. The **absolute value** of a is defined by

$$|a| = \begin{cases} a \text{ if } 0 \leq a \\ -a \text{ if } a < 0 \end{cases}$$

Proposition 1.36. *The following algebraic rules hold in every ordered field.*

$a \cdot (-b) = (-a) \cdot b = -(a \cdot b)$

$a \cdot 0 = 0,$	$0 < 1$
$-(-a) = a,$	$(a^{-1})^{-1} = a$ *if* $a \neq 0$
$-(a - b) = b - a,$	$(a/b)^{-1} = b/a$ *if* $a, b \neq 0$
$\lvert -a \rvert = \lvert a \rvert,$	$\lvert a \cdot b \rvert = \lvert a \rvert \cdot \lvert b \rvert$

If $a < b$ *then* $-b < -a$.
If $0 < a < b$ *then* $0 < b^{-1} < a^{-1}$.

Proposition 1.37. *In an ordered field, if* $b, d \neq 0$ *then*

$$\frac{a}{b} + \frac{c}{d} = \frac{a \cdot d + b \cdot c}{b \cdot d}, \qquad \frac{a}{b} \cdot \frac{c}{d} = \frac{a \cdot c}{b \cdot d}$$

Complete Ordered Field: An ordered field F is complete if every nonempty subset $X \subseteq F$ which has an upper bound in F has a least upper bound in F.

Order Isomorphic: Two ordered fields F, G are order isomorphic if there is an isomorphism h from F onto G such that for any $a, b \in F$, $a < b$ if and only if $h(a) < h(b)$.

Theorem 1.38. *There is a complete ordered field, and any two complete ordered fields are order isomorphic.*

This is an important theorem which shows that there is exactly one complete ordered field up to order isomorphism. The complete ordered field is called the field $\mathbb{R}$ of **real numbers**. The theorem has two parts, existence and uniqueness. The uniqueness part, that any two complete ordered fields are order isomorphic, is easy to prove. Given two complete ordered fields F and G, one may assume that F and G have the same subfields of rational numbers. Then the order isomorphism is the mapping h such that $h(x) = y$ if and only if for every rational number q, $q < x$ if and only if $q < y$. It also follows that this isomorphism is unique.

There are several well-known ways to prove the existence part, that there exists a complete ordered field. Each of these proofs shows somewhat more; it gives what is called a **definable complete ordered field**. For beginning calculus students, this is usually done informally in pre-calculus courses, where the real numbers are constructed by taking a positive real number to be a natural number followed by a decimal point and an infinite sequence of decimal digits which does not end in a sequence of 9's. In more advanced courses this is done more carefully in other ways, such as constructing the real numbers as the set of equivalence classes of Cauchy sequences of rationals.

We now turn to the natural numbers. The existence of the set of natural numbers, and the Principle of Induction, are part of the underlying set theory. We identify natural numbers with elements of an ordered field in the usual way.

NATURAL NUMBER: The set $\mathbb{N} = \{0, 1, 2, \dots\}$ of natural numbers in an ordered field F is the smallest subset X of F such that $0 \in X$, and $x \in X$ implies $x + 1 \in X$. Thus $\mathbb{N}$ is the set of all elements of F formed by adding 1 to itself zero or more times.

PRINCIPLE OF INDUCTION: If $0 \in Y$, and $n \in Y$ implies $n + 1 \in Y$, then $\mathbb{N} \subseteq Y$.

COROLLARY 1.39. *Every nonempty subset of* $\mathbb{N}$ *has a least element.*

ARCHIMEDEAN PROPERTY: An ordered field F has the Archimedean Property if every element $x \in F$ is less than some natural number $n \in \mathbb{N}$.

We saw in Section 1D that $\mathbb{R}$ has the Archimedean Property but $\mathbb{R}^*$ does not. The next theorem generalizes these facts.

THEOREM 1.40. *An ordered field G has the Archimedean property if and only if it is order isomorphic to a subfield of* $\mathbb{R}$.

Since we often refer to intervals in the real line, we review the definition and notation here.

INTERVAL: A real interval is a set I of real numbers such that if $a, b \in I$ and $a < c < b$, then $c \in I$.

It follows from the Completeness Property that every real interval is of one of the types below:

Bounded closed intervals

$$[a, b] = \{x \colon a \le x \le b\}.$$

Bounded open intervals

$$(a, b) = \{x \colon a < x < b\}.$$

Bounded half-open intervals

$$[a, b), \qquad (a, b].$$

Unbounded open intervals

$$(a, \infty) = \{x \colon a < x\},$$

$$(-\infty, b) = \{x \colon x < b\},$$

$$(-\infty, \infty) = \mathbb{R}.$$

Unbounded half-open intervals

$$[a, \infty), \qquad (-\infty, b].$$

1G. Building the Hyperreal Numbers

This is an optional section for the reader who wants to see where the hyperreal numbers come from. It will not be needed in the body of the monograph, until Chapter 15 at the end (which is also optional).

The existence and uniqueness theorem for complete ordered fields, Theorem 1.38, has an analogue for the hyperreal number systems. Moreover, as discovered by Kanovei and Shelah [KS 2004], there are definable hyperreal number systems, just as there are definable complete ordered fields. In practice, this does not matter for the calculus course, where one just uses the axioms, but it is important for the foundations of the subject.

In this section we will first build a hyperreal number system the easy way, using what is called an ultrapower. This does not give a definable object, because it depends on an arbitrary choice of an ultrafilter. We will then describe a more elaborate method, the iterated ultrapower, which does give a definable hyperreal number system.

The uniqueness theorem for hyperreal number systems will be postponed until the last chapter, Chapter 15. The reason for this is that to have uniqueness one needs one more axiom in addition to Axioms A–E for the hyperreal numbers, called the Saturation Axiom. Saturation has an appeal similar to completeness, and is important in more advanced applications of hyperreal numbers. But the Saturation Axiom is not needed at the beginning calculus level, and so it was not included in our present list. For this reason we say "a hyperreal number system" rather than "the hyperreal system".

The Ultrapower. We will now build a hyperreal number system as an ultrapower of the real number system. This will prove that there exists a triple $(*, \mathbb{R}, \mathbb{R}^*)$ which satisfies Axioms A—E. We will then be able to conclude that any statement about the real numbers which follows from the axioms is true of the real numbers. The hyperreal numbers can be regarded as a tool which facilitate the study of the real numbers.

Historically, ultrapowers were first applied to the natural numbers by Skolem [Skolem 1934]. Hewitt [Hewitt 1948] studied ultrapowers of the real number field, and the ultrapower was applied to arbitrary structures by Łoś in [Łoś 1955]. Since then ultrapowers have had a variety of applications in several areas of mathematics.

We will use a form of the Axiom of Choice called Zorn's Lemma. A nonempty set X of sets is called a **chain** if for any two sets $x, y \in X$, either $x \subseteq y$ or $y \subseteq x$.

Zorn's Lemma. *Let Y be a nonempty set of sets such that for any chain $X \subseteq Y$, the union of X belongs to Y. Then Y has a maximal element y, that is, a set $y \in Y$ such that no member of Y properly contains y.*

We begin with the definition of an ultrafilter over an infinite set I. We call I the **index set**.

DEFINITION 1.41. *A* **filter** *U over I is a set of subsets of I such that:*

(i) U is closed under supersets; if $X \in U$ and $X \subseteq Y \subseteq I$ then $Y \in U$.
(ii) U is closed under finite intersections; if $X \in U$ and $Y \in U$ then $X \cap Y \in U$.
(iii) $I \in U$ but $\emptyset \notin U$.

An **ultrafilter** *over I is a filter U over I with the additional property that for each $X \subseteq I$, exactly one of the sets $X, I \setminus X$ belongs to U.*
A **free ultrafilter** *is an ultrafilter U such that no finite set belongs to U.*

THEOREM 1.42. *For every infinite set I, there exists a free ultrafilter over I.*

PROOF. The set of all cofinite (complements of finite) subsets of I is a filter over I (called the Frèchet filter). Let A be the set of all filters F over I such that F contains all cofinite subsets of I. Then A is nonempty and A is closed under unions of chains. By Zorn's Lemma, A has a maximal element U (in fact, infinitely many maximal elements). U is a filter and contains no finite set, because U contains all cofinite sets but $\emptyset \notin U$. To show that U an ultrafilter, we consider an arbitrary set $X \subseteq I$ and prove that there is a filter $V \supseteq U$ which contains either X or $I \setminus X$, so by maximality, $X \in U$ or $I \setminus X \in U$.

Case 1: For all $Y \in U$, $X \cap Y$ is infinite. X and each $Y \in U$ belong to the set

$$V = \{Z \subseteq I \colon Z \supseteq X \cap Y \text{ for some } Y \in U\}.$$

V is a filter over I, because V is obviously closed under supersets and finite intersections, and the hypothesis of Case 1 guarantees that each $Z \in V$ is infinite.

Case 2: For some $Y \in U$, $X \cap Y$ is finite. Then for every $W \in U$, $(I \setminus X) \cap W$ is infinite, for otherwise $Y \cap W \in U$ would be finite. Case 1 applies to $I \setminus X$, so the set

$$V = \{Z \subseteq I \colon Z \supseteq (I \setminus X) \cap Y \text{ for some } Y \in U\}$$

is a filter over I such that $V \subseteq U$, $I \setminus X \in V$. $\dashv$

Hereafter we let U be a free ultrafilter over I. Let $\mathbb{R}^I$ be the set of all functions $a \colon I \to \mathbb{R}$. The elements $a \in \mathbb{R}^I$ will be called I-sequences, and we write a_i for the value of a at an element $i \in I$..

DEFINITION 1.43. *Two I-sequences a, b in $\mathbb{R}^I$ are said to be* **U-equivalent***, in symbols $a =_U b$, if*

$$\{i \colon a_i = b_i\} \in U.$$

LEMMA 1.44. *The relation $=_U$ is an equivalence relation on the set $\mathbb{R}^I$.*

Proof. The Reflexive and Symmetric Laws for $=_U$ are obvious. We prove the Transitive Law. Assume $a =_U b$ and $b =_U c$. Let

$$X = \{i \colon a_i = b_i\},\ Y = \{i \colon b_i = c_i\},\ Z = \{i \colon a_i = c_i\}.$$

Then $X \in U$ and $Y \in U$, so $X \cap Y \in U$. But $X \cap Y \subseteq Z$, so $Z \in U$, and hence $a =_U c$. $\dashv$

Our next step is to define the ultrapower $\prod_U \mathbb{R}$, which will be the set $\mathbb{R}^*$ of hyperreal numbers built from U. The idea is to take the set of all U-equivalence classes of I-sequences and modify it by replacing the U-equivalence class of a constant I-sequence by the constant itself. This makes $\prod_U \mathbb{R}$ an extension of $\mathbb{R}$.

Definition 1.45. *Let a be an I-sequence. If a is U-equivalent to a constant I-sequence $\langle r, r, \dots\rangle$ where $r \in \mathbb{R}$, we define $a_U = r$. Otherwise, we define a_U to be the U-equivalence class of a,*

$$a_U = \{b \colon a =_U b\}.$$

The **ultrapower** *of the set $\mathbb{R}$ modulo U is the set*

$$\prod_U \mathbb{R} = \{a_U \colon a \in \mathbb{R}^I\}.$$

Lemma 1.46. *(i) $\mathbb{R} \subseteq \prod_U \mathbb{R}$.*
(ii) $a =_U b$ if and only if $a_U = b_U$.

Proof. (i) We have $\langle r, r, \dots\rangle_U = r \in \mathbb{R}^*$ for each $r \in \mathbb{R}$.
(ii) This follows from the fact that $=_U$ is an equivalence relation. $\dashv$

Definition 1.47. *The natural extension of the order relation $<$ is the relation $<^*$ on $\prod_U \mathbb{R}$ such that for all $a, b \in \mathbb{R}^I$,*

$$a_U <^* b_U \text{ if and only if } \{i \colon a_i < b_i\} \in U.$$

Lemma 1.48. *The relation $<^*$ is well-defined, that is, if $a =_U c$ and $b =_U d$ then*

$$\{i \colon a_i < b_i\} \in U \text{ iff } \{i \colon c_i < d_i\} \in U.$$

Proof. Suppose $\{i \colon a_i < b_i\} \in U$. Then

$$\{i \colon c_i < d_i\} \supseteq \{i \colon a_i < b_i\} \cap \{i \colon a_i = c_i\} \cap \{i \colon b_i = d_i\}.$$

The right side belongs to U, so the left side belongs to U, as required. $\dashv$

Definition 1.49. *An element $x \in \prod_U \mathbb{R}$ is* ***positive infinite*** *if $n <^* x$ for every natural number n.*

Definition 1.50. *A set I is* **countable** *if there is a one to one function a from I onto $\mathbb{N}$.*

Lemma 1.51. *Suppose the index set I is countable. Then there are positive infinite elements in $\prod_U \mathbb{R}$.*

PROOF. Let a be a one to one function from I onto $\mathbb{N}$. Then $a_U \in \prod_U \mathbb{R}$. However, for each $n \in \mathbb{N}$, the set $\{i\colon n < a_i\}$ is cofinite hence belongs to U. Therefore $n <^* a_U$, so a_U is positive infinite. ⊣

DEFINITION 1.52. *We use vector notation for n-tuples in the obvious way. Let f be a real function of n variables. The natural extension of f is the function f^* of n variables on $\prod_U \mathbb{R}$ such that whenever $a_i, \dots, a_n, c \in \mathbb{R}^I$, we have*

$$f^*(\vec{a}_U) = c_U \textit{ if and only if } \{i\colon f(\vec{a})_i = c_i\} \in U.$$

LEMMA 1.53. *For each real function f of n variables, the natural extension f^* is well-defined. That is, whenever $\vec{a} =_U \vec{b}$ and $c =_U d$ we have*

$$\{i\colon f(\vec{a})_i = c_i\} \in U \textit{ if and only if } \{i\colon f(\vec{b})_i = d_i\} \in U.$$

The proof is similar to that of Lemma 1.48.

DEFINITION 1.54. *The hyperreal number system built from the ultrafilter U is the structure $(*, \mathbb{R}, \mathbb{R}^*)$ where $\mathbb{R}^* = \prod_U \mathbb{R}$, $<^*$ is the natural extension of $<$, and f^* is the natural extension of f for each real function f.*

THEOREM 1.55. *For each free ultrafilter U on a countable index set I, the hyperreal number system $(*, \mathbb{R}, \mathbb{R}^*)$ built from U satisfies Axioms A–E.*

PROOF. Axiom A, that $\mathbb{R}$ is a complete ordered field, is satisfied by definition.

Axiom D, the Function Axiom, follows from Lemmas 1.48 and 1.53, which show that $<^*$ and f^* are well-defined.

We now prove the Transfer Axiom E. From the definition of the natural extensions $<^*$ and f^*, a tuple $\vec{x} = \vec{a}_U$ of hyperreal numbers is a solution of an equation or inequality S if and only if

$$\{i\colon \vec{a}_i \text{ is a solution of } S\} \in U.$$

Since $X \cap Y \in U$ if and only if $X \in U$ and $Y \in U$, it follows that this also holds for each finite system of formulas S. Suppose every real solution of a system of formulas S is a solution of T, and let $\vec{x} = \vec{a}_U$ be a hyperreal solution of S. Then

$$\{i\colon \vec{a}_i \text{ is a solution of } S\} \subseteq \{i\colon \vec{a}_i \text{ is a solution of } T\} \in U,$$

so $\vec{x}$ is a solution of T. This proves the Transfer Axiom.

Axiom B says that $\mathbb{R}^*$ is an ordered field extension of $\mathbb{R}$. By definition, the set $\mathbb{R}^*$ is an extension of the set $\mathbb{R}$, and it follows from $I \in U$ that $<^*$ is an extension of $<$ and that for each real function f, f^* is an extension of f. Each of the ordered field axioms except for the Trichotomy Law is a statement saying that every real solution of some system of formulas S is a solution of T, and hence follows from the Transfer Axiom. For the Trichotomy Law, let

$x = a_U$ and $y = b_U$. It is easy to see from the definition of an ultrafilter that exactly one of the sets

$$\{i\colon a_i < b_i\}, \quad \{i\colon a_i = b_i\}, \quad \{i\colon b_i < a_i\}$$

belongs to U. Therefore exactly one of

$$x <^* y, \quad x = y, \quad y <^* x$$

holds, as required.

Finally Axiom C, that $\mathbb{R}^*$ has a positive infinitesimal, follows from Lemma 1.51 and the fact that in a hyperreal ordered field reciprocals of positive infinite elements are positive infinitesimals. $\dashv$

Examples of Infinitesimals. When we build a hyperreal number system as an ultrapower, we can take the index set I to be any countable set. Let us now take I to be the set of natural numbers $\mathbb{N}$. The elements of $\mathbb{R}^* \setminus \mathbb{R}$ are now U-equivalence classes of sequences of reals, and we can give explicit examples of sequences of reals whose equivalence classes are hyperreal numbers with various properties.

Positive infinitesimals:

$$\left\langle 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots, \frac{1}{n+1}, \dots \right\rangle_U$$

$$\left\langle 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots, 2^{-n}, \dots \right\rangle_U$$

Infinite hyperintegers:

$$\langle 1, 2, 3, 4, \dots, n, \dots \rangle_U$$

$$\langle 1, 2, 6, 24, \dots, n!, \dots \rangle_U$$

Elements of the monad of π:

$$\left\langle 3, 3.1, 3.14, \dots, \frac{[10^n \pi]}{10^n} \dots \right\rangle_U$$

$$\left\langle \pi - 1, \pi - \frac{1}{2}, \pi - \frac{1}{3}, \dots, \pi - \frac{1}{n}, \dots, \right\rangle_U$$

The next theorem verifies these assertions.

Theorem 1.56. *Let $\langle a_1, a_2, a_3, \dots \rangle$ be a sequence of real numbers and let $r \in \mathbb{R}$*

(i) $\langle a_1, a_2, a_3, \dots \rangle_U \approx r$ for every free ultrafilter U over $\mathbb{N}$ if and only if $\lim_{n\to\infty} a_n = r$.

(ii) $\langle a_1, a_2, a_3, \dots \rangle_U$ is positive infinite for every free ultrafilter U over $\mathbb{N}$ if and only if $\lim_{n\to\infty} a_n = \infty$.

Proof. We prove (i). First assume that $\lim_{n\to\infty} a_n = r$. Then for each positive real ε, we have $|a_n - r| < \varepsilon$ for all but finitely many $n \in \mathbb{N}$, and hence

$$\{n \in \mathbb{N} \colon |a_n - r| < \varepsilon\} \in U.$$

Then $|a_U -^* r|^* <^* \varepsilon$. Since ε is arbitrary, we have $a_U \approx r$.

Now suppose it is not the case that $\lim_{n\to\infty} a_n = r$. By the ε, δ condition 5.9 there is a real $\varepsilon > 0$ such that the set

$$X = \{n \in \mathbb{N} \colon |a_n - r| \geq \varepsilon\}$$

is infinite. Using the proof of Theorem 1.42 one can show that there is a free ultrafilter U over $\mathbb{N}$ such that $X \in U$. Then $|a_U -^* r|^* \geq^* \varepsilon$, so $a_U \not\approx r$. ⊣

Here are some examples of sequences a such that the behavior of a_U in $\mathbb{R}^*$ depends on the ultrafilter U.

The equivalence class

$$\langle 1, -1, 1, -1, \dots, (-1)^n, \dots \rangle_U$$

is equal to 1 if $\{n \colon n \text{ is even}\} \in U$, and is equal to -1 if $\{n \colon n \text{ is odd}\} \in U$.

The equivalence class

$$\left\langle 1, \frac{1}{2}, 3, 1, \frac{1}{5}, 6, 1, \frac{1}{7}, 8, \dots \right\rangle_U$$

is either one, infinitesimal, or infinite, depending on which congruence class modulo 3 belongs to U.

For each $r \in [-1, 1]$ there is a free ultrafilter U over $\mathbb{N}$ such that

$$\langle \sin 0, \sin 1, \sin 2, \dots, \sin n, \dots \rangle_U \approx r.$$

The following theorem can be used to verify these examples. Its proof is similar to the proof of Theorem 1.56.

Theorem 1.57. *Let $a \colon \mathbb{N} \to \mathbb{R}$ and $r \in \mathbb{R}$.*

(i) $\langle a_0, a_1, a_2, \dots \rangle_U \approx r$ for some free ultrafilter U over $\mathbb{N}$ if and only if a has a subsequence converging to r.

(ii) $\langle a_0, a_1, a_2, \dots \rangle_U$ is positive infinite for some free ultrafilter U over $\mathbb{N}$ if and only if a has a subsequence diverging to ∞.

A Definable Hyperreal Number System. The ultrapower of the real number system produces a hyperreal number system $(*, \mathbb{R}, \mathbb{R}^*)$ which satisfies Axioms A–E, but which depends on a free ultrafilter U over a countable index set I.

We will next show how to modify the ultrapower to get a hyperreal number system which satisfies Axioms A–E and is definable in set theory. Our purpose here is only to explain why there is a definable hyperreal system. The details will not be needed in the rest of this monograph. For this reason, we will skip the proofs of some lemmas along the way.

First, we give some comments about the notion of being definable. In set theory, we say that a set X is **definable** by a first order formula $\theta(v)$ if we can prove that X is the unique set such that $\theta(X)$ holds. For example, the sets $\mathbb{N}$ of natural numbers, $\mathcal{P}(\mathbb{N})$ of sets of natural numbers, and $\mathbb{R}$ of equivalence classes of Cauchy sequences of rationals, are definable. A definable structure, such as an ordered field or a hyperreal number system, can be thought of as a structure that be described explicitly by a formula. By contrast, the ultrapower construction gives us a nonempty class of isomorphic structures, each depending on an ultrafilter U.

We remark that definable sets can have elements which are not definable. In fact, there must be real numbers $r \in \mathbb{R}$ and sets of natural numbers $X \in \mathcal{P}(\mathbb{N})$ which are not definable, because there are only countably many statements in the language of set theory, but the sets $\mathbb{R}$ and $\mathcal{P}(\mathbb{N})$ have uncountably many elements.

It should be no surprise that something similar happens when we build a definable hyperreal number system. The set of all free ultrafilters over the set $\mathbb{N}$ is definable, but some (and possibly every) free ultrafilter over $\mathbb{N}$ is not definable. We are going to build a big but definable hyperreal number system from the set of all free ultrafilters over $\mathbb{N}$. The idea will be to amalgamate all ultrapowers of $\mathbb{R}$ with index set $\mathbb{N}$ together into one large structure.

The starting point is a product operation on finite sequences of ultrafilters, which can be used to amalgamate finitely many ultrapowers.

Let U, V be ultrafilters over index sets I, J. The product $U \otimes V$ is the set

$$U \otimes V = \{Z \subseteq I \times J \colon \{j \colon \{i \colon \langle i, j\rangle \in Z\} \in U\} \in V\}.$$

Warning: in general, $U \otimes V$ will be different from $V \otimes U$. The finite product $U_1 \otimes \cdots \otimes U_n$ is defined inductively by

$$U_1 \otimes \cdots \otimes U_n = (U_1 \otimes \cdots \otimes U_{n-1}) \otimes U_n.$$

Lemma 1.58. *Given free ultrafilters $U_1, \ldots, U_n$ over index sets $I_1, \ldots, I_n$, the product $U_1 \otimes \cdots \otimes U_n$ is a free ultrafilter over $I_1 \times \cdots \times I_n$.*

The next ingredient is a definable function which maps a linearly ordered set onto the set of all free ultrafilters over $\mathbb{N}$. This is the key idea that was introduced in the paper of Kanovei and Shelah [KS 2004].

To get this function we need ordinal numbers in the sense of Von Neumann. These are defined in such a way that each ordinal number is equal to the set of all smaller ordinals numbers. Let $\mathfrak{c}$ be the least ordinal number whose cardinality is the continuum $2^{\aleph_0}$, that is, the least ordinal number which can be mapped onto the set $\mathcal{P}(\mathbb{N})$.

We define A to be the set of all functions $a \colon \mathfrak{c} \to \mathcal{P}(\mathbb{N})$ such that range(a) is a free ultrafilter U_a over $\mathbb{N}$. The set A is nonempty because, by Theorem 1.42, free ultrafilters over $\mathbb{N}$ exist.

For $X, Y \in \mathcal{P}(\mathbb{N})$, we write $X <_p Y$ if and only if $\sum_{n\in X} 3^{-n} < \sum_{n\in Y} 3^{-n}$. We define the lexicographic order $<_A$ on A as follows. For $a, b \in A$, $a <_A b$ if and only if $a(\alpha) <_p b(\alpha)$ where α is the least ordinal such that $a(\alpha) \neq b(\alpha)$.

Lemma 1.59. *$<_p$ is a linear ordering of $\mathcal{P}(\mathbb{N})$, $<_A$ is a linear ordering of A, and $\{U_a : a \in A\}$ is the set of all free ultrafilters over $\mathbb{N}$.*

We now build a whole definable family of hyperreal number systems, one for each nonempty finite subset σ of A. Arrange σ in increasing order,

$$\sigma = \{a_1 <_A \ldots <_A a_n\}.$$

Let

$$U_\sigma = U_{a_1} \otimes \cdots \otimes U_{a_n}$$

be the product of the corresponding ultrafilters, and let $\mathbb{R}^{(\sigma)}$ be the ultrapower of $\mathbb{R}$ modulo U_σ. By Theorem 1.55, the ultrapower yields a hyperreal number system $((\sigma), \mathbb{R}, \mathbb{R}^{(\sigma)})$ with an order relation $<^{(\sigma)}$ and a natural extension $f^{(\sigma)}$ for each real function f. We also put $\mathbb{R}^{(\emptyset)} = \mathbb{R}$ where $\emptyset$ is the empty set.

For each pair of finite subsets $\sigma \subseteq \tau$ of A, there is a natural embedding $h_{\sigma\tau} : \mathbb{R}^{(\sigma)} \to \mathbb{R}^{(\tau)}$. To illustrate, we give the definition of $h_{\sigma\tau}$ in the case that $\sigma = \{a_1 <_A a_3\}$ and $\tau = \{a_1 <_A a_2 <_A a_3\}$. U_σ is an ultrafilter over $\mathbb{N} \times \mathbb{N}$ and U_τ is an ultrafilter over $\mathbb{N} \times \mathbb{N} \times \mathbb{N}$. Given an element $x_{U_\sigma} \in \mathbb{R}^{(\sigma)}$ where $x : \mathbb{N} \times \mathbb{N} \to \mathbb{R}$, $h_{\sigma\tau}(x_{U_\sigma})$ is the element $y_{U_\tau} \in \mathbb{R}^{(\tau)}$ such that

$$y(n_1, n_2, n_3) = x(n_1, n_3).$$

In particular, $h_{\sigma\sigma}$ is the identity map on $\mathbb{R}^{(\sigma)}$.

Lemma 1.60. *Let τ be a finite subset of A.*
(i) If $\sigma \subseteq \rho \subseteq \tau$ then $h_{\sigma\tau}$ is the composition $h_{\sigma\tau}(x) = h_{\rho\tau}(h_{\sigma\rho}(x))$.
(ii) For each $x \in \mathbb{R}^{(\tau)}$, there is a unique smallest subset $\sigma \subseteq \tau$ such that x is in the range of $h_{\sigma\tau}$.
(iii) If $\sigma \subseteq \tau$ then for each system of formulas S and tuple $\vec{x}$ in $\mathbb{R}^{(\sigma)}$, $\vec{x}$ is a solution of S in $\mathbb{R}^{(\sigma)}$ if and only if $h_{\sigma\tau}(\vec{x})$ is a solution of S in $\mathbb{R}^{(\tau)}$.

The proofs of Lemmas 1.58 and 1.60 can be found in Section 6.5 of the book [CK 1990].

With the above lemmas, we can now amalgamate the hyperreal fields $\mathbb{R}^{(\sigma)}$ into one large hyperreal field $\mathbb{R}^\bullet$. The intuitive idea is to identify each element $x \in \mathbb{R}^{(\sigma)}$ with its image $h_{\sigma\tau}(x) \in \mathbb{R}^{(\tau)}$, and then take $\mathbb{R}^\bullet$ to be the union of the sets $\mathbb{R}^{(\sigma)}$. Formally, we can carry out this idea by introducing, for each finite $\sigma \subseteq A$ and $x \in \mathbb{R}^{(\sigma)}$, a new object called the **thread of** x, defined by

$$h_\sigma(x) = \{(\rho, y) : \rho \subseteq \sigma,\ h_{\rho\sigma}(y) = x\} \cup \{(\tau, y) : \sigma \subseteq \tau,\ h_{\sigma\tau}(x) = y\}.$$

One can easily check that

Lemma 1.61. *If τ is a finite subset of A and $\sigma \subseteq \tau$, then h_σ is the composition $h_\sigma(x) = h_\tau(h_{\sigma\tau}(x))$.*

We then define $\mathbb{R}^\bullet$ to be the set of all threads,

$$\mathbb{R}^\bullet = \{h_\sigma(x)\colon \sigma \subseteq A \text{ and } x \in \mathbb{R}^{(\sigma)}\},$$

and define the natural extensions $<^\bullet$ and $f^\bullet$ in the obvious way on $\mathbb{R}^\bullet$. This gives the desired result.

Theorem 1.62. *The hyperreal structure* $(\bullet, \mathbb{R}, \mathbb{R}^\bullet)$ *is definable and satisfies Axioms A–E.*

CHAPTER 2

DIFFERENTIATION

Permanent Assumption *Throughout this monograph it will be understood that $f, g, \ldots$ denote real functions.*

We will use the hyperreal numbers as a tool to define the notions of limit, derivative, continuity, and integral for real functions.

Section 2B contains a rigorous treatment of infinitesimal microscopes and telescopes suggested by Keith Stroyan [Stroyan 1997]. The remaining material in this chapter is also given in *Elementary Calculus* but is repeated here to make this monograph complete.

2A. Derivatives (§2.1, §2.2)

Definition 2.1. *A real number S is said to be the* **slope** *of a real function f at a real point a if*

$$S = st\left(\frac{f(a+\triangle x) - f(a)}{\triangle x}\right)$$

for every nonzero infinitesimal $\triangle x$.

The **derivative** *of a real function f is the real function f' such that:*
$f'(x) =$ slope of f at x if it exists,
$f'(x)$ is undefined otherwise.

We will show in Chapter 5 that this definition is equivalent to the standard definition of derivative. We say that f is **differentiable** at a if the slope of f at a exists. Here are some easy consequences of the definition.

Corollary 2.2. *f is differentiable at a real number a if and only if*
(a) $f(x)$ is defined for all $x \approx a$, and
(b) The quotient $(f(a+\triangle x) - f(a))/\triangle x$ is finite and has the same standard part for all nonzero $\triangle x \approx 0$.

Corollary 2.3. *If f is differentiable at a real point a, then $f(x)$ is defined for all real x in some neighborhood of a.*

Proof. By part (a) above and Corollary 1.30. $\dashv$

To take full advantage of the Leibniz notation we use independent and dependent variables. We first make these notions precise. If we are given a system of formulas which has the same solution set as the simple equation $y = f(x)$, we say that y is a **function of** x, or that y **depends on** x, and we call x the **independent variable** and y the **dependent variable**. There can also be more than one independent variable. When y depends on x, we introduce one new independent variable $\triangle x$ and two new dependent variables $\triangle y$ and dy. $\triangle y$ is called the **increment of** y and its dependence on x and $\triangle x$ is given by the equation

$$\triangle y = f(x + \triangle x) - f(x).$$

Thus when $f'(x)$ exits, its value is

$$f'(x) = st\left(\frac{\triangle y}{\triangle x}\right).$$

We call dy the **differential of** y, and its dependence on x and $\triangle x$ is given by

$$dy = f'(x)\triangle x,$$

with the understanding that dy exists only when $f'(x)$ exists. As usual we write

$$dx = \triangle x,$$

so we have the familiar equations

$$dy = f'(x)\,dx, \quad f'(x) = \frac{dy}{dx}.$$

Geometrically, we define the **tangent line** to the curve $y = f(x)$ at a real point (a, b) on the curve to be the line through (a, b) with slope $f'(a)$. Thus

$\triangle y =$ change in y along curve,

$dy =$ change in y along tangent line.

Usually we will be interested in the case where $\triangle x$ is infinitesimal. Remember that the increment $\triangle y$ and differential dy are dependent variables which depend on x and $\triangle x$. The next theorem shows the relationship between dy and $\triangle y$ when $\triangle x \approx 0$.

DEFINITION 2.4. *Let $\triangle x$ be a nonzero infinitesimal. We say that u and v are* **infinitely close compared to** $\triangle x$*, in symbols*

$$u \approx v \quad \textit{(compared to } \triangle x\textit{)},$$

if

$$\frac{u}{\triangle x} = \frac{v}{\triangle x}.$$

The relation $\approx$ (compared to $\triangle x$) is obviously an equivalence relation on the hyperreal numbers. It is closely related to the classical large and small oh notation. In the present hyperreal setting, for each nonzero infinitesimal $\triangle x$, we can define

$$O(\triangle x) = \{u \colon u/\triangle x \text{ is finite}\},$$

$$o(\triangle x) = \{u \colon u/\triangle x \text{ is infinitesimal}\}.$$

Both $O(\triangle x)$ and $o(\triangle x)$ are ideals in galaxy(0), and

$$o(\triangle x) \subseteq O(\triangle x) \subseteq \text{monad}(0).$$

One can see from the definitions that

$$u \approx v \quad \text{(compared to } \triangle x) \text{ iff } u - v \in o(\triangle x).$$

THEOREM 2.5. *(Increment Theorem) Suppose x is real, $y = f(x)$, $f'(x)$ exists, and $\triangle x$ is a nonzero infinitesimal. Then*

$$\triangle y = f'(x)\triangle x + \varepsilon\triangle x = dy + \varepsilon\triangle x$$

for some infinitesimal ε. In other words,

$$\triangle y \approx dy \text{ (compared to } \triangle x).$$

PROOF. Take

$$\varepsilon - \frac{\triangle y}{\triangle x} - f'(x).$$

Then $\varepsilon \approx 0$. Multiplying by $\triangle x$,

$$\varepsilon\triangle x = \triangle y - f'(x)\triangle x,$$

$$\triangle y = f'(x)\triangle x + \varepsilon\triangle x.$$

$\dashv$

2B. Infinitesimal Microscopes and Infinite Telescopes

In *Elementary Calculus* we frequently used the pictorial devices of infinitesimal microscopes and infinite telescopes to illustrate definitions and theorems about hyperreal numbers. The following precise definition was suggested by Keith Stroyan [Stroyan 1997] and may help the reader to consistently apply the device in new situations.

Given a point $P(a, b)$ in the hyperreal plane and a positive hyperreal number ε, the **ε-disc** around P is defined as the set of all hyperreal points (x, y) at distance at most ε from P, that is,

$$(x - a)^2 + (y - b)^2 \leq \varepsilon^2.$$

DEFINITION 2.6. *Let $P(a, b)$ be a point in the hyperreal plane and δ be a positive infinitesimal. The* **δ-infinitesimal microscope** *aimed at P is the mapping M from the 2δ-disc around P onto the 2-disc around the origin given by the formula*

$$M(a + \delta x, b + \delta y) = (x, y) \text{ where } x^2 + y^2 \leq 4.$$

Thus M maps (a, b) to $(0, 0)$, magnifies distances by $1/\delta$, and preserves directions.

The 2δ-disc around P is called the **field of view** of the microscope. A drawing of a δ-infinitesimal microscope will distinguish two points if and only if the distance between them is not infinitesimal compared to δ. Thus a point (x, y) is infinitely close to (a, b) if and only if it is in the field of view of some infinitesimal microscope aimed at (a, b). As an example we discuss the use of infinitesimal microscopes in giving a picture of the slope of a function.

By the Increment Theorem 2.5, for each real point a we have $f'(a) = S$ if and only if for every nonzero infinitesimal $\triangle x$, the curve at $a + \triangle x$ is infinitely close to the tangent line at $a + \triangle x$ compared to $\triangle x$, that is,

$$f(a + \triangle x) = f(a) + S\triangle x \quad \text{(compared to } \triangle x).$$

It follows that $f'(a) = S$ at a real point a if and only if the curve $y = f(x)$ looks like a straight line with slope S in the field of view of any infinitesimal microscope aimed at the point $(a, f(a))$. This is illustrated in Figure 2a.

One would also like to use infinitesimal microscopes to illustrate the difference between the tangent line and the curve when the slope exists. In *Elementary Calculus* we simply used artistic license and drew a picture like Figure 2b with the curvature exaggerated. A more sophisticated but accurate picture can be drawn by using a more powerful $(\triangle x)^2$-infinitesimal microscope within a $(\triangle x)$-infinitesimal microscope, as in Figure 2c. The curve and tangent line are indistinguishable in the $(\triangle x)$-infinitesimal microscope but can usually be distinguished in the $(\triangle x)^2$-infinitesimal microscope aimed at the hyperreal point $(a + \triangle x, f(a + \triangle x))$.

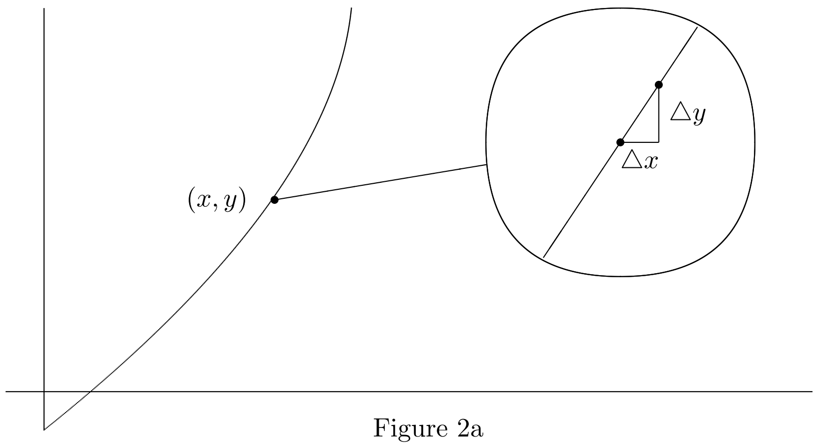

Figure 2a

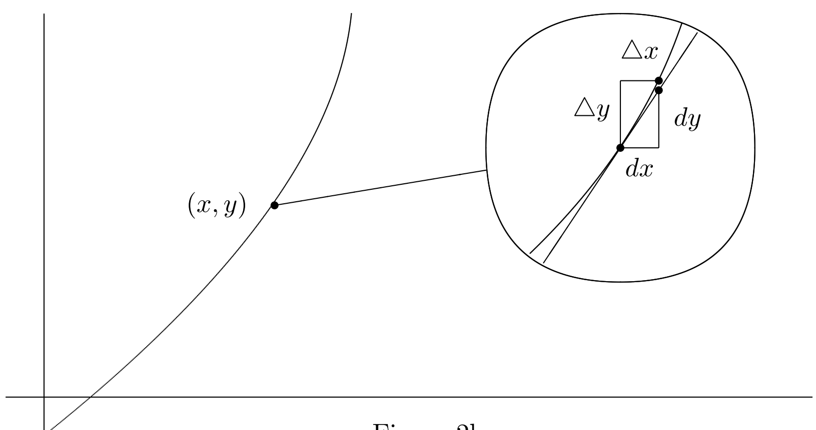

Figure 2b

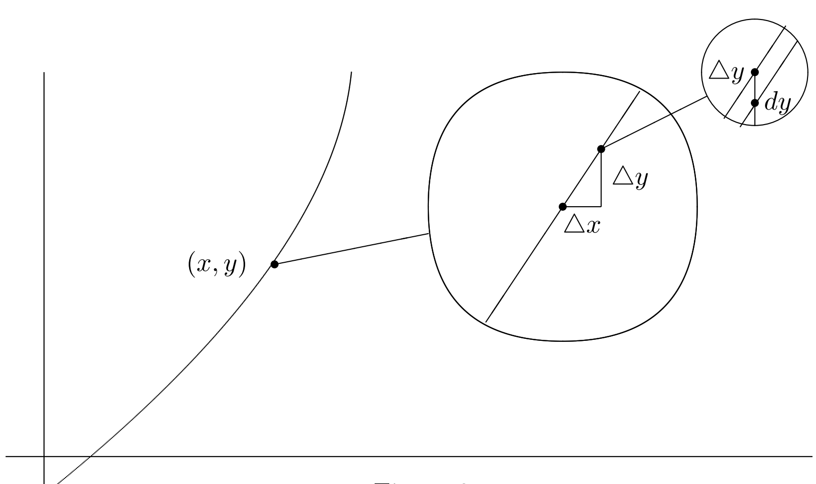

Figure 2c

We now turn to the notion of an infinite telescope.

DEFINITION 2.7. *Let* (a, b) *be a point in the hyperreal plane which is infinitely far from the origin, that is,* $a^2 + b^2$ *is infinite. By the* **infinite telescope** *aimed at* (a, b) *we mean the mapping* T *from the 2-disc around* (a, b) *onto the 2-disc around the origin given by*

$$T(a + x, b + y) = (x, y).$$

Thus T simply translates the 2-disc around (a, b) to the 2-disc around the origin, and preserves distances and directions. In *Elementary Calculus*, limits of the form

$$\lim_{x \to \infty} f(x) = L$$

are illustrated with an infinitesimal microscope within an infinite telescope aimed at a point (H, L) where H is positive infinite. See Chapter 5 of this monograph for the hyperreal definition of infinite limits.

2C. Properties of Derivatives (§2.3, §2.4)

The familiar rules for derivatives can be obtained quite easily from the rules for standard parts in Section 1B.

Given a term τ, we write $d\tau$ for dy, and $d\tau/dx$ for dy/dx, where y is the dependent variable given by the equation $y = \tau$. This often saves space. For example, we can write $d(u + v)$ without introducing a new variable $y = u + v$.

THEOREM 2.8. *Suppose* u *and* v *depend on the independent variable* x*. Then for any real value of* x *where* du/dx *and* dv/dx *exist, we have*

(i) (Sum Rule)

$$\frac{d(u + v)}{dx} = \frac{du}{dx} + \frac{dv}{dx}.$$

(ii) (Constant Rule) For any real number c*,*

$$\frac{d(cu)}{dx} = c\frac{du}{dx}.$$

(iii) (Product Rule)

$$\frac{d(uv)}{dx} = u\frac{dv}{dx} + v\frac{du}{dx}.$$

(iv) (Quotient Rule) If $v \neq 0$*,*

$$\frac{d(u/v)}{dx} = \frac{v(du/dx) - u(dv/dx)}{v^2}.$$

PROOF. For each part we let $\triangle x$ be a nonzero infinitesimal.

Sum Rule: Let $y = u + v$. Then

$$\begin{aligned}
&\triangle y = (u + \triangle u) + (v + \triangle v) - (u + v) = \triangle u + \triangle v,\\
&\frac{\triangle y}{\triangle x} = \frac{\triangle u}{\triangle x} + \frac{\triangle v}{\triangle x},\\
&st\left(\frac{\triangle y}{\triangle x}\right) = st\left(\frac{\triangle u}{\triangle x} + \frac{\triangle v}{\triangle x}\right) = st\left(\frac{\triangle u}{\triangle x}\right) + st\left(\frac{\triangle v}{\triangle x}\right),\\
&\frac{dy}{dx} = \frac{du}{dx} + \frac{dv}{dx}.
\end{aligned}$$

Constant Rule: Let $y = cu$. Then

$$\begin{aligned}
&\triangle y = c(u + \triangle u) - cu = c\triangle u,\\
&\frac{\triangle y}{\triangle x} = c\frac{\triangle u}{\triangle x},\\
&st\left(\frac{\triangle y}{\triangle x}\right) = st\left(c\frac{\triangle u}{\triangle x}\right) = c \cdot st\left(\frac{\triangle u}{\triangle x}\right),\\
&\frac{dy}{dx} = c\frac{du}{dx}.
\end{aligned}$$

Product Rule: Let $y = uv$.

$$\begin{aligned}
&\triangle y = (u + \triangle u)(v + \triangle v) - uv = u\triangle v + v\triangle u + \triangle u\triangle v,\\
&\frac{\triangle y}{\triangle x} = u\frac{\triangle v}{\triangle x} + v\frac{\triangle u}{\triangle x} + \triangle u\frac{\triangle v}{\triangle x},\\
&st\left(\frac{\triangle y}{\triangle x}\right) = st\left(u\frac{\triangle v}{\triangle x} + v\frac{\triangle u}{\triangle x} + \triangle u\frac{\triangle v}{\triangle x}\right),\\
&st\left(\frac{\triangle y}{\triangle x}\right) = u \cdot st\left(u\frac{\triangle v}{\triangle x}\right) + v \cdot st\left(\frac{\triangle u}{\triangle x}\right) + 0 \cdot st\left(\frac{\triangle v}{\triangle x}\right),\\
&\frac{dy}{dx} = u\frac{dv}{dx} + v\frac{du}{dx}.
\end{aligned}$$

Quotient Rule: Let $y = u/v, v \neq 0$.

$$\begin{aligned}
\triangle y &= \frac{u + \triangle u}{v + \triangle v} - \frac{u}{v} \\
&= \frac{(u + \triangle u)v - (v + \triangle v)u}{v(v + \triangle v)} \\
&= \frac{v\triangle u - u\triangle v}{v(v + \triangle v)}, \\
\frac{\triangle y}{\triangle x} &= \frac{v(\triangle u/\triangle x) - u(\triangle v/\triangle x)}{v(v + \triangle v)}, \\
st\left(\frac{\triangle y}{\triangle x}\right) &= \frac{v \cdot \mathrm{st}(\triangle u/\triangle x) - u \cdot \mathrm{st}(\triangle v/\triangle x)}{v^2}, \\
\frac{dy}{dx} &= \frac{v(du/dx) - u(dv/dx)}{v^2}.
\end{aligned}$$

$\dashv$

THEOREM 2.9. *(Power Rule) If x is a positive real number and r is any rational number, then*

$$\frac{d(x^r)}{dx} = rx^{r-1}.$$

PROOF. *Case* 1: r is a positive integer. The proof is an easy induction using the Product Rule.

Case 2: $r = 1/n$ for some positive integer n. Let $y = x^{1/n}$ and let $\triangle x$ be a nonzero infinitesimal. Consider

$$\triangle y = (x + \triangle x)^{1/n} - x^{1/n}.$$

$\triangle y \neq 0$ because $x + \triangle x \neq x$. $\triangle y$ is infinitesimal because

$$\mathrm{st}(\triangle y) = \mathrm{st}((x + \triangle x)^{1/n}) - \mathrm{st}(x^{1/n}) = x^{1/n} - x^{1/n} = 0.$$

Now

$$\begin{aligned}
x &= y^n, \\
\frac{dx}{dy} &= ny^{n-1}, \\
\frac{\triangle x}{\triangle y} &\approx ny^{n-1}.
\end{aligned}$$

Therefore

$$\begin{aligned}
\frac{\triangle y}{\triangle x} &\approx \frac{1}{ny^{n-1}} = \frac{1}{n}x^{(1/n)-1}, \\
\frac{dy}{dx} &= \frac{1}{n}x^{(1/n)-1}.
\end{aligned}$$

Case 3: r is a positive rational. This follows from Cases 1 and 2 using the fact that

$$x^{m/n} = x^{(1/n)m}.$$

Case 4: r is a negative rational. This follows from Case 3 using the Quotient Rule. $\dashv$

If $r = m/n$ where n is odd, the above Power Rule also holds for negative values of x. If $r = m/n$ where n is even, then x^r is undefined in the real number system when x is negative .

2D. Chain Rule (§2.6, §2.7)

The Chain Rule can be proved in a natural way using the Increment Theorem.

THEOREM 2.10. *(Chain Rule) Let f and G be real functions and let g be the composition*

$$g(t) = G(f(t)).$$

For any real value of t where $f'(t)$ and $G'(f(t))$ exist, $g'(t)$ also exists and

$$g'(t) = G'(f(t))f'(t).$$

PROOF. Let $x = f(t), y = g(t) = G(x)$. Let $\triangle t \neq 0$ be infinitesimal. By the Increment Theorem 2.5 for $x = f(t)$, $\triangle x$ is infinitesimal. By the Increment Theorem for $y = G(x)$,

$$\triangle y = G'(x)\triangle x + \varepsilon \triangle x$$

for some infinitesimal ε. Dividing by $\triangle t$ and taking standard parts,

$$\begin{aligned}
&\frac{\triangle y}{\triangle t} = G'(x)\frac{\triangle x}{\triangle t} + \varepsilon\frac{\triangle x}{\triangle t},\\
&\frac{dy}{dt} = G'(x)\frac{dx}{dt} + 0,\\
&g'(t) = G'(f(t))f'(t).
\end{aligned}$$

$\dashv$

The Chain Rule with dependent variables is stated in the following form.

Let $x = f(t), y = g(t) = G(x)$, and suppose $g'(t)$ and $G'(x)$ exist. Then

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt}$$

where dx/dt and dy/dt are computed with t as the independent variable, and dy/dx is computed with x as the independent variable.

As in the standard calculus treatment, the Chain Rule is not trivial because dy has one meaning when x is the independent variable, $dy = G'(x)dx$, and a different meaning when t is the independent variable, $dy = g'(t)dt$.

Higher derivatives are defined in the usual way. Thus

$$f'' \text{ is the derivative of } f',$$

$$f^{(n+1)} \text{ is the derivative of } f^{(n)}.$$

The **n-th differential** of y can be considered separately and is defined by

$$d^n y = f^{(n)}(x)dx^n$$

where dx^n means $(dx)^n$. Notice that when dx is infinitesimal, dx^2 is a much smaller infinitesimal and d^2y is the product of the real number $f''(x)$ and dx^2.

CHAPTER 3

CONTINUOUS FUNCTIONS

In this chapter we go beyond the treatment given in Chapter 3 of *Elementary Calculus*. Several proofs which were omitted or only sketched there are given fully here.

3A. Limits and Continuity (§3.3, §3.4)

We will now use hyperreal numbers to define the notions of a limit and of a continuous function. In Chapter 5 we will show that these definitions are equivalent to the standard definitions.

DEFINITION 3.1. *Let L, c be real numbers. L is the* **limit** *of $f(x)$ as x approaches c, in symbols*

$$L = \lim_{x \to c} f(x),$$

if whenever $x \approx c$ but $x \neq c$, we have $f(x) \approx L$. If there is no such L we say that the limit does not exist.

In *Elementary Calculus* we made the intuitive statement that whenever $\lim_{x \to c} f(x) = L$, we can see the entire part of the hyperreal graph of $f(x)$, where $x \approx c$ but $x \neq c$, in an infinitesimal microscope aimed at $(c.L)$. This is a simplification that does not conform to the precise definition of infinitesimal microscope given in Section 2B, because an ε-infinitesimal microscope only has a field of radius 2ε. A more exact statement is as follows: $\lim_{x \to c} f(x) = L$ if and only if whenever $x \approx c$ but $x \neq c$, the point $(x, f(x))$ belongs to the field of some infinitesimal microscope aimed at (c, L).

Limits can often be formed by computing standard parts. We see from the definition that if $\mathrm{st}(f(x)) = L$ for all x infinitely close but not equal to c, then

$$\lim_{x \to c} f(x) = L.$$

COROLLARY 3.2. *If $\lim_{x \to c} f(x)$ exists then $f(x)$ is defined for all real $x \neq c$ in some neighborhood of c.*

PROOF. Let $Y = \mathrm{domain}(f) \cup \{c\}$. Then $Y^* = \mathrm{domain}(f^*) \cup \{c\}$, by Proposition 1.27. By the definition of limit, $f(x)$ must be defined for all $x \neq c$ in the monad of c, so Y^* contains the monad of c. Then by Theorem 1.28, Y contains a neighborhood of c, so $f(x)$ must be defined for all real $x \neq c$ in that neighborhood. ⊣

The next corollary follows at once from the definitions.

Corollary 3.3. *The slope of f at a is given by the limit*

$$f'(a) = \lim_{\triangle x \to 0} \frac{f(a+\delta x) - f(a)}{\triangle x}.$$

The limit with respect to a set $Y \subseteq \mathbb{R}$ is defined as follows.

Definition 3.4. *Let L and c be real numbers. L is the* **limit** *of $f(x)$ as x approaches c in Y,*

$$L = \lim_{x \to c, x \in Y} f(x),$$

if whenever $x \in Y^*$ and $x \approx c$ but $x \neq c$, we have $f(x) \approx L$.

Important special cases are the one-sided limits, defined by

$$\lim_{x \to c^-} f(x) = \lim_{x \to c, x<c} f(x),$$

$$\lim_{x \to c^+} f(x) = \lim_{x \to c, x>c} f(x).$$

The following result is easy.

Proposition 3.5. $\lim_{x \to c} f(x)$ *exists if and only if both one-sided limits exist and are equal,*

$$\lim_{x \to c^-} f(x) = \lim_{x \to c^+} f(x).$$

The rules for standard parts lead at once to the following rules for limits.

Theorem 3.6. *(Rules for Limits) Suppose the limits*

$$\lim_{x \to c} f(x), \quad \lim_{x \to c} g(x)$$

both exist.

(i) For any constant k, $\lim_{x \to c}(kf(x)) = k \lim_{x \to c} f(x)$.

(ii) $\lim_{x \to c}(f(x) + g(x)) = \lim_{x \to c} f(x) + \lim_{x \to c} g(x)$.

(iii) $\lim_{x \to c} f(x)g(x) = (\lim_{x \to c} f(x))(\lim_{x \to c} g(x))$.

(iv) If $\lim_{x \to c} g(x) \neq 0$, $\lim_{x \to c}(f(x)/g(x)) = (\lim_{x \to c} f(x))/(\lim_{x \to c} g(x))$.

Proof. To illustrate we prove (ii). Let $x \approx c$ but $x \neq c$. Then by Theorem 1.12,

$$\lim_{x \to c}(f(x)+g(x)) = \mathrm{st}(f(x)+g(x)) = \mathrm{st}(f(x))+\mathrm{st}(g(x)) = \lim_{x \to c} f(x)+\lim_{x \to c} g(x).$$

⊣

Definition 3.7. *f is* **continuous** *at a real point c if $f(c)$ is defined and whenever x is infinitely close to c, $f(x)$ is infinitely close to $f(c)$.*

As an immediate consequence of the definitions, we have the usual condition for continuity in terms of limits.

COROLLARY 3.8. *f is continuous at a real point c if and only if $f(c)$ is defined and* $\lim_{f\to c} f(x) = f(c)$.

COROLLARY 3.9. *If f is continuous at c, then $f(x)$ is defined for all real x in some neighborhood of c.*

PROOF. By Corollary 3.2, $f(x)$ is defined for all $x \neq c$ in some neighborhood of c. By definition, $f(x)$ is also defined at $x = c$. ⊣

It follows from Theorem 3.6 that sums, products, and quotients of continuous functions are continuous, provided that the denominator is not 0.

THEOREM 3.10. *If f is differentiable at c then f is continuous at c.*

PROOF. Let f be differentiable at c. Then $f(c)$ is defined. Let $x \approx c$ but $x \neq c$. Then

$$\frac{f(x) - f(c)}{x - c}$$

is finite and $x - c$ is infinitesimal. It follows that $f(x) - f(c)$ is infinitesimal, so $f(x) \approx f(c)$. ⊣

PROPOSITION 3.11. *Compositions of continuous functions are continuous. That is, if f is continuous at c, and G is continuous at $f(c)$, then $g(x) = G(f(x))$ is continuous at c.*

PROOF. Let $x \approx c$. Then $f(x) \approx f(c)$, so

$$g(x) = G(f(x)) \approx G(f(c)) = g(c).$$

⊣

We now define continuity and uniform continuity on a set Y of real numbers.

DEFINITION 3.12. *Let Y be a subset of the domain of f. f is* **continuous** *on Y if whenever $c \in Y$, $x \approx c$, and $x \in Y^*$, we have $f(x) \approx f(c)$.*

f is **uniformly continuous** *on Y if whenever $x, y \in Y^*$ and $x \approx y$, we have $f(x) \approx f(y)$.*

COROLLARY 3.13. *If f is uniformly continuous on Y then f is continuous on Y.*

PROOF. Suppose $c \in Y, x \in Y^*, x \approx c$. By Theorem 1.27, $Y \subseteq Y^*$, so $c \in Y^*$. Then, since f is uniformly continuous on Y, $f(x) \approx f(c)$, and therefore f is continuous on Y. ⊣

A set Y of reals is said to be **compact** if it is closed and bounded. For example, each closed interval $[a, b]$ is compact.

COROLLARY 3.14. *A set Y of reals is compact if and only if for every $y \in Y^*$, y is finite and* $\mathrm{st}(y) \in Y$.

PROOF. By Corollary 1.29 and Theorem 1.31. ⊣

THEOREM 3.15. *Let Y be a compact set of reals. If f is continuous on Y then f is uniformly continuous on Y.*

PROOF. Suppose f is continuous on Y. Let $x, y \in Y^*$ and $x \approx y$. By Corollary 3.14, x is finite and $c \in Y$ where $c = \mathrm{st}(x)$. Since $x \approx y$, $c = \mathrm{st}(y)$. By the continuity of f on Y,

$$f(x) \approx f(c), \quad f(y) \approx f(c).$$

Thus $f(x) \approx f(y)$, and f is uniformly continuous on Y. ⊣

If f is uniformly continuous on a set Y then f is obviously uniformly continuous on any subset of Y. The following theorem allows us to extend the domain of a uniformly continuous function from an interval to the whole real line.

THEOREM 3.16. *(i) Let f be uniformly continuous on an interval I. Then there is a function g which agrees with f on I and is uniformly continuous on the whole real line.*

(ii) Suppose the derivative f' of f is uniformly continuous on an interval I. Then there is a function g which agrees with f on I such that g' is uniformly continuous on the whole real line.

PROOF. We give the proof for the case where I is a half-open interval of the form $[a, b)$. The other cases are similar.

(i) We first show that $\lim_{x \to b^-} f(x)$ exists. Let $x < b$ and $x \approx b$. Assume for the moment that $f(x)$ is infinite. Then every real $u < b$ is a partial hyperreal solution of

$$u < y, \quad y < b, \quad |f(u) - f(y)| \geq 1. \tag{12}$$

Let $u_1 < b, u_1 \approx b$. By the Partial Solution Theorem 1.20 there is a hyperreal y_1 such that (12) holds. But then $u_1 \approx y_1$ but not $f(u_1) \approx f(y_1)$, contradicting the uniform continuity of f. Therefore $f(x)$ could not have been infinite. So $f(x)$ is finite, and has a standard part B. For all $y < b$ with $y \approx x$ we have $f(y) \approx f(x)$ and hence $f(y) \approx B$. This shows that $B = \lim_{x \to b^-} f(x)$. Now let g be the function

$$g(x) = \begin{cases} f(a), & \text{if } x < a \\ f(x), & \text{if } x \in [a, b) \\ B, & \text{if } x \geq b. \end{cases}$$

Then g agrees with f on $[a, b)$ and is uniformly continuous on the whole real line.

(ii) From the proof of (i), the limits

$$B = \lim_{x \to b^-} f(x), \quad C = \lim_{x \to b^-} f'(x)$$

both exist. The function

$$g(x) = \begin{cases} f(a) + f'(a)(x-a) \text{ if } x < a \\ f(x), \text{ if } x \in [a,b) \\ B + C(x-b), \text{ if } x \geq b \end{cases}$$

agrees with f on $[a,b)$ and has a uniformly continuous derivative on the whole real line. $\dashv$

3B. Hyperintegers (§3.8)

Hyperintegers are a basic tool in several areas of the calculus, including integration and infinite series. In this chapter they are used in the proofs of the Intermediate and Extreme Value Theorems.

Recall from Section 1E that $\mathbb{Z}$ denotes the set of integers, and that $x \in \mathbb{Z}$ is the formula $C_{\mathbb{Z}}(x) = 1$, where C_Z is the characteristic function of $\mathbb{Z}$.

Definition 3.17. *The set of integers is denoted by $\mathbb{Z}$. The natural extension $\mathbb{Z}^*$ of $\mathbb{Z}$ is called the set of* **hyperintegers**.

Note that by Theorem 1.27 (i), $\mathbb{Z}^* \cap \mathbb{R} = \mathbb{Z}$, that is, a real number is an integer if and only if it is a hyperinteger.

In *Elementary Calculus* the hyperintegers were defined in a different way using the function $[x] =$ the greatest integer $n \leq x$. We now show that the two definitions are equivalent.

Theorem 3.18. *$\mathbb{Z}^*$ is the set of all hyperreal numbers y such that $y = [x]$ for some hyperreal x.*

Proof. We first note that the set of integers $\mathbb{Z}$ is the set of all real solutions of the equation $y = [y]$. By Proposition 1.26, the natural extension $\mathbb{Z}^*$ is the set of all hyperreal solutions of $y = [y]$. Thus if $y \in \mathbb{Z}^*$ then $y = [y]$, and hence $y = [x]$ for some hyperreal number x.

Now suppose that $y = [x]$ for some hyperreal number x. The equation $[x] = [[x]]$ holds for all real numbers. By Transfer, it holds for all hyperreal numbers. Therefore $y = [x] = [[x]] = [y]$, so $y = [y]$ and hence $y \in \mathbb{Z}^*$. $\dashv$

It follows that for each hyperreal number x, $[x]$ is a hyperinteger. Moreover, since $[x] \leq x < [x] + 1$ for all real x, we see from the Transfer Axiom that $[x] \leq x < [x] + 1$ for all hyperreal x.

Theorem 3.19. *(i) $\mathbb{Z}^*$ is a subring of $\mathbb{R}^*$. That is, sums, differences, and products of hyperintegers are hyperintegers.*

(ii) For each hyperreal number x, $[x]$ is the greatest hyperinteger $\leq x$, and

$$[x] \leq x < [x] + 1.$$

(iii) There are positive infinite and negative infinite hyperintegers.

(iv) Every finite hyperinteger is an integer, that is,

$$\mathbb{Z}^* \cap \text{galaxy}(0) = \mathbb{Z}.$$

Proof. (i) Every real solution of

$$x \in \mathbb{Z}, \quad y \in \mathbb{Z} \tag{13}$$

is a solution of

$$x + y \in \mathbb{Z}. \tag{14}$$

Hence every hyperreal solution of (13) is a solution of (14), so the sum of two hyperintegers is a hyperinteger. The proofs for differences and products are similar.

(ii) For each real number x, $[x]$ is the greatest integer $\leq x$. Then $[x] \leq x$ for all real x, and by Corollary 1.16, $[x] \leq x$ for all hyperreal x. Moreover, every real solution of

$$n \in \mathbb{Z}, \quad n \leq x \tag{15}$$

is a solution of

$$n \leq [x]. \tag{16}$$

By Transfer, every hyperreal solution of (15) is a solution of (16). This shows $[x]$ is the greatest hyperinteger $\leq x$ for every hyperreal x.

Finally, since the system of formulas

$$[x] \leq x < [x] + 1$$

holds for all real x, it holds for all hyperreal x by Corollary 1.16.

(iii) Let H be a positive infinite hyperreal number. Then $K = [H] + 1$ is a hyperinteger which is greater than H and hence is positive infinite.

(iv) Let K be a finite hyperinteger. Then $\text{st}(K)$ is real, so there is an integer n with

$$n \leq \text{st}(K) < n + 1.$$

It follows that

$$0 \leq |n - K| < 1.$$

However, $|n - K|$ is a hyperinteger by (i), so by (ii) we must have $|n - K| = 0$, and thus $n = K \in \mathbb{Z}$. $\dashv$

A frequent construction in the calculus is the partition of a closed interval $[a, b]$ into infinitely many subintervals of equal infinitesimal length. When x and y are hyperreal numbers, we call the set

$$[x, y]^* = \{x \in \mathbb{R}^* \colon x \leq z \leq y\}$$

a **hyperreal closed interval**. If $a \leq x \leq y \leq b$ we call $[x, y]^*$ a **hyperreal subinterval** of $[a, b]^*$. When $0 < \triangle x$ and $\triangle x \approx 0$, we call $[x, x + \triangle x]^*$ an **infinitesimal interval**. Since no ambiguity can arise, we sometimes drop the star on an infinitesimal interval, writing $[x, x + \triangle x]$ for $[x, x + \triangle x]^*$. In *Elementary Calculus* we always wrote $[x, x+\triangle x]$ instead of $[x, x+\triangle x]^*$. Given a hyperinteger $H > 0$, the closed hyperreal interval $[a, b]^*$ may be partitioned into subintervals of length $\delta = (b - a)/H$. The partition points are

$$a, a+\delta, a+2\delta, \dots, a+K\delta, \dots, a+H\delta = b$$

where K runs over the hyperintegers from 0 to H. If H is infinite, each subinterval will have infinitesimal length $\delta = (b - a)/H$, and the partition is called an **infinite partition** of $[a, b]^*$, or sometimes an infinite partition of $[a, b]$. A typical subinterval has the form

$$[a + K\delta, a + (K+1)\delta]^*.$$

Corollary 3.20. *Given a closed real interval $[a, b]$ and a positive hyperinteger H, let $\delta = (b - a)/H$. Then $[a, b]^*$ is the union of the subintervals*

$$[a + K\delta, a + (K+1)\delta)^*$$

where $K \in \mathbb{Z}^*$ and $0 \leq K < H$.

Proof. Let $x \in [a, b]^*$. Let $K = [(x - a)/\delta]$. Then $K \in \mathbb{Z}^*$ and

$$K \leq \frac{x-a}{\delta} < K + 1.$$

It follows that

$$0 \leq K < \frac{b-a}{\delta} = H, \qquad a + K\delta \leq x < a + (K+1)\delta.$$

⊣

3C. Properties of Continuous Functions (§3.5–§3.8)

We use hyperintegers to prove the Intermediate and Extreme Value Theorems.

Theorem 3.21. *(Intermediate Value Theorem) Suppose f is continuous on the closed interval $[a, b]$. Then for every real number D between $f(a)$ and $f(b)$ there is a point $c \in [a, b]$ such that $f(c) = D$.*

Proof. We may assume that $f(a) \leq f(b)$. The result is trivial if $D = f(a)$ or $D = f(b)$, so we assume that $a < b$ and $f(a) < D < f(b)$. Consider a positive integer n and the finite partition

$$a, a+\delta, a+2\delta, \dots, a+n\delta = b$$

where $\delta = (b - a)/n$. The value of f must cross D in one of the subintervals, so there is an integer m such that

$$0 \leq m < n, \quad f(a + m\delta) < D \leq f(a + (m+1)\delta). \tag{17}$$

Thus every real solution of the system of formulas

$$n \in \mathbb{Z}, \quad 0 < n, \quad \delta = (b - a)/n$$

is a partial real solution of (17). Now let n_1 be a positive infinite hyperinteger and let $\delta_1 = (b-a)/n_1$. By the Partial Solution Theorem, there exists m_1 such that (17) holds. Let $c = \mathrm{st}(a + m_1\delta_1)$. We show that $a \le c \le b$ and $f(c) = D$. We have

$$a \le a + m_1\delta_1 \le a + (m_1+1)\delta_1 \le a + n_1\delta_1 = b.$$

Taking standard parts, $a \le c \le b$. Since f is continuous on $[a,b]$,

$$f(c) = \mathrm{st}(f(a + m_1\delta_1)) \le D,$$

$$f(c) = \mathrm{st}(f(a + (m_1+1)\delta_1)) \ge D.$$

Therefore $f(c) = D$. ⊣

DEFINITION 3.22. *A function f is called* **increasing** *if $f(x) < f(y)$ whenever $x < y$ and $x, y \in$* domain(f)*.*

f is called **decreasing** *if $f(x) > f(y)$ whenever $x < y$ and $x, y \in$* domain(f)*.*

Here is a useful consequence about continuous one to one functions.

THEOREM 3.23. *Suppose f is a continuous one to one function whose domain is an interval I. Then f is either increasing or decreasing.*

PROOF. We prove:

(1) Whenever $a < b < c$ in I, either $f(a) < f(b) < f(c)$ or $f(a) > f(b) > f(c)$.

Proof of (1): Suppose $a < b < c$ in I. Since f is one to one, the values $f(a), f(b)$, and $f(c)$ are all different.

We consider two cases. *Case* 1: $f(a) < f(b)$. In this case we must show that $f(b) < f(c)$. We do this by assuming instead that $f(b) > f(c)$ and arriving at a contradiction. Under this assumption, we may pick a value x such that $f(a) < x < f(b)$ and $f(c) < x < f(b)$. By the Intermediate Value Theorem 3.21, there are points $a_1 \in (a,b)$ and $c_1 \in (b,c)$ such that $f(a_1) = x$ and $f(c_1) = x$. But then $a_1 < c_1$ but $f(a_1) = f(c_1)$, contradicting our hypothesis that f is one to one. This shows that $f(a) < f(b) < f(c)$, and proves (1) in Case 1.

Case 2: $f(a) > f(b)$. A similar argument shows that $f(a) > f(b) > f(c)$.

We next prove:

(2) Whenever $a < b < c < d$ in I, either $f(a) < f(b) < f(c) < f(d)$ or $f(a) > f(b) > f(c) > f(d)$.

Proof of (2): By (1), either $f(a) < f(b) < f(c)$ or $f(a) > f(b) > f(c)$. If $f(a) < f(b) < f(c)$, then we cannot have $f(b) > f(c) > f(d)$, so by (1) again, $f(a) < f(b) < f(c) < f(d)$. Similarly, if $f(a) > f(b) > f(c)$, then $f(a) > f(b) > f(c) > f(d)$.

We now use (2) to prove that f is either increasing or decreasing. Pick any two points $x < y$ in I. We again consider two cases.

Case A: $f(x) < f(y)$. In this case we show that f is increasing. To do this, let $u < v$ in I. Arrange the points x, y, u, v in increasing order, giving us four points $a < b < c < d$ in I with $\{x, y, u, v\} \subseteq \{a, b, c, d\}$. Since $f(x) < f(y)$ and $x, y \in \{a, b, c, d\}$, we cannot have $f(a) > f(b) > f(c) > f(d)$. Then by (2), we must have $f(a) < f(b) < f(c) < f(d)$. Since $u, v \in \{a, b, c, d\}$, $f(u) < f(v)$, so f is increasing.

Case B: $f(x) > f(y)$. A similar argument shows that f is decreasing. ⊣

Definition 3.24. *f has a* **maximum** *at c if $f(c) \geq f(x)$ for all x in the domain of f. A* **minimum** *of f is defined analogously.*

Proposition 3.25. *Suppose f has a maximum at c. Then the natural extension f^* also has a maximum at c, that is, $f(c) \geq f(x)$ for all hyperreal x in the domain of f^*.*

Proof. Every real solution of the formula

$$f(x) \text{ is defined}$$

is a solution of $f(c) \geq f(x)$. By Transfer, every hyperreal solution of the first formula is a solution of the second, so f^* has a maximum at c. ⊣

Definition 3.26. *f has a* **local maximum** *at c if c has a real neighborhood $(c-r, c+r)$ such that $f(x)$ is defined and $f(c) \geq f(x)$ for all $x \in (c-r, c+r)$. Local minima are defined in a similar way.*

Theorem 3.27. *f has a local maximum at c if and only if $f(x)$ is defined and $f(c) \geq f(x)$ for all hyperreal $x \approx c$.*

Proof. Suppose f has a local maximum at c. Then $f(c) \geq f(x)$ for all x in some real neighborhood $(c-r, c+r)$ of c. By Proposition 3.25, $f(c) \geq f(x)$ for all hyperreal x such that $c - r < x < c + r$, and therefore $f(c) \geq f(x)$ for all hyperreal $x \approx c$.

Now suppose f does not have a local maximum at c. *Case* 1: There is no real neighborhood of c on which f is defined. By Corollary 1.30, there is a hyperreal number $x \approx c$ at which $f(x)$ is undefined. *Case* 2: f is defined on some real neighborhood $(c - r, c + r)$ of c. Since f does not have a local maximum at c, every real solution of

$$0 < s < r \tag{18}$$

is a partial real solution of

$$c - s < x < c + s, \quad f(c) < f(x). \tag{19}$$

Let s_1 be positive infinitesimal. Then s_1 is a hyperreal solution of (18). By the Partial Solution Theorem , s_1 is a partial hyperreal solution of (19). Hence there exists $x_1 \approx c$ with $f(c) < f(x_1)$. ⊣

Theorem 3.28. *(Extreme Value Theorem) Suppose that the domain of f is a closed interval $[a, b]$ and f is continuous on $[a, b]$. Then f has a maximum and a minimum.*

PROOF. The result is trivial if $a = b$, so we assume $a < b$. Let n be a positive integer and consider the finite partition

$$a, a+\delta, a+2\delta, \dots, a+n\delta = b$$

where $\delta = (b-a)/n$. Let $f(a+m\delta)$ be the greatest of the values

$$f(a), f(a+\delta), \dots, f(a+n\delta),$$

and let g be the function on the set of positive integers such that $m = g(n)$. Then every real solution of

$$n \in \mathbb{Z}, \quad 0 < n, \quad \delta = (b-a)/n, \quad m = g(n) \tag{20}$$

is a solution of

$$a \le a + m\delta \le b. \tag{21}$$

Furthermore, every real solution of (20) plus

$$k \in \mathbb{Z}, \quad 0 \le k \le n \tag{22}$$

is a solution of

$$f(a+m\delta) \ge f(a+k\delta). \tag{23}$$

Let n_1 be a positive infinite hyperinteger and let $\delta_1 = (b-a)/n_1$ and $m_1 = g(n_1)$. We show that f has a maximum at

$$c = \mathrm{st}(a + m_1\delta_1).$$

The triple (n_1, δ_1, m_1) is a hyperreal solution of (20), so by Transfer it is a solution of (21). Thus

$$a \le a + m_1\delta_1 \le b.$$

Taking standard parts,

$$a \le c \le b.$$

Consider any real number $x \in [a,b]$. By Corollary 3.20, x belongs to an infinitesimal subinterval of $[a,b]$ of the form

$$[a + k_1\delta_1, a + (k_1+1)\delta_1]^*$$

where k_1 is a hyperinteger between 0 and n_1. Then

$$x = \mathrm{st}(a + k_1\delta_1).$$

The quadruple $(n_1, \delta_1, m_1, k_1)$ is a hyperreal solution of (22), and by Transfer it is also a solution of (23). Thus

$$f(a + m_1\delta_1) \ge f(a + k_1\delta_1).$$

Since f is continuous on $[a,b]$,

$$f(c) = \mathrm{st}(f(a+m_1\delta_1)) \ge \mathrm{st}(f(a+k_1\delta_1)) = f(x).$$

Thus f has a maximum at c. $\dashv$

Theorem 3.29. *(Critical Point Theorem) Suppose the domain of f is an interval I, f is continuous on I, and f has a maximum or minimum at a point c in I. Then one of the following occurs:*
(i) c is an endpoint of I,
(ii) $f'(c)$ is undefined,
(iii) $f'(c) = 0$.

Proof. Assume neither (i) nor (ii) holds, so c is not an endpoint of I and $f'(c)$ exists. We show that $f'(c) = 0$, so that (iii) holds. Suppose f has a maximum at c. Let $\triangle x > 0$ be infinitesimal. Then

$$f(c + \triangle x) \le f(c), \quad f(c - \triangle x) \le f(c),$$

$$\frac{f(c + \triangle x) - f(c)}{\triangle x} \le 0 \le \frac{f(c - \triangle x) - f(c)}{\triangle x},$$

$$st\left(\frac{f(c + \triangle x) - f(c)}{\triangle x}\right) \le 0 \le st\left(\frac{f(c - \triangle x) - f(c)}{\triangle x}\right),$$

$$f'(c) \le 0 \le f'(c),$$

$$f'(c) = 0.$$

⊣

We call a point c where (i), (ii), or (iii) occurs a **critical point** of f. A critical point which is not an endpoint of I is called an **interior critical point** of f.

Theorem 3.30. *(Mean Value Theorem) Assume that $a < b$ and f is continuous on the closed interval $[a, b]$ and differentiable on (a, b). Then there is a point $c \in (a, b)$ such that the slope of f at c equals the average slope of f,*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

The proof of the Mean Value Theorem from the Extreme Value and Critical Point Theorems is elementary and uses standard methods only, so we will not give it here. Here is a corollary of the Mean Value Theorem which is often used in the calculus.

Corollary 3.31. *(i) If f is continuous on an interval I and $f'(x) = 0$ whenever x is in the interior of I, then $f(x)$ is constant on I.*

(ii) If f is continuous on an interval I and $f'(x) \ge 0$ whenever x is in the interior of I, then $f(x) \le f(y)$ whenever $x \le y$ in I.

(iii) If f is continuous on an interval I and $f'(x) > 0$ whenever x is in the interior of I, then f is increasing in I.

The Intermediate, Extreme, and Mean Value Theorems have the following useful consequences which involve hyperreal numbers. In these theorems we start with a real function f with at least one variable x, but allow the possibility of other variables as well. We state the results for the case that there is one extra variable. Given a function $f(x,s)$ of two variables x, s, for each real constant a we get a function $f(x,a)$ of one variable x. The natural extension $f^*(x,s)$ is a hyperreal function of two variables x and s, and for each hyperreal constant a, $f^*(x,a)$ is a hyperreal function of one variable x.

In each of the following theorems we suppose that $f(x,s)$ is a real function of two variables such that for each real constant a, $f(x,a)$ considered as a function of x is continuous on an interval I.

THEOREM 3.32. *(Hyperreal Intermediate Value Theorem) For each hyperreal constant a and each $x < y$ in I^*, if u is a hyperreal number between $f^*(x,a)$ and $f^*(y,a)$, then there is a hyperreal z such that $x \le z \le y$ and $f^*(z,a) = u$.*

THEOREM 3.33. *(Hyperreal Extreme Value Theorem) For each hyperreal constant a and each $x < y$ in I^*, $f^*(z,a)$ has a maximum and minimum on the hyperreal closed interval $[x,y]^*$. That is, there is a hyperreal number z between x and y such that whenever $x \le u \le y$, $f^*(z,a) \ge f^*(u,a)$.*

THEOREM 3.34. *(Hyperreal Mean Value Theorem) Suppose that for each real a, $f(x,a)$ as a function of x is differentiable on the interior of I, and let $g(x,a) = f'(x,a)$. Then for each hyperreal constant a and each $x < y$ in I^*, there is a hyperreal number z such that*

$$x < z < y \text{ and } g^*(z,a) = \frac{f^*(y,a) - f^*(x,a)}{y - x}.$$

PROOF. To illustrate the method we prove the Hyperreal Extreme Value Theorem. By the real Extreme Value Theorem, for each a, on every closed real subinterval $[x_0, y_0]$ of I, $f(z,a)$ has a maximum at some point $z_0 = g(x_0, y_0, a)$. Thus each real solution of

$$x_0 \in I, \quad y_0 \in I, \quad x_0 \le u_0 \le y_0, \quad a = a \tag{24}$$

is a solution of

$$x_0 \le g(x_0, y_0, a) \le y_0, \quad f(g(x_0, y_0, a), a) \ge f(u_0, a). \tag{25}$$

By Transfer, every hyperreal solution of (24) is also a solution of (25). Therefore for each hyperreal constant a, $f^*(z,a)$ has a maximum in $[x,y]^*$ at $z = g^*(x,y,a)$. ⊣

We conclude this section with two applications of the Hyperreal Mean Value Theorem.

THEOREM 3.35. *(Second Derivative Test)*
(i) If $f'(c) = 0$ and $f''(c) < 0$, then f has a local maximum at c.
(ii) If $f'(c) = 0$ and $f''(c) > 0$, then f has a local minimum at c.

PROOF. We prove (i). Since $f''(c)$ exists, f and f' are defined on some real neighborhood of c. Let $x \approx c$. By Theorem 3.27 it suffices to prove that

$f(c) \geq f(x)$. We assume $f(c) < f(x)$ and arrive at a contradiction. Say $c < x$. By the Hyperreal Mean Value Theorem there is a hyperreal point t such that

$$c < t < x, \quad f'(t) = \frac{f(x) - f(c)}{x - c}.$$

Then $f'(t) > 0$. Since $f'(c) = 0$, we have

$$\frac{f'(t) - f'(c)}{t - c} = \frac{f'(t)}{t - c} > 0.$$

Taking standard parts,

$$f''(c) = st\left(\frac{f'(t) - f'(c)}{t - c}\right) \geq 0.$$

This contradicts the hypothesis $f''(c) < 0$. The case $c > x$ is similar. We conclude that $f(c) \geq f(x)$. ⊣

Definition 3.36. *A real function f is said to be* **uniformly differentiable** *at a real point c if $f'(c)$ exists and whenever $x \approx c$ and $\triangle x$ is nonzero infinitesimal,*

$$f'(c) \approx \frac{f(x + \triangle x) - f(x)}{\triangle x}.$$

Note that uniform differentiability implies differentiability. Uniform differentiability will be useful when we study inverse functions in Chapter 7 and partial derivatives in Chapter 11. The next theorem compares uniform differentiability with continuous differentiability.

Theorem 3.37. *(i) If f is continuously differentiable at c, then f is uniformly differentiable at c.*

(ii) f is uniformly differentiable at every point of an open interval I, then f is continuously differentiable at every point of I.

Proof. (i) Assume f is continuously differentiable at c. Then f' is continuous at c, and by Corollary 3.9, f' is defined and hence f is differentiable at every point of some open neighborhood I of c. Let $x \approx c$ and let $\triangle x$ be nonzero infinitesimal. By Theorem 1.28, $x, x + \triangle x \in I^*$. By the Hyperreal Mean Value Theorem there is a hyperreal number t between x and $x + \triangle x$ such that

$$f'(t) = \frac{f(x + \triangle x) - f(x)}{\triangle x}.$$

Since $t \approx x \approx c$, we have $f'(c) \approx f'(t)$, so f is uniformly differentiable at c.

(ii) Assume f is uniformly differentiable at every point of an open interval I. Every real solution of

$$x \in I, \quad \varepsilon > 0 \tag{26}$$

is a partial hyperreal solution of

$$0 < \triangle x < \varepsilon, \quad \left| f'(x) - \frac{f(x + \triangle x) - f(x)}{\triangle x} \right| < \varepsilon, \tag{27}$$

for we may take $\triangle x$ to be infinitesimal. Let $x_1 \approx c \in I$ and let ε_1 be positive infinitesimal. Since I is open, $x_1 \in I^*$, hence (x_1, ε_1) is a hyperreal solution of (26). By the Partial Solution Theorem, every hyperreal solution of (26) is a partial hyperreal solution of (27), Therefore there is a $\triangle x_1$ such that (27) holds. Then

$$f'(x_1) \approx \frac{f(x_1 + \triangle x_1) - f(x_1)}{\triangle x_1}.$$

But $\triangle x_1 \approx 0$, so by (ii),

$$f'(c) \approx \frac{f(x_1 + \triangle x_1) - f(x_1)}{\triangle x_1}.$$

Therefore $f'(x_1) \approx f'(c)$, and (i) holds. ⊣

The following corollary shows that if f has a continuous derivative then the hyperreal formula for $f'(x)$ holds for finite hyperreal x as well as for real x.

COROLLARY 3.38. *Suppose the derivative of f is continuous on an interval I (not necessarily open), $c \in I, c \approx x \approx x + \triangle x$, $\triangle x \neq 0$, and $x, x + \triangle x \in I^*$. Then*

$$f'(x) \approx \frac{f(x + \triangle x) - f(x)}{\triangle x}.$$

PROOF. By Theorem 3.37, f is uniformly differentiable at c, so

$$f'(x) \approx f'(c) \approx \frac{f(x + \triangle x) - f(x)}{\triangle x}.$$

⊣

Here is an example of a function f which is uniformly differentiable at 0 but not continuously differentiable at 0. Let $f(x)$ be the function with domain $[-1, 1]$ such that $f(0) = 0$, $f(-x) = f(x)$ for each x, $f(1/n) = n^{-2}$ for every positive integer n, and the graph of $f(x)$ is a straight line on each subinterval $[1/(n+1), 1/n]$. Then f is uniformly differentiable at 0. But f is not differentiable at $1/n$ for each positive integer n, so f' is not defined on an open neighborhood of 0 and thus f' cannot be continuous at 0.

This example shows that uniform differentiability of f at c does not imply that f is differentiable on some open neighborhood of c. But it does imply that f is continuous on some open neighborhood of c.

THEOREM 3.39. *If f is uniformly differentiable at c, then f is continuous on some open neighborhood I of c.*

PROOF. Let $L = |f'(c)| + 1$. It suffices to prove that on some real neighborhood $(c - r, c + r)$ of c, we have the Lipschitz condition

$$|f(y) - f(x)| \leq L|y - x| \text{ for all } x, y \in (c - r, c + r). \tag{28}$$

f is differentiable at c, so by Theorem 3.10 and Corollary 3.9, f is defined on some open neighborhood of c. Suppose there is no real number $r > 0$ such that (28) holds. Then every real solution of $r > 0$ is a partial solution of

$$|f(y) - f(x)| > L|y - x|, \quad x, y \in (c - r, c + r). \tag{29}$$

Let r_1 be a positive infinitesimal. By the Partial Solution Theorem, there are hyperreal x_1, y_1 such that (29) holds. But then

$$\left|\frac{f(y_1) - f(x_1)}{y_1 - x_1}\right| > |f'(c)| + 1, \quad x_1 \approx c, \quad y_1 \approx c,$$

contradicting the uniform differentiability of f at c. $\dashv$

CHAPTER 4

INTEGRATION

In this chapter we use hyperreal numbers to develop the Riemann Integral. To keep the theory as elementary as possible, we restrict ourselves to continuous real functions.

PERMANENT ASSUMPTION *We assume throughout this chapter that f and g are real functions which are continuous on an interval I.*

4A. The Definite Integral (§4.1)

Given a positive real function f, consider the region in the plane bounded by the lines $x = a$, $x = b$, $y = 0$, and the curve $y = f(x)$. We call this the **region under the curve** $y = f(x)$ from a to b. The area of this region may be regarded as a real function $A(a, b)$ of two variables. In this and the next section we will define the definite integral

$$\int_b^a f(x)\,dx$$

and show that it is equal to the area of the region under $y = f(x)$ from a to b. Our plan is as follows. First we list some properties which the intuitive concept of area has. Second we prove that the definite integral has these properties. Third we prove that the definite integral is the only function with these properties.

DEFINITION 4.1. *By an* **area function** *for f we mean a real function $A(u, v)$, whose domain is the set of ordered pairs of elements of I, such that*

(i) A has the Addition Property:

$$A(a, c) = A(a, b) + A(b, c) \text{ for all } a, b, c \in I.$$

(ii) A has the Rectangle Property:

$$m(b - a) \leq A(a, b) \leq M(b - a)$$

whenever $a < b$ in I and f has minimum value m and maximum value M on $[a, b]$.

The Rectangle Property states that the area of the region is between the areas of the inscribed and circumscribed rectangles. It follows at once from the Addition Property that

$$A(a,a) = 0, \quad A(b,a) = -A(a,b).$$

Thus to specify an area function we need only specify the values of $A(a,b)$ for $a < b$ in I.

We now introduce the notion of a Riemann sum of a function with respect to a partition of an interval. For simplicity we will consider only partitions of $[a,b]$ in which all subintervals except the last subinterval have the same length.

DEFINITION 4.2. *Let $[a,b]$ be a subinterval of I and let $\triangle x$ be a positive real number. The* **Riemann sum** $\sum_a^b f(x)\triangle x$ *is defined as the sum*

$$\sum_a^b f(x)\triangle x = f(x_0)\triangle x + f(x_1)\triangle x + \cdots + f(x_{n-1})\triangle x + f(x_n)(b - x_n)$$

where n is the largest integer such that $a + n\triangle x < b$, and

$$x_0 = a,\ x_1 = a + \triangle x,\ \ldots,\ x_n = a + n\triangle x.$$

Geometrically, the interval $[a,b]$ is partitioned into subintervals of equal length $\triangle x$, except that if $\triangle x$ does not evenly divide $b - a$ then the last subinterval $[x_n, b]$ is shorter than $\triangle x$. The Riemann sum is equal to the sum of the areal of the vertical strips over each subinterval with height equal to the value of $f(x)$ at the left end of the subinterval.

The Riemann sum $\sum_a^b f(x)\triangle x$ is a real function of the three variables $a, b, \triangle x$. If a and b are held fixed it becomes a function of the single variable $\triangle x$. If we replace the positive real $\triangle x$ in this function by a positive infinitesimal dx, the natural extension gives us the infinite Riemann sum.

DEFINITION 4.3. *Given a continuous real function f on I and a subinterval $[a,b]$ of I, let*

$$S(\triangle x) = \sum_a^b f(x)\triangle x$$

be the finite Riemann sum. The **infinite Riemann sum** *is the natural extension*

$$S^*(dx) = \sum_a^b f(x)\,dx.$$

Since the finite Riemann sum is defined for all real $\triangle x > 0$, the infinite Riemann sum is defined for all hyperreal $dx > 0$. Our plan is to define the integral as the standard part of the infinite Riemann sum. First we must prove that this sum is finite, so its standard part exists.

LEMMA 4.4. *Let $a < b$ in I and let dx be positive infinitesimal. Then the infinite Riemann sum $\sum_a^b f(x)\,dx$ is a finite hyperreal number.*

PROOF. By the Extreme Value Theorem 3.28, f has a minimum value m and a maximum value M on $[a, b]$. For each positive real $\triangle x$ we have

$$\sum_a^b m\triangle x \leq \sum_a^b f(x)\triangle x \leq \sum_a^b M\triangle x,$$

and

$$\sum_a^b m\triangle x = m(b-a), \quad \sum_a^b M\triangle x = M(b-a).$$

Therefore every real solution of $\triangle x > 0$ is a solution of

$$m(b-a) \leq \sum_a^b f(x)\triangle x \leq M(b-a). \tag{30}$$

By Transfer, since $dx > 0$, dx is a hyperreal solution of (30), and therefore $\sum_a^b f(x)\,dx$ is finite. ⊣

DEFINITION 4.5. *Let $a < b$ in I and let dx be positive infinitesimal. The* **definite integral** *of f from a to b with respect to dx is the standard part of the infinite Riemann sum,*

$$\int_a^b f(x)\,dx = st\left(\sum_a^b f(x)\,dx\right).$$

Moreover,

$$\int_a^a f(x) = 0, \quad \int_b^a f(x)\,dx = -\int_a^b f(x)\,dx.$$

For each fixed positive infinitesimal dx, the definite integral $\int_u^w f(x)\,dx$ is a real function of two variables u and w. It does not depend on the dummy variable x. We always use matching symbols for the dummy variable x and the infinitesimal dx. This convention identifies the dummy variable when integrating a function of two or more variables. For example,

$$\int_0^1 x^2 t\,dx = \frac{1}{3}t, \quad \int_0^1 x^2 t\,dt = \frac{1}{2}x^2.$$

We now develop some properties of the definite integral.

THEOREM 4.6. *Let $a < b$ in I, let c be a real constant, and let dx be positive infinitesimal. Then*

(i) $\int_a^b c\,dx = c(b-a)$
(ii) $\int_a^b c\,f(x)\,dx = c\int_a^b f(x)\,dx$
(iii) $\int_a^b (f(x)+g(x))\,dx = \int_a^b f(x)\,dx + \int_a^b g(x)\,dx$
(iv) If $f(x) \leq g(x)$ for all $x \in [a, b]$, then $\int_a^b f(x)\,dx \leq \int_a^b g(x)\,dx$

PROOF. For each case the proof has three steps. First, verify the analogous formula for finite Riemann sums. Second, use the Transfer Axiom to prove the formula for infinite Riemann sums. Third, take standard parts. ⊣

The following theorem shows that the definite integral $\int_a^b f(x)\,dx$ does not depend on the infinitesimal dx.

THEOREM 4.7. *Let $a < b$ in I and let dx and du be positive infinitesimals. Then*

$$\int_a^b f(x)\,dx = \int_a^b f(u)\,du.$$

PROOF. It suffices to prove that for every positive real number r,

$$\int_a^b f(x)\,dx \le \int_a^b f(u)\,du + r.$$

Let $c = r/(b-a)$. We will show that

$$\sum_a^b f(x)\,dx \le \sum_a^b (f(u)+c)\,du, \tag{31}$$

whence by Theorem 4.6,

$$\int_a^b f(x)\,dx \le \int_a^b (f(u)+c)\,du = \int_a^b f(u)\,du + r.$$

In *Elementary Calculus*, Formula (31) was justified intuitively. To give a rigorous proof of (31) we use the Partial Solution Theorem 1.20. Let $\triangle x$ and $\triangle u$ be positive real numbers. If

$$\sum_a^b f(x)\triangle x > \sum_a^b (f(u)+c)\triangle u,$$

there must be a point at which a rectangle in $\sum_a^b f(x)\triangle x$ is above a rectangle in $\sum_a^b (f(u)+c)\triangle u$. So there must be a pair of points x, u in $[a,b]$ such that

$$x - \triangle u \le u \le x + \triangle x, \quad f(x) > f(u) + c.$$

Thus every real solution of

$$\triangle x > 0, \quad \triangle u > 0, \quad \sum_a^b f(x)\triangle x > \sum_a^b (f(u)+c)\triangle u \tag{32}$$

is a partial real solution of

$$a \le x \le b, \quad a \le u \le b, \quad x - \triangle u \le u \le x + \triangle x, \quad f(x) > f(u) + c. \tag{33}$$

Now suppose (31) fails for dx and du, so

$$\sum_a^b f(x)\,dx > \sum_a^b (f(u)+c)\,du.$$

Then (dx, du) is a hyperreal solution of (32). By the Partial Solution Theorem 1.20, there is a hyperreal solution (dx, du, x_1, u_1) of (33). Since dx and du are infinitesimal, (33) implies that $x_1 \approx u_1$ and $f(x_1) \not\approx f(u_1)$, contradicting the continuity of f. We conclude that (31) is true. ⊣

Corollary 4.8.

$$\int_a^b f(x)\,dx = \lim_{\triangle x \to 0^+} \sum_a^b f(x)\triangle x.$$

Proof. By Theorem 4.7 and Definition 4.5, for every positive infinitesimal $dx = \triangle x$ we have

$$\int_a^b f(u)\,du = \int_a^b f(x)\,dx = st\left(\sum_a^b f(x)\triangle x\right).$$

⊣

From now on when we write $\int_a^b f(x)\,dx$, it is to be understood that dx is some positive infinitesimal. By Theorem 4.7, it doesn't matter which one.

Theorem 4.9. *The definite integral* $\int_a^b f(x)\,dx$ *is an area function for* f.

Proof. The Rectangle Property follows at once from Theorem 4.6, which gives

$$m \le f(x) \le M,$$

$$m(b-a) = \int_a^b m\,dx \le \int_a^b f(x)\,dx \le \int_a^b M\,dx = M(b-a).$$

By Theorem 4.7, it suffices to prove the Addition Property for one positive infinitesimal dx. Let $a < b < c$ in I. Let n be a positive integer and let $\triangle x = (b-a)/n$. Then $b = n\triangle x$ is an endpoint of one of the subintervals of length $\triangle x$ in the partition of $[a, c]$, so

$$\sum_a^c f(x)\triangle x = \sum_a^b f(x)\triangle x + \sum_b^c f(x)\triangle x. \tag{34}$$

Thus every real solution of

$$n \in \mathbb{Z}, \quad 0 < n, \quad \triangle x = (b-a)/n \tag{35}$$

is a solution of (34). Now let n_1 be a positive infinite hyperinteger and let $dx = (b-a)/n_1$. (n_1, dx) is a hyperreal solution of (35). By the Transfer Axiom, dx is a solution of (34), so

$$\sum_a^c f(x)\,dx = \sum_a^b f(x)\,dx + \sum_b^c f(x)\,dx.$$

Since dx is positive infinitesimal, we may take standard parts and obtain the Addition Property for f and dx,

$$\int_a^c f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx.$$

$\dashv$

4B. Fundamental Theorem of Calculus (§4.2)

We will prove that the definite integral is the only area function for f, and use that fact to prove the Fundamental Theorem of Calculus. We first introduce lower and upper Riemann sums. For simplicity we consider only the case where $\triangle x$ evenly divides the interval $[a, b]$.

DEFINITION 4.10. *Let $\triangle x = (b - a)/n$ for some positive integer n. The* **lower Riemann sum** *of f is the sum*

$$\sum_a^b m(x, x + \triangle x)\triangle x = m(x_0, x_1)\triangle x + m(x_1, x_2)\triangle x + \cdots + m(x_{n-1}, x_n)\triangle x,$$

and the **upper Riemann sum** *of f is the sum*

$$\sum_a^b M(x, x + \triangle x)\triangle x = M(x_0, x_1)\triangle x + M(x_1, x_2)\triangle x + \cdots + M(x_{n-1}, x_n)\triangle x,$$

where

$$\begin{aligned} m(x, x + \triangle x) &= \text{ minimum value of } f \text{ on } [x, x + \triangle x], \\ M(x, x + \triangle x) &= \text{ maxumum value of } f \text{ on } [x, x + \triangle x], \\ x_k &= a + k\triangle x. \end{aligned}$$

Geometrically, the lower Riemann sum is the sum of the inscribed rectangles, and the upper Riemann sum is the sum of the circumscribed rectangles. For a given real function f continuous on $[a, b]$, the lower and upper Riemann sums

$$\sum_a^b m(x, x + \triangle x)\triangle x, \quad \sum_a^b M(x, x + \triangle x)\triangle x, \quad \triangle x = (b - a)/n$$

are real functions of n, and are defined whenever n is a positive integer. By Transfer, their natural extensions are defined for any positive hyperinteger. Given a positive infinite hyperinteger H, we let $dx = (b - a)/H$ and call the natural extensions

$$\sum_a^b m(x, x + dx)\,dx, \quad \sum_a^b M(x, x + dx)\,dx$$

the **infinite lower and upper Riemann sums** of f with respect to dx.

LEMMA 4.11. *The infinite lower and upper Riemann sums are infinitely close to each other,*

$$\sum_a^b m(x, x+dx)\,dx \approx \sum_a^b M(x, x+dx)\,dx.$$

PROOF. For each positive integer n with $\triangle x = (b-a)/n$, we have

$$\sum_a^b m(x, x+\triangle x)\triangle x \le \sum_a^b M(x, x+\triangle x)\triangle x,$$

so by Transfer,

$$\sum_a^b m(x, x+dx)\,dx \le \sum_a^b M(x, x+dx)\,dx.$$

We must show that for each positive real r,

$$\sum_a^b M(x, x+dx)\,dx < \sum_a^b m(x, x+dx)\,dx + r. \tag{36}$$

The proof is like that of Theorem 4.7. Let $c = r/(b-a)$. Consider a positive integer n with $\triangle x = (b-a)/n$. If

$$\sum_a^b M(x, x+\triangle x)\triangle x \ge \sum_a^b m(x, x+\triangle x)\triangle x + r,$$

then there must be an x such that

$$M(x, x+\triangle x) \ge m(x, x+\triangle x) + c.$$

But $M(x, x+\triangle x) = f(y)$ and $m(x, x+\triangle x) = f(z)$ for some $y, z \in (x, x+\triangle x)$. Thus any real solution of

(37)

$$n \in \mathbb{Z}, \quad 0 < n, \quad \triangle x = (b-a)/n, \quad \sum_a^b M(x, x+\triangle x)\triangle x \ge \sum_a^b m(x, x+\triangle x)\triangle x + r$$

is a partial real solution of

$$a \le y \le b, \quad a \le z \le b, \quad |y - z| \le \triangle x, \quad f(y) \ge f(z) + c. \tag{38}$$

Now let $dx = (b-a)/n_1$ where n_1 is a positive infinite hyperinteger. Suppose (36) fails,

$$\sum_a^b M(x, x+dx)\,dx \ge \sum_a^b m(x, x+dx)\,dx + r.$$

Then (n_1, dx) is a hyperreal solution of (37). By the Partial Solution Theorem, (n_1, dx) is a partial hyperreal solution of (38), so there is a solution (n_1, dx, y_1, z_1) of (38). It follows that

$$y_1 \approx z_1, \quad f(y_1) \not\approx f(z_1),$$

contradicting the continuity of f. We conclude that (36) holds. ⊣

THEOREM 4.12. *The definite integral is the only area function for f.*

PROOF. Let $A(u,v)$ and $B(u,v)$ be area functions for f. Let $a < b$ in I. For each positive integer n with $\triangle x = (b-a)/n$, the Rectangle Property gives

$$m(x_k, x_{k+1})\triangle x \le A(x_k, x_{k+1}) \le M(x_k, x_{k+1})\triangle x, \quad k = 0, 1, \dots, n-1,$$

and by the Addition Property,

$$\sum_a^b m(x, x+\triangle x)\triangle x \le A(a,b) \le \sum_a^b M(x, x+\triangle x)\triangle x.$$

Let n_1 be a positive infinite hyperinteger and $dx = (b-a)/n_1$. By Transfer,

$$\sum_a^b m(x, x+dx)\,dx \le A(a,b) \le \sum_a^b M(x, x+dx)\,dx.$$

Similarly,

$$\sum_a^b m(x, x+dx)\,dx \le B(a,b) \le \sum_a^b M(x, x+dx)\,dx.$$

By Lemma 4.11,

$$A(a,b) \approx B(a,b),$$

and since $A(a,b)$ and $B(a,b)$ are real they must be equal. If $b \le a$ they are still equal because $A(a,a) = B(a,a) = 0$, and

$$A(b,a) = -A(a,b) = -B(a,b) = B(b,a).$$

⊣

DEFINITION 4.13. *Suppose the domain of f is an open interval I. A function F is said to be the* **antiderivative** *of f on I if f is the derivative of F on I.*

THEOREM 4.14. *(Fundamental Theorem of Calculus) Suppose I is an open interval, $a \le b$ in I, and F is the antiderivative of f on I. Then*

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

PROOF. Let $D(a,b) = F(b) - F(a)$. We show that D is an area function for f. Since the definite integral is the only area function for f by Theorem 4.12, it will follow that D is equal to the definite integral of f.

Addition Property:

$$D(a,c) = F(c) - F(a) = (F(b) - F(a)) + (F(c) - F(b)) = D(a,b) + D(b,c).$$

Rectangle Property: By the Mean Value Theorem 3.30 there is a point $c \in (a,b)$ such that

$$f(c) = F'(c) = \frac{F(b) - F(a)}{b - a},$$

$$f(c)(b-a) = F(b) - F(a) = D(a,b).$$

By the Extreme Value Theorem 3.28, f has a minimum value m and a maximum value M on $[a,b]$. Then

$$m \le f(c) \le M,$$
$$m(b-a) \le f(c)(b-a) \le M(b-a),$$
$$m(b-a) \le D(a,b) \le M(b-a).$$

Thus D has the Rectangle Property and hence D is an area function for f. ⊣

Theorem 4.15. *Suppose the domain of f is an open interval I, and F is an antiderivative of f. Then the set of all antiderivatives of f is equal to the set of all functions which differ from F by a constant.*

Proof. For any constant C_0, $F(x) + C_0$ has the same derivative as F and hence is an antiderivative of f. Suppose G is an antiderivative of f, and let $H(x) = F(x) - G(x)$. We prove that H is constant. We have

$$H'(x) = F'(x) - G'(x) = f(x) - f(x) = 0.$$

By the Mean Value Theorem 3.30, whenever $a < b$ in I there is a point c in (a,b) such that

$$0 = H'(c) = \frac{H(b) - H(a)}{b-a}.$$

Therefore $H(b) = H(a)$, and H is a constant function. ⊣

Definition 4.16. *The set of all antiderivatives of f is called the* **indefinite integral** *of f, and is written $\int f(x)\,dx$.*

The set of all functions which differ from F by a constant is denoted by

$$F(x) + C.$$

If F is an antiderivative of f, then Theorem 4.15 states that

$$\int f(x)\,dx = F(x) + C.$$

The techniques for evaluating integrals in this treatment are exactly the same as in the standard calculus treatment. As usual, one can evaluate a definite integral by finding an antiderivative and using the Fundamental Theorem of Calculus. Notice that we have not yet proved that every continuous real function f has an antiderivative. This will be done in the next section.

4C. Second Fundamental Theorem of Calculus (§4.2)

In this section let I be an arbitrary interval and suppose f has domain I and is continuous on I. We may regard the definite integral $\int_u^w f(t)\,dt$ as a real function of two variables defined for all $u, w \in I$. By the Function Axiom, $\int_u^w f(t)\,dt$ is also defined when u and w are hyperreal numbers in I^*.

THEOREM 4.17. *(Second Fundamental Theorem of Calculus) Let $a \in I$ and define $F(x)$ for all $x \in I$ by*

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

(i) F is continuous on I.
(ii) F is an antiderivative of f on the interior of I.

PROOF. (i) Let $c \in I$ and let $\triangle x$ be an infinitesimal such that $c + \triangle x \in I^*$. We must show that

$$F(c + \triangle x) \approx F(c). \tag{39}$$

Suppose $\triangle x > 0$, and let b be a point in I such that $c < c + \triangle x < b$. By the Extreme Value Theorem 3.28, f has a minimum value m and a maximum value M in $[c, b]$. For each positive real number $u < b - c$ we have

$$F(c + u) - F(c) = \int_c^{c+u} f(t)\,dt,$$

$$mu \le \int_c^{c+u} f(t)\,dt \le Mu,$$

and hence

$$mu \le F(c + u) - F(c) \le Mu.$$

By Transfer we have

$$m\triangle x \le F(c + \triangle x) - F(c) \le M\triangle x,$$

and (39) follows. The case $\triangle x < 0$ is similar.

(ii) Let c be an interior point of I and let $\triangle x$ be nonzero infinitesimal. Again suppose $\triangle x > 0$ and let $b \in I$, $\quad c < c + \triangle x < b$. Consider a real number $u \in (0, b - c)$. By the Extreme Value Theorem 3.28, in the interval $[c, c + u]$ the function f has a minimum at some point y and a maximum at some point z. Then

$$F(c + u) - F(c) = \int_c^{c+u} f(t)\,dt,$$

$$f(y)u \le \int_c^{c+u} f(t)\,dt \le f(z)u.$$

Then any real solution of

$$0 < u < b - c \tag{40}$$

is a partial real solution of

$$c \le y \le c + u, \quad c \le z \le c + u, \quad f(y)u \le F(c + u) - F(c) \le f(z)u. \tag{41}$$

$u = \triangle x$ is a hyperreal solution of (40). By the Partial Solution Theorem, it is a partial hyperreal solution of (41). Thus there are hyperreal numbers y_1, z_1 such that

$$y_1 \approx c, \quad z_1 \approx c, \quad f(y_1)\triangle x \le F(c + \triangle x) - F(c) \le f(z_1)\triangle x.$$

Then

$$f(y_1) \le \frac{F(c + \triangle x) - F(c)}{\triangle x} \le f(z_1).$$

Using the continuity of f,

$$f(c) = \mathrm{st}(f(y_1)) \le st\left(\frac{F(c + \triangle x) - F(c)}{\triangle x}\right) \le \mathrm{st}(f(z_1)) = f(c),$$

whence

$$f(c) = st\left(\frac{F(c + \triangle x) - F(c)}{\triangle x}\right).$$

A similar argument works when $\triangle x < 0$. It follows that

$$f(c) = F'(c).$$

$\dashv$

We now use the Second Fundamental Theorem to give a short alternate proof of the Fundamental Theorem of Calculus in Section 4B. This proof does not depend on the fact that the area function for f is unique.

ALTERNATE PROOF OF THE FUNDAMENTAL THEOREM OF CALCULUS Let G be an antiderivative of f. Since any two antiderivatives differ by a constant,

$$G(x) = \int_a^x f(t)\,dt + C_0$$

for some constant C_0. Then

$$G(b) - G(a) = \left(\int_a^b f(t)\,dt + C_0\right) - \left(\int_a^a f(t)\,dt + C_0\right) = \int_a^b f(t)\,dt.$$

CHAPTER 5

LIMITS

5A. ε, δ Conditions for Limits (§5.8, §5.1)

We show that the infinitesimal definitions of limit, continuity, uniform continuity, differentiability, and uniform differentiability are equivalent to the standard ε, δ definitions. These equivalence theorems will be useful later on. In this section, f is a real function. We start with the equivalence theorem for finite limits.

THEOREM 5.1. *Let c, L be real numbers. The following are equivalent.*

(i) $\lim_{x \to c} f(x) = L$. That is, whenever $x \approx c$, we have $f(x) \approx L$.

(ii) There exists a hyperreal $\delta > 0$ such that whenever $0 < |x - c| < \delta$, we have $f(x) \approx L$.

(iii) The ε, δ condition: For every real $\varepsilon > 0$ there is a real $\delta > 0$ such that whenever x is real and $0 < |x - c| < \delta$, we have $|f(x) - L| < \varepsilon$.

PROOF. (i) obviously implies (ii), with δ being any positive infinitesimal. To prove (ii) implies (iii), assume (iii) fails for some real $\varepsilon > 0$. Then every real $\delta > 0$ is a partial real solution of

$$0 < |x - c| < \delta, \qquad |f(x) - L| \geq \varepsilon. \tag{42}$$

Let $\delta_1 > 0$ be hyperreal. By the Partial Solution Theorem 1.20, there is a hyperreal x_1 such that (42) holds, and therefore

$$0 < |x_1 - c| < \delta_1, \qquad f(x_1) \not\approx L.$$

Thus if (iii) fails then (ii) fails, so (ii) implies (iii).

Assume (iii). Let $x_1 \approx c$. Let ε be any positive real number, and let δ be the corresponding positive real number in the ε, δ condition. Then every real solution of

$$0 < |x - c| < \delta$$

is a solution of

$$|f(x) - L| < \varepsilon.$$

We have $0 < |x_1 - c| < \delta$, and then by Transfer, $|f(x_1) - L| < \varepsilon$. Since this holds for all positive real ε, $f(x_1) \approx L$. ⊣

Condition (ii) in the above theorem is sometimes easier to verify than (i). Here is the equivalence theorem for continuity.

COROLLARY 5.2. *Let c a real number. The following are equivalent.*
(i) f is continuous at c. That is, whenever $x \approx c$, we have $f(x) \approx f(c)$.
(ii) There is a hyperreal $\delta > 0$ such that whenever $|x - c| < \delta$, we have $f(x) \approx f(c)$.
(iii) The ε, δ condition: For every real $\varepsilon > 0$ there is a real $\delta > 0$ such that for all real $x \in (c - \delta, c + \delta)$, we have $|f(x) - f(c)| < \varepsilon$.

We next give the equivalence theorem for differentiability.

COROLLARY 5.3. *Let c, S be real numbers. The following are equivalent.*
(i) $f'(c) = S$. That is, whenever $\triangle x \approx 0$ but $\triangle x \neq 0$, we have

$$\frac{f(c + \triangle x) - f(c)}{\triangle x} \approx S.$$

(ii) There is a hyperreal $\delta > 0$ such that whenever $0 < |\triangle x| < \delta$, we have

$$\frac{f(c + \triangle x) - f(c)}{\triangle x} \approx S.$$

(iii) The ε, δ condition: For every real $\varepsilon > 0$ there is a real $\delta > 0$ such that whenever $\triangle x$ is real and $0 < |\triangle x| < \delta$, we have

$$\left| \frac{f(c + \triangle x) - f(c)}{\triangle x} - S \right| < \varepsilon.$$

Here is the equivalence theorem for uniform differentiability.

THEOREM 5.4. *Let c, S be real numbers. The following are equivalent.*
(i) $f'(c) = S$ and f is uniformly differentiable at c.
(ii) There is a hyperreal $\delta > 0$ such that whenever $0 < |\triangle x| < \delta$ and $|x - c| < \delta$, we have

$$\frac{f(x + \triangle x) - f(x)}{\triangle x} \approx S.$$

(iii) The ε, δ condition: For every real $\varepsilon > 0$ there is a real $\delta > 0$ such that for all real $\triangle x, x$ with $0 < |\triangle x| < \delta$ and $|x - c| < \delta$, we have

$$\left| \frac{f(x + \triangle x) - f(x)}{\triangle x} - S \right| < \varepsilon.$$

PROOF. The proof is similar to Theorem 5.1. (i) clearly implies (ii) where δ is any positive infinitesimal. Assume (iii) fails for some real $\varepsilon > 0$. Then any real $\delta > 0$ is a partial real solution of

$$0 < |\triangle x| < \delta, \quad |x - c| < \delta, \quad \left| \frac{f(x + \triangle x) - f(x)}{\triangle x} - S \right| \geq \varepsilon. \tag{43}$$

Let δ_1 be hyperreal. By the Partial Solution Theorem there are hyperreal $\triangle x_1$ and x_1 such that (43) holds, and therefore (ii) fails. Thus (ii) implies (iii).

Assume (iii), let $\triangle x_1 \approx 0$, $\triangle x_1 \neq 0$, and $x_1 \approx c$. Let $\varepsilon > 0$ be real and take the corresponding real $\delta > 0$. Then every real solution of

$$|x - c| < \delta, \quad 0 < |\triangle x| < \delta$$

is a solution of

$$\left|\frac{f(x + \triangle x) - f(x)}{\triangle x} - S\right| < \varepsilon.$$

By Transfer, this also holds for $\triangle x_1$ and x_1, and therefore (i) holds. ⊣

We now state, without proof, versions of Theorem 5.1 and Corollary 5.2 restricted to a set $Y \subseteq \mathbb{R}$.

THEOREM 5.5. *Let $Y \subseteq \mathbb{R}$ and $c \in Y$, and let L be real. The following are equivalent.*

(i) $\lim_{x \to c, x \in Y} f(x) = L$. That is, whenever $x \in Y^*, x \approx c$, and $x \neq c$, we have $f(x) \approx L$.

(ii) There is a hyperreal $\delta > 0$ such that whenever $x \in Y^*$ and $0 < |x - c| < \delta$, we have $f(x) \approx L$.

(iii) The ε, δ condition: For every real $\varepsilon > 0$ there is a real $\delta > 0$ such that whenever $x \in Y$ and $0 < |x - c| < \delta$, we have $|f(x) - L| < \varepsilon$.

COROLLARY 5.6. *Let $Y \subseteq \mathbb{R}$. The following are equivalent.*

(i) f is continuous on Y. That is, whenever $c \in Y, x \in Y^*, x \approx c$, and $x \neq c$, we have $f(x) \approx f(c)$.

(ii) For each $c \in Y$ there exists a hyperreal $\delta > 0$ such that whenever $x \in Y^*$ and $0 < |x - c| < \delta$, we have $f(x) \approx f(c)$.

(iii) The ε, δ condition: For every real $\varepsilon > 0$ and $c \in Y$ there is a real $\delta > 0$ such that whenever $x \in Y$ and $0 < |x - c| < \delta$, we have $|f(x) - f(c)| < \varepsilon$.

Our next result is an equivalence theorem for uniform continuity.

THEOREM 5.7. *Let $Y \subseteq \mathbb{R}$. The following are equivalent.*

(i) f is uniformly continuous on Y. That is, whenever $x, y \in Y^*$ and $x \approx y$, we have $f(x) \approx f(y)$.

(ii) There is a hyperreal $\delta > 0$ such that whenever $x, y \in Y^*$ and $|x - y| < \delta$, we have $f(x) \approx f(y)$.

(iii) The ε, δ condition: For every real $\varepsilon > 0$ there is a real $\delta > 0$ such that whenever $x, y \in Y$ and $0 < |x - y| < \delta$, we have $|f(x) - f(y)| < \varepsilon$.

PROOF. (i) trivially implies (ii). Assume (iii) fails for some real $\varepsilon > 0$. Then every real $\delta > 0$ is a partial real solution of

$$x \in Y, \quad y \in Y, \quad |x - y| < \delta, \quad |f(x) - f(y)| \geq \varepsilon. \tag{44}$$

Hence every hyperreal $\delta_1 > 0$ is a partial hyperreal solution of (44). So there exist $x_1, y_1 \in Y^*$ such that $|x_1 - y_1| < \delta_1$ but $f(x_1) \not\approx f(y_1)$. This shows that the failure of (iii) implies the failure of (ii), so (ii) implies (iii).

Assume (iii). Let $x_1, y_1 \in Y^*$ and $x_1 \approx y_1$. Let ε be any positive real and let $\delta > 0$ be the corresponding number in the ε, δ condition. Every real solution of

$$x \in Y, \quad y \in Y, \quad |x - y| < \delta \tag{45}$$

is a solution of

$$|f(x) - f(y)| < \varepsilon. \tag{46}$$

(x_1, y_1) is a hyperreal solution of (45). By Transfer, (x_1, y_1) is a hyperreal solution of (46). Since this holds for all real $\varepsilon > 0$, $f(x_1) \approx f(y_1)$. ⊣

We conclude this section with a discussion of infinite limits.

DEFINITION 5.8. *Let c and L be real numbers.*
$\lim_{x\to\infty} f(x) = L$ *if $f(H) \approx L$ for every positive infinite H.*
$\lim_{x\to c} f(x) = \infty$ *if $f(x)$ is positive infinite whenever $x \approx c$ but $x \neq c$.*

We state equivalence theorems for these limits without proof.

THEOREM 5.9. *The following are equivalent, where L is real.*
(i) $\lim_{x\to\infty} f(x) = L$.
(ii) There is a hyperreal number K such that whenever $H > K$, we have $f(H) \approx L$.
(iii) The ε, M condition: For every real $\varepsilon > 0$ there is a real number M such that whenever x is real and $x > M$, we have $|f(x) - L| < \varepsilon$.

THEOREM 5.10. *The following are equivalent, where c is real.*
(i) $\lim_{x\to c} f(x) = \infty$.
(ii) There is a hyperreal $\delta > 0$ such that whenever $0 < |x - c| < \delta$, $f(x)$ is positive infinite.
(iii) The M, δ condition: For every real number M there is a real $\delta > 0$ such that whenever x is real and $0 < |x - c| < \delta$, we have $f(x) > M$.

Limits such as $\lim_{x\to\infty} f(x) = \infty$, negative infinite limits, and infinite limits restricted to a set $Y \subseteq \mathbb{R}$, are defined analogously and have similar equivalence theorems.

5B. L'Hospital's Rule (§5.2)

In this section f and g are real functions. The proof of l'Hospital's Rule uses the Generalized Mean Value Theorem, whose proof is elementary and can be found in *Elementary Calculus*.

THEOREM 5.11. *(Generalized Mean Value Theorem) Suppose f and g are continuous on the closed interval $[a, b]$ and differentiable on the open interval*

(a, b). Assume further that $g'(x) \neq 0$ for $x \in (a, b)$. Then there is a point $t \in (a, b)$ such that

$$\frac{f'(t)}{g'(t)} = \frac{f(b) - f(a)}{g(b) - g(a)}.$$

Theorem 5.12. *(L'Hospital's Rule for $0/0$). Suppose that for all x in some real open interval (c, b), $f'(x)$ and $g'(x)$ exist and $g'(x) \neq 0$. Assume that*

$$\lim_{x \to c^+} f(x) = 0, \qquad \lim_{x \to c^+} g(x) = 0.$$

If $\lim_{x \to c^+}(f'(x)/g'(x))$ exists, then

$$\lim_{x \to c^+} \frac{f(x)}{g(x)} = \lim_{x \to c^+} \frac{f'(x)}{g'(x)}.$$

Proof. Let $\lim_{x \to c^+}(f'(x)/g'(x)) = L$. We may set $f(c) = 0$ and $g(c) = 0$, so f and g are continuous on $[c, b)$. By the Generalized Mean Value Theorem, every real solution of

$$c < x < b$$

is a partial real solution of

$$c < t < x, \qquad \frac{f'(t)}{g'(t)} = \frac{f(x) - f(c)}{g(x) - g(c)},$$

which simplifies to

$$c < t < x, \qquad \frac{f'(t)}{g'(t)} = \frac{f(x)}{g(x)}. \tag{47}$$

Now let $x_1 \approx c$ and $x_1 > c$. By the Partial Solution Theorem there exists t_1 such that (47) holds. Thus

$$t_1 \neq c, \qquad t_1 \approx c, \qquad \frac{f'(t_1)}{g'(t_1)} = \frac{f(x_1)}{g(x_1)}.$$

Since $L = \lim_{t \to c^+}(f'(t)/g'(t))$, we have

$$L \approx \frac{f'(t_1)}{g'(t_1)} = \frac{f(x_1)}{g(x_1)}, \qquad L = \lim_{x \to c^+} \frac{f(x)}{g(x)}.$$

⊣

L'Hospital's Rule also holds when either c or L or both are replaced by ∞ or $-\infty$, and for $x \to c^-$ and $x \to c$, with only routine changes in the proof. The proof of l'Hospital's Rule for ∞/∞ is more difficult.

Theorem 5.13. *(L'Hospital's Rule for ∞/∞) Suppose that for all x in some real open interval (c, b), $f'(x)$ and $g'(x)$ exist and $g'(x) \neq 0$. Assume that*

$$\lim_{x \to c^+} f(x) = \infty, \qquad \lim_{x \to c^+} g(x) = \infty.$$

If $\lim_{x\to c^+}(f'(x)/g'(x))$ *exists, then*

$$\lim_{x\to c^+}\frac{f(x)}{g(x)}=\lim_{x\to c^+}\frac{f'(x)}{g'(x)}.$$

PROOF. We will use Theorem 5.1. Let $L=\lim_{x\to c^+}(f'(x)/g'(x))$. By the Generalized Mean Value Theorem every real solution of

$$c<x<y<c+r \tag{48}$$

is a partial solution of

$$x<t<y, \qquad \frac{f'(t)}{g'(t)}=\frac{f(y)-f(x)}{g(y)-g(x)},$$

which we rewrite as

$$x<t<y, \qquad \frac{f'(t)}{g'(t)}=\frac{\frac{f(x)}{g(x)}-\frac{f(y)}{g(x)}}{1-\frac{g(y)}{g(x)}}. \tag{49}$$

Let $y_1\approx c$ and $y_1>c$. Then $f(y_1)$ and $g(y_1)$ are positive infinite, and so is their product $K=f(y_1)g(y_1)$. By the M,δ condition for $\lim_{x\to c^+}g(x)=\infty$, for every real M there is a real $\delta(M)$ such that every real solution of

$$c<x<c+\delta(M)$$

is a solution of $g(x)>M$. Let δ_1 be such that $0<\delta_1\le\delta(K)$ and $c+\delta_1<y_1$. Consider any x_1 with $c<x_1<c+\delta_1$. By Transfer, $g(x_1)>K$. Moreover,

$$c<x_1<y_1<c+x,$$

so by the Partial Solution Theorem there is a t_1 such that (49) holds. Then

$$t_1\approx c, \qquad \frac{f'(t_1)}{g'(t_1)}\approx L.$$

Also,

$$\left|\frac{f(y_1)}{g(x_1)}\right|\le\left|\frac{f(y_1)}{K}\right|=\left|\frac{f(y_1)}{f(y_1)g(y_1)}\right|\approx 0$$

so $(f(y_1)/g(x_1))\approx 0$. Similarly $(g(y_1)/g(x_1))\approx 0$. Taking standard parts in (49), we have

$$st\left(\frac{f'(t_1)}{g'(t_1)}\right)=st\left(\frac{\frac{f(x_1)}{g(x_1)}-\frac{f(y_1)}{g(x_1)}}{1-\frac{g(y_1)}{g(x_1)}}\right),$$

whence

$$L\approx\frac{f'(t_1)}{g'(t_1)}\approx\frac{f(x_1)}{g(x_1)}.$$

Since this holds for all $c<x_1<c+\delta_1$, we see from Theorem 5.1 that

$$L=\lim_{x\to c^+}\frac{f(x)}{g(x)}.$$

$\dashv$

CHAPTER 6

APPLICATIONS OF THE INTEGRAL

6A. Infinite Sum Theorem (§6.1, §6.2, §6.6)

The Infinite Sum Theorem, which is a nonstandard version of Duhamel's Principle, is a simple and extremely useful criterion for a quantity to be equal to the definite integral of a function. It can be used to justify many familiar applications of the integral in geometry and physics. It captures the intuitive idea of obtaining an integration formula by considering a typical infinitesimal element and adding up.

Recall from Chapter 2 that given a nonzero infinitesimal $\triangle x$,

$$u \approx v \text{ (compared to } \triangle x)$$

means that

$$\frac{u}{\triangle x} \approx \frac{v}{\triangle x}.$$

Theorem 6.1. *(Infinite Sum Theorem) Assume that*
(i) h is a real function which is continuous on the interval $[a, b]$.
(ii) $B(u, w)$ is a real function which has the **Addition Property***,*

$$B(u, w) = B(u, v) + B(v, w) \textit{ for } u < v < w \textit{ in } [a, b].$$

(iii) For any infinitesimal subinterval $[x, x + \triangle x]^*$ of $[a, b]^*$,

$$B(x, x + \triangle x) = h(x)\triangle x \text{ (compared to } \triangle x) .$$

Then

$$B(a, b) = \int_a^b h(x)\, dx.$$

We will write

$$\triangle B = B(x, x + \triangle x).$$

The Infinite Sum Theorem intuitively says that if each infinitesimal piece $\triangle B$ is infinitely close to $h(x)\triangle x$ (compared to $\triangle x$), then the sum $B(a, b)$ of all the $\triangle B$'s is infinitely close to the Riemann sum $\sum_a^b h(x)\triangle x$.

PROOF. Let n_1 be a positive infinite hyperinteger and let $\triangle x_1 = (b-a)/n_1$. We will prove that for every positive real number r,

$$\sum_a^b (h(x) - r)\triangle x_1 < B(a,b) < \sum_a^b (h(x) + r)\triangle x_1.$$

This will show that

$$\sum_a^b h(x)\triangle x_1 - r(b-a) < B(a,b) < \sum_a^b h(x)\triangle x_1 + r(b-a),$$

so

$$\sum_a^b h(x)\triangle x_1 \approx B(a,b)$$

and

$$\int_a^b h(x)\,dx = B(a,b).$$

Consider a positive integer n and let $\triangle x = (b-a)/n$. By the Addition Property,

$$B(a,b) = \sum_a^b B(x, x + \triangle x).$$

It follows that every real solution of

$$(50) \qquad n \in \mathbb{Z}, \quad 0 < n, \quad \triangle x = (b-a)/n, \quad B(a,b) \geq \sum_a^b (h(x) + r)\triangle x$$

is a partial real solution of

$$(51) \qquad a \leq x < x + \triangle x \leq b, \quad B(x, x + \triangle x) \geq (h(x) + r)\triangle x.$$

We may rewrite (51) as

$$(52) \qquad a \leq x < x + \triangle x \leq b, \quad \frac{B(x, x + \triangle x)}{\triangle x} \geq h(x) + r.$$

By hypothesis (iii), there is no hyperreal number x_1 such that (52) holds for $x_1, \triangle x_1$. By the Partial Solution Theorem 1.20, (50) cannot hold for $n_1, \triangle x_1$, so

$$B(a,b) < \sum_a^b (h(x) + r)\triangle x_1.$$

The proof that

$$B(a,b) > \sum_a^b (h(x) - r)\triangle x_1$$

is similar. ⊣

A standard form of the Infinite Sum Theorem can be found, for example, in the book Buck [B], Section 3.5. This treatment also covers the two variable case, which is given in Chapter 12 of this monograph. The Infinite Sum Theorem is actually a sharp form of Theorem 4.12 which states that the definite integral of f is the unique area function for f. As a first illustration of how the Infinite Sum Theorem is used, we give a second proof of the uniqueness of the area function.

Theorem 4.12 (Repeated) *The definite integral is the only area function for a continuous function f on $[a,b]$.*

Second Proof. We show that for any area function $B(u,v)$ for f,

$$B(a,b) = \int_a^b f(x)\,dx.$$

$B(u,v)$ has the Addition Property. Consider any infinitesimal subinterval $[x, x+\triangle x]^*$ of $[a,b]^*$. By the Hyperreal Extreme Value Theorem 3.33, f has a minimum value m and a maximum value M in $[x, x+\triangle x]^*$. By the Rectangle Property in Definition 4.1 and Transfer,

$$m\triangle x \le M\triangle x, \quad m \le \triangle B/\triangle x \le M.$$

Since f is continuous,

$$m \approx f(x) \approx M,$$

so

$$f(x) \approx \triangle B/\triangle x,$$

$$f(x)\triangle x \approx \triangle B \text{ (compared to } \triangle x).$$

Then by the Infinite Sum Theorem,

$$B(a,b) = \int_a^b f(x)\,dx.$$

⊣

At this point we need the real constant π.

Definition 6.2. *π is the area of the unit circle $x^2 + y^2 = 1$,*

$$\pi = \int_{-1}^{1} 2\sqrt{1-x^2}\,dx.$$

It follows by a change of variables that the circle $x^2 + y^2 = r^2$ of radius r has area $A = \pi r^2$.

We will now use the Infinite Sum Theorem to justify several formulas in geometry and physics. In each case we start with intuitively reasonable assumptions about the infinitesimal elements and apply the Infinite Sum Theorem to get an integration formula. We first justify the formulas for volumes of solids of revolution.

DEFINITION 6.3. *Let D be a basic region in the plane of the form*

$$D = \{(x, y) \colon a \leq x \leq b, \quad 0 \leq y \leq g(x)\}.$$

(i) The solid formed by revolving D about the x axis has volume

$$V = \int_a^b \pi(g(x))^2 \, dx.$$

(ii) The solid formed by revolving D about the y axis has volume

$$V = \int_a^b 2\pi \, x \, g(x) \, dx.$$

JUSTIFICATION. (i) We use the disc method. Let $V(u, w)$ be the volume of the solid generated around the x axis by the region between 0 and g over $[u, w]$. We assume:

(a) $V(u, w)$ has the Addition Property.

(b) Subset Property: If S and T are solids of revolution and $S \subseteq T$, the volume of S is at most the volume of T.

(c) A right circular cylinder with base of radius r and thickness h has volume $\pi r^2 h$.

We use (b) and (c) for solids generated by regions over an infinitesimal subinterval $[x, x+\triangle x]^*$ of $[a, b]^*$. The region between 0 and g over $[x, x+\triangle x]^*$ generates an infinitely thin disc of volume $\triangle V$. g has minimum value m and maximum value M in $[x, x + \triangle x]^*$, so $\triangle V$ has a inscribed cylinder of radius m and a circumscribed cylinder of radius M. Both cylinders have thickness $\triangle x$, so by (b) and (c),

$$\pi m^2 \triangle x \leq \triangle V \leq \pi M^2 \triangle x.$$

By continuity, $m \approx g(x) \approx M$, so

$$\triangle V \approx \pi(g(x))^2 \triangle x \text{ (compared to } \triangle x\text{)}.$$

The Infinite Sum Theorem gives

$$V(a, b) = \int_a^b \pi(g(x))^2 \, dx.$$

(ii) We use the cylindrical shell method. This time $V(u, w)$ denotes the volume generated about the y axis by the region between 0 and g over $[u, w]$. We assume the same properties (a)—(c). The region between 0 and g over $[x, x + \triangle x]^*$ generates a volume $\triangle V$ shaped like an infinitely thin cylindrical shell. By (a) and (c), the inscribed cylindrical shell has volume

$$\pi(x + \triangle x)^2 m - \pi x^2 m = \pi(2x + \triangle x)\triangle x \, m.$$

Using a similar formula for the circumscribed shell and (b), we have

$$\pi(2x + \triangle x)\triangle x \, m \leq \triangle V \leq \pi(2x + \triangle x)\triangle x \, M,$$

$$\pi(2x + \triangle x)m \leq \triangle V/\triangle x \leq \pi(2x + \triangle x)M.$$

By continuity, $m \approx g(x) \approx M$, so

$$\triangle V/\triangle x \approx \pi(2x+\triangle x)g(x) \approx 2\pi\, x\, g(x),$$

$$\triangle V \approx 2\pi\, x\, g(x)\triangle x \text{ (compared to } \triangle x).$$

The Infinite Sum Theorem gives

$$V(a,b) = \int_a^b 2\pi\, x\, g(x)\, dx.$$

$\dashv$

We give just one example from physics here. Other examples are sketched in *Elementary Calculus*, and the detailed use of the Infinite Sum Theorem can be readily worked out.

Definition 6.4. *Suppose a plane object occupies the basic region*

$$a \le x \le b, \quad 0 \le y \le g(x)$$

and its density (mass per unit area) at (x, y) is a continuous function $\rho(x)$ of x alone. The mass of the object is

$$mass = \int_a^b g(x)\, \rho(x)\, dx.$$

Justification. Let $m(u, w)$ be the mass of the piece of the object from $x = u$ to $x = v$. Our assumptions are:

(a) $m(u, w)$ has the Addition Property.

(b) If the region occupied by a plane object S is a subset of the region occupied by a plane object T, and the density of S is everywhere at most the density of T, then the mass of S is at most the mass of T.

(c) The mass of a vertical rectangular object of constant density is the product of the base, the height, and the density.

On an infinitesimal subinterval $[x, x+\triangle x]^*$ of $[a, b]^*$, the continuous functions g and ρ have minimum values g_{min}, ρ_{min} and maximum values g_{max}, ρ_{max}. $\triangle m$ is an infinitely thin strip of width $\triangle x$. Using (b) and (c),

$$g_{min}\, \rho_{min}\, \triangle x \le \triangle m \le g_{max}\, \rho_{max}\, \triangle x.$$

By continuity,

$$g_{min} \approx g(x) \approx g_{max}, \quad \rho_{min} \approx \rho(x) \approx \rho_{max},$$

and hence

$$\triangle m \approx g(x)\, \rho(x)\, \triangle x \text{ (compared to } \triangle x).$$

Applying the Infinite Sum Theorem,

$$m(a,b) = \int_a^b g(x)\, \rho(x)\, dx.$$

$\dashv$

6B. Lengths of Curves (§6.3, §6.4)

Definition 6.5. *A real curve*

$$y = f(x), \quad x \in [a, b]$$

is said to be **smooth** *if the derivative* dy/dx *is continuous on* $[a, b]$*. The* **length** *of a smooth curve is defined as*

$$s = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx.$$

Justification. Let $s(u, w)$ be the intuitive length of the curve segment over the interval $[u, w]$. Our assumptions are:

(a) $s(u, w)$ has the Addition Property.

(b) If the slope of a smooth curve is infinitely close to the slope of the chord from $P(x, y)$ to $Q(x + \triangle x, y + \triangle y)$ at every point of $[x, x + \triangle x]^*$, then

$$\frac{\triangle s}{\sqrt{\triangle x^2 + \triangle y^2}} \approx 1.$$

Here $\triangle s$ is the length of the curve segment over $[x, x+\triangle x]^*$ and $\sqrt{\triangle x^2 + \triangle y^2}$ is the length of the chord.

On an infinitesimal subinterval $[x, x + \triangle x]^*$, the infinitesimal curve segment $\triangle s$ connects $P(x, y)$ to $Q(x + \triangle x, y + \triangle y)$. The chord from P to Q has slope $\triangle y/\triangle x$. It follows from Corollary 3.38 on continuous derivatives that $f'(x) \approx \triangle y/\triangle x$. Since f' is continuous, $f'(u) \approx \triangle y/\triangle x$ for all $u \in [x, x + \triangle x]^*$. By (b) and (c),

$$\frac{\triangle s}{\sqrt{\triangle x^2 + \triangle y^2}} \approx 1.$$

Then

$$\begin{aligned}
\frac{\triangle s}{\triangle x} &= \frac{\triangle s}{\sqrt{\triangle x^2 + \triangle y^2}} \cdot \frac{\sqrt{\triangle x^2 + \triangle y^2}}{\triangle x} \\
&= \frac{\triangle s}{\sqrt{\triangle x^2 + \triangle y^2}} \sqrt{1 + \left(\frac{\triangle y}{\triangle x}\right)^2} \\
&\approx \sqrt{1 + \left(\frac{dy}{dx}\right)^2}
\end{aligned}$$

Since $dy/dx \approx \triangle y/\triangle x$,

$$\frac{\triangle s}{\triangle x} \approx \sqrt{1 + \left(\frac{\triangle y}{\triangle x}\right)^2} \approx \sqrt{1 + \left(\frac{dy}{dx}\right)^2},$$

$$\triangle s \approx \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, \triangle x \text{ (compared to } \triangle x\text{)}.$$

By the Infinite Sum Theorem,

$$s(a,b) = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx.$$

$\dashv$

As an application of the length formula we obtain the classical formula for the area of a sector of a circle.

THEOREM 6.6. *On the circle $x^2 + y^2 = r^2$ of radius r, let P be the point $(r, 0)$ and let Q be a point (x, y) in the first quadrant. Then the area A of the sector POQ is given by the formula $A = \frac{1}{2} r s$, where s is the length of the arc PQ.*

PROOF. Since the circle is vertical at P, we will take y as the independent variable instead of x. In the first quadrant the circle has the equation

$$x = \sqrt{r^2 - y^2}.$$

Then

$$s = \int_0^y \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy,$$

$$\frac{dx}{dy} = -\frac{y}{\sqrt{r^2 - y^2}} = -\frac{y}{x},$$

$$\sqrt{1 + \left(\frac{dx}{dy}\right)^2} = \sqrt{1 + \frac{y^2}{x^2}} = \frac{\sqrt{x^2 + y^2}}{x} = \frac{r}{x}.$$

Let A be the area of the sector POQ,

$$A = \int_0^y \sqrt{r^2 - u^2}\, du - \frac{1}{2}\, x\, y.$$

Letting y vary, we have

$$\frac{dA}{dy} = \sqrt{r^2 - y^2} - \frac{1}{2}\left(x + y\frac{dx}{dy}\right)$$

$$= x - \frac{1}{2}\, x - \frac{1}{2}\, y\left(-\frac{y}{x}\right) = \frac{x^2 + y^2}{2x} = \frac{r^2}{2x}.$$

Therefore

$$\frac{dA}{dy} = \frac{1}{2}\, r\sqrt{1 + \left(\frac{dx}{dy}\right)^2}.$$

Integrating from 0 to y we obtain

$$A = \frac{1}{2}\, r \int_0^y \sqrt{1 + \left(\frac{dx}{dy}\right)^2}\, dy = \frac{1}{2}\, r\, s.$$

$\dashv$

COROLLARY 6.7. *The circle $x^2 + y^2 = r^2$ has circumference $2\pi r$.*

PROOF. This follows from the formula $A = \pi r^2$ for the area of the circle. By symmetry, the first quadrant of the circle has area $A = \frac{1}{4}\pi r^2$. By Theorem 6.6, $A = \frac{1}{2}\, rs$ where s is the arc length of the first quadrant. Solving for s we get $s = \frac{1}{2}\,\pi r$. By symmetry again, the circumference is $C = 4s = 2\pi r$. ⊣

We next define the length of a parametric curve.

DEFINITION 6.8. *A* **parametric curve** *is a pair of continuous functions*

$$x = f(t), \quad y = g(t), \quad t \in [a, b].$$

It is said to be **smooth** *if at every point $t \in [a, b]$, f' and g' are continuous and at least one of $f'(t), g'(t)$ is nonzero. The* **path** *of the parametric curve is the set*

$$\{(f(t), g(t))\colon a \le t \le b\}.$$

The **length** *of a smooth parametric curve is the integral*

$$s = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt.$$

The justification of the parametric length formula is exactly like the justification of the length formula for curves of the form $y = f(x)$. Notice that when $x = t$, the parametric equation and length formula reduce to the special case

$$y = g(x), \quad a \le x \le b, \quad s = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx.$$

DEFINITION 6.9. *A smooth parametric curve*

$$C\colon x = f(t), \quad y = g(t), \quad t \in [a, b]$$

is said to be **simple** *if C maps $[a, b]$ one to one onto its path and there is no t with $f'(t) = g'(t) = 0$. C_1 is a* **reparametrization** *of C if C and C_1 are simple parametric curves with the same path P.*

The next theorem shows that the length of a simple parametric curve depends only on its path.

THEOREM 6.10. *(Reparametrization Theorem) Let*

$$C\colon x = f(t), \quad y = g(t), \quad t \in [a, b]$$

be a simple parametric curve.

(i) A parametric curve

$$C_1\colon x = F(u), \quad y = G(u), \quad u \in [A, B]$$

is a reparametrization of C if and only if there is a smooth function $t = h(u)$ mapping $[A, B]$ onto $[a, b]$ such that $h'(u)$ is never zero and for all $u \in [A, B]$,

$$F(u) = f(h(u)), \quad G(u) = g(h(u)).$$

(ii) If C_1 is a reparametrization of C then C and C_1 have the same length.

Proof. (i) Suppose C_1 is a reparametrization of C. There are two cases for the endpoints.

Case 1: $F(A) = f(a), \quad G(A) = g(a)$.

Case 2: $F(A) = f(b), \quad G(A) = g(b)$.

We give the proof in Case 1, where both curves move in the same direction. Case 2 is similar, but with the curves moving in opposite directions. Since C and C_1 are one to one functions onto the same path P, there is a function h mapping $[A, B]$ onto $[a, b]$ such that whenever $t = h(u)$,

$$(F(u), G(u)) = (f(t), g(t)).$$

The function h is one to one because C and C_1 are one to one.

We claim that h is continuous. To see this, we assume that $u, v \in [A, B]^*$ and $\mathrm{st}(u) = \mathrm{st}(v)$, and prove that $\mathrm{st}(h(u)) = \mathrm{st}(h(v))$. Since F and f are continuous, and $F(u) = f(h(u))$, we have

$$f(\mathrm{st}(h(u))) = \mathrm{st}(f(h(u))) = \mathrm{st}(F(u)) = F(\mathrm{st}(u)) = F(\mathrm{st}(v)).$$

Using the same argument at v,

$$f(\mathrm{st}(h(v))) = \mathrm{st}(f(h(v))) = \mathrm{st}(F(v)) = F(\mathrm{st}(v)).$$

Thus $f(\mathrm{st}(h(u))) = f(\mathrm{st}(h(v)))$. A similar fact holds for g, so

$$(f(\mathrm{st}(h(u))), g(\mathrm{st}(h(u)))) = (f(\mathrm{st}(h(v))), g(\mathrm{st}(h(v)))).$$

Since the curve C maps $[a, b]$ one to one onto the path P, we have $\mathrm{st}(h(u)) = \mathrm{st}(h(v))$ as required.

In Case 1 we have $h(A) = a$. Since h is continuous and maps $[A, B]$ one to one onto $[a, b]$, it follows from Theorem 3.23 that h is increasing.

We now show that h is smooth and $h'(u)$ is never zero. Let $A \le u < u + \triangle u \le B$ with $\triangle u \approx 0$, and let $\triangle x, \triangle y$, and $\triangle t$ be the corresponding changes in $x = F(u)$, $y = G(u)$, and $t = h(u)$. Since F, G, and h are continuous, $\triangle x, \triangle y$, and $\triangle t$ are infinitesimal. By Transfer, h^* is increasing, so $\triangle t > 0$. Since C is a simple curve, either $F'(\mathrm{st}(u)) \ne 0$ or $G'(\mathrm{st}(u)) \ne 0$, say $F'(\mathrm{st}(u)) \ne 0$. The function F is smooth, so its derivative F' is continuous. Then by Theorem 3.37, $F'(\mathrm{st}(u)) \approx \triangle x / \triangle u$, and therefore $\triangle x \ne 0$. It follows that $\sqrt{\triangle x^2 + \triangle y^2}$ is positive infinitesimal. We thus have

$$\begin{aligned} \frac{\triangle t}{\triangle u} &= \frac{\sqrt{\triangle x^2 + \triangle y^2}/\triangle u}{\sqrt{\triangle x^2 + \triangle y^2}/\triangle t} = \frac{\sqrt{(\triangle x/\triangle u)^2 + (\triangle y/\triangle u)^2}}{\sqrt{(\triangle x/\triangle t)^2 + (\triangle y/\triangle t)^2}} \\ &= \frac{\sqrt{F'(\mathrm{st}(u))^2 + G'(\mathrm{st}(u))^2}}{\sqrt{f'(\mathrm{st}(t))^2 + g'(\mathrm{st}(t))^2}}, \end{aligned}$$

so for real u, $h'(u)$ exists and

$$h'(u) = \frac{\sqrt{F'(u)^2 + G'(u)^2}}{\sqrt{f'(h(u))^2 + g'(h(u))^2}} \ne 0.$$

Since h, F', G', f', and g' are continuous, the derivative h' is continuous.

We now prove the converse. Since h' is never zero, h maps $[A, B]$ one to one onto $[a, b]$. Therefore C_1 maps $[A, B]$ one to one onto the path P. By the Chain Rule,

$$F'(u) = f'(h(u))h'(u), \quad G'(u) = g'(h(u))h'(u).$$

Thus F' and G' are continuous and are never both zero, so C_1 is a simple parametric curve.

(ii) We give the proof in Case 1, where $(F(A), G(A)) = (f(a), g(a))$. Let h be as in part (i). Then $h'(u) > 0$, so

$$\begin{aligned}\text{length of } C &= \int_a^b \sqrt{f'(t)^2 + g'(t)^2}\, dt \\ &= \int_A^B \sqrt{f'(h(u))^2 + g'(h(u))^2}\, h'(u)\, du \\ &= \int_A^B \sqrt{F'(u)^2 + G'(u)^2}\, du = \text{ length of } C_1.\end{aligned}$$

The proof in Case 2, where $(F(A), G(A)) = (f(b), g(b))$, is similar but with $h'(u) < 0$ and the integral from B to A. ⊣

We now justify the formula for the area of a surface of revolution. Our starting point is the elementary formula

$$A = \pi(r_1 + r_2)\ell$$

for the surface area of a circular cone frustum of slant height ℓ and bases of radius r_1 and r_2.

DEFINITION 6.11. *The* **surface area** *formed by revolving a smooth curve*

$$y = f(x), \quad a \le x \le b$$

about the y axis is

$$A = \int_a^b 2\pi\, x \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx.$$

The ***surface area*** *formed by revolving the curve about the x axis is*

$$A = \int_a^b 2\pi\, y \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx.$$

JUSTIFICATION. We justify the area formula for the surface of revolution about the y axis. Let $s(u, v)$ and $A(u, v)$ be the length of the curve and the area of the surface of revolution for $u \le x \le v$. Assume that $A(u, v)$ has the Addition Property. Let $[x, x + \triangle x]^*$ be an infinitesimal subinterval of $[a, b]^*$. When the line segment from (x, y) to $(x + \triangle x, y + \triangle y)$ is revolved about the y axis it generates an infinitely thin cone frustum of slant height $\sqrt{\triangle x^2 + \triangle y^2}$ and bases of radius x and $x + \triangle x$. This frustum has surface area

$$\pi(x + (x + \triangle x))\sqrt{\triangle x^2 + \triangle y^2}.$$

The curve $y = f(x)$ has value and slope infinitely close to the line segment from (x, y) to $(x + \triangle x, y + \triangle y)$ throughout the interval $[x, x + \triangle x]^*$. It is thus reasonable to assume that the surface area $\triangle A = A(x, x + \triangle x)$ and the frustum area have a ratio infinitely close to one,

$$\frac{\triangle A}{\pi(x + (x + \triangle x))\sqrt{\triangle x^2 + \triangle y^2}} \approx 1.$$

Since $\triangle y/\triangle x \approx dy/dx$,

$$\begin{aligned}\frac{\triangle A}{\pi(x + (x + \triangle x))\sqrt{\triangle x^2 + \triangle y^2}} &\approx \frac{\triangle A}{2\pi\, x\sqrt{\triangle x^2 + \triangle y^2}} \\ &= \frac{\triangle A/\triangle x}{2\pi\, x\sqrt{1 + \left(\frac{\triangle y}{\triangle x}\right)^2}} \approx \frac{\triangle A/\triangle x}{2\pi\, x\sqrt{1 + \left(\frac{dy}{dx}\right)^2}}.\end{aligned}$$

Therefore

$$\triangle A \approx 2\pi\, x\sqrt{1 + \left(\frac{dy}{dx}\right)^2}\,\triangle x \text{ (compared to } \triangle x),$$

and by the Infinite Sum Theorem,

$$A(a, b) = \int_a^b 2\pi\, x\sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx.$$

The justification of the area formula for the surface of revolution about the x axis is similar. ⊣

Given a simple parametric curve

$$x = f(t), \quad y = g(t), \quad t \in [a, b],$$

the area of the surface of revolution about the y axis is defined by

$$A = \int_a^b 2\pi\, x\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt.$$

6C. Improper Integrals (§6.7)

In Chapter 4 we only considered integrals of continuous functions over closed intervals. Using improper integrals we can integrate functions with at most finitely many discontinuities over arbitrary intervals.

Definition 6.12. *Suppose f is continuous on the half-open interval $[a, b)$. The* **improper integral** *of f from a to b is defined as the limit*

$$\int_a^b f(x)\, dx = \lim_{u \to b^-} \int_a^u f(x)\, dx.$$

If the limit exists, the integral is said to **converge**. *Otherwise the integral is said to* **diverge**. *If the limit is* ∞ *we say that the integral* **diverges to** ∞ *and write*

$$\int_a^b f(x)\,dx = \infty.$$

Other types of improper integrals are defined analogously. For example: If f *is continuous on* $(a, b]$,

$$\int_a^b f(x)\,dx = \lim_{u\to a^+} \int_u^b f(x)\,dx.$$

If f *is continuous on* $[a, \infty)$,

$$\int_a^\infty f(x)\,dx = \lim_{u\to\infty} \int_a^u f(x)\,dx.$$

The definitions can be rephrased using the infinitesimal definition of limit. We state four cases.

Let f be continuous on $[a, b)$.
$\int_a^b f(x)\,dx = L$ if and only if whenever $u < b$ but $u \approx b$, $\int_a^u f(x)\,dx \approx L$.
$\int_a^b f(x)\,dx = \infty$ if and only if whenever $u < b$ but $u \approx b$, $\int_a^u f(x)\,dx$ is positive infinite.

Let f be continuous on $[a, \infty)$.
$\int_a^\infty f(x)\,dx = L$ if and only if whenever H is positive infinite, $\int_a^H f(x)\,dx \approx L$.
$\int_a^\infty f(x)\,dx = \infty$ if and only if whenever H is positive infinite, $\int_a^H f(x)\,dx$ is positive infinite.

THEOREM 6.13. *If* f *is continuous on* $[a, b]$, *then the improper integral of* f *from* a *to* b *converges and equals the definite integral.*

PROOF. Let $F(u) = \int_a^u f(x)\,dx$. By the Second Fundamental Theorem of Calculus 4.17, F is continuous on $[a, b]$. Therefore

$$\int_a^b f(x)\,dx = F(b) = \lim_{u\to b^-} F(u) = \lim_{u\to b^-} \int_a^u f(x)\,dx.$$

$\dashv$

THEOREM 6.14. *Let* f *be continuous and nonnegative for* x *in* $[a, \infty)$. *Then the improper integral* $\int_a^\infty f(x)\,dx$ *either converges to some finite value or diverges to* ∞.

PROOF. Let $F(u) = \int_a^u f(x)\,dx$. By the Second Fundamental Theorem of Calculus, F is continuous on $[a, \infty)$. Since f is nonnegative, it follows from Theorem 4.9 that F is nondecreasing on $[a, \infty)$.

Case 1: The range of F is bounded. Then it has a least upper bound L. Let H be positive infinite. Since every real $u \geq a$ is a solution of $F(u) \leq L$, Transfer gives $F(H) \leq L$. Consider a real number $r < L$. There is a real $u \geq a$

such that $F(u) > r$. Any real solution of $v \geq u$ is a solution of $F(v) \geq F(u)$, and hence $F(v) > r$. Therefore $F(H) > r$. Since this holds for all real $r < L$, we conclude that $F(H) \approx L$, so $\int_a^\infty f(x)\,dx = L$.

Case 2: The range of F is not bounded. Again let H be positive infinite. For every real r there is a real $u \geq a$ such that $F(u) > r$. Since $F(H) \geq F(u)$, we have $F(H) > r$ and hence $F(H)$ is positive infinite. Thus $\int_a^\infty f(x)\,dx = \infty$. $\dashv$

CHAPTER 7

TRIGONOMETRIC FUNCTIONS

In the study of trigonometric functions we use the notion of arc length on the unit circle, which depends on the integration formula for curve length from Chapter 6,

$$s = \int_a^b \sqrt{1 + \left(\frac{dy}{dx}\right)^2}\, dx.$$

To prepare the way for this study we introduce inverse functions.

7A. Inverse Function Theorem (§7.3)

Given a binary relation $X \subseteq \mathbb{R}^2$ on the reals, the **inverse relation** of X is the binary relation

$$Y = \{(y, x)\colon (x, y) \in X\}.$$

Obviously, if Y is the inverse relation of X then X is the inverse relation of Y. If f and g are real functions and g is the inverse relation of f, we call g the **inverse function** of f. Here are some simple facts about inverse functions.

LEMMA 7.1. *(i) A real function f has an inverse function if and only if f is one to one.*

(ii) g is the inverse function of f if and only if the equations

$$f(x) = y, \quad g(y) = x$$

have the same solutions.

(iii) g is the inverse function of f if and only if

$$\mathrm{domain}(g) = \mathrm{range}(f)$$

and

$$g(f(x)) = x \textit{ for all } x \in \mathrm{domain}(f).$$

We remark that if f is increasing, then f is one to one and therefore has an inverse function. Similarly, if f is decreasing then f has an inverse function.

THEOREM 7.2. *Let f be increasing and continuous on its domain which is an interval I, and let g be the inverse function of f. Then g is increasing, the domain of g is an interval J, and g is continuous on J.*

PROOF. Let J be the domain of g. Suppose $A, B \in J$ and $A < B$. Let $a = g(A)$, $b = g(B)$. Then $f(a) = A, f(b) = B$. We cannot have $b \leq a$ because that would imply that $f(b) \leq f(a), B \leq A$. Therefore $a < b$, so g is increasing.

Suppose $C \in (A, B)$. By the Intermediate Value Theorem 3.28 for the continuous function f, there is a point $c \in (a, b)$ such that $f(c) = C$. Therefore C belongs to the range of f, which is the domain J of g. This shows that J is an interval. To show that g is continuous, let $A \in J, X \in J^*, A \approx X$, and put $a = g(A), x = g(X)$. We must show that $a \approx x$. $C \in J$ and $c = g(C)$ implies $c \in I$ and $f(c) = C$, so by Transfer we have $x \in I^*$ and $f(x) = X$. Assume first that $A < X$. Since g is increasing, g^* is increasing by Transfer, so $a < x$. If $a \not\approx x$, then there is a real number $c \in I$ such that $a < c < x$, hence $f(a) < f(c) < f(x)$, contradicting $f(a) = A \approx X = f(x)$. This shows that $a \approx x$. The case $X > A$ is similar, so g is continuous on J. $\dashv$

THEOREM 7.3. *(Inverse Function Theorem) Suppose f is continuous and increasing on an open interval I, and g is the inverse function of f. For any point $x \in I$ where f is differentiable and $f'(x) \neq 0$, g is differentiable at $y = f(x)$ and $g'(y) = 1/f'(x)$.*

PROOF. Let $\triangle y$ be a nonzero infinitesimal. We must show that

$$\frac{g(y + \triangle y) - g(y)}{\triangle y} \approx \frac{1}{f'(x)}.$$

Let J be the range of f and the domain of g. Since I is an open interval, J is an open interval. Thus $y = f(x)$ is an interior point of J. Therefore $g(y + \triangle y)$ is defined. We have $x = g(y)$, and we put

$$\triangle x = g(y + \triangle y) - g(y).$$

By Theorem 7.2, g is continuous and increasing, so $\triangle x$ is infinitesimal and not zero. Also,

$$x + \triangle x = g(y + \triangle y), \quad f(x + \triangle x) = y + \triangle y.$$

Then

$$f'(x) \approx \frac{f(x + \triangle x) - f(x)}{\triangle x} = \frac{\triangle y}{\triangle x},$$

and since $f'(x) \neq 0$,

$$\frac{1}{f'(x)} \approx \frac{\triangle x}{\triangle y} = \frac{g(y + \triangle y) - g(y)}{\triangle y}.$$

$\dashv$

We now apply the Inverse Function Theorem to show that in a simple parametric curve we may take the curve length itself as the independent variable.

THEOREM 7.4. *Let*

$$C \colon x = f(t), \quad y = g(t), \quad t \in [a, b]$$

be a simple parametric curve. Then C has a reparametrization

$$C_1 \colon x = F(s), \quad y = G(s), \quad s \in [0, L]$$

such that s is equal to the length of the curve from 0 to s.

Proof. Let $\ell(t)$ be the curve length from a to t,

$$\ell(t) = \int_a^t \sqrt{f'(u)^2 + g'(u)^2}\, du.$$

By the Second Fundamental Theorem of Calculus, 4.17, $\ell'(t) = \sqrt{f'(t)^2 + g'(t)^2}$. This is always positive and continuous, and ℓ maps $[a, b]$ onto $[0, L]$ where L is the length of C. By the Inverse Function Theorem 7.3, the inverse function h of ℓ maps $[0, L]$ onto $[a, b]$ and has the derivative $h'(s) = 1/\ell'(h(s)) > 0$. Since h and ℓ' are continuous, h' is continuous. By the Reparametrization Theorem 6.10, the curve

$$C_1 \colon x = f(h(s)), \quad y = g(h(s)), \quad s \in [0, L]$$

is a reparametrization of C, and the length of C_1 from 0 to s is $\ell(h(s)) = s$. ⊣

We will now prove a Local Inverse Function Theorem, due to Behrens (see the book of Stroyan and Luxemburg [SL 1976]). It does not require the hypothesis that f is continuous on a neighborhood, and will be convenient when we study two variables later.

Recall from Chapter 3 that a real function f is uniformly differentiable at a real point c if $f'(c)$ exists and whenever $x \approx c$ and $\triangle x$ is nonzero infinitesimal,

$$f'(c) \approx \frac{f(x + \triangle x) - f(x)}{\triangle x}.$$

Theorem 7.5. *(Local Inverse Function Theorem) Suppose f is defined in a neighborhood of a real point c, increasing, uniformly differentiable at c, and that $f'(c) \neq 0$. Then the inverse function g of f is uniformly differentiable at $d = f(c)$ and $g'(d) = 1/f'(c)$.*

Proof. Since f is increasing, g is increasing. By Theorem 3.39, f is continuous on some real neighborhood I of c. Then by the Inverse Function Theorem 7.3, $g'(d)$ exists and $g'(d) = 1/f'(c)$. To show that g is uniformly differentiable at d, suppose $y \approx d$ and $\triangle y$ is nonzero infinitesimal. Let $x = g(y)$ and $\triangle x = g(y + \triangle y) - g(y)$. Since g increasing and continuous, $\triangle x$ is a nonzero infinitesimal. f is uniformly differentiable at c, so

$$f'(c) \approx \frac{f(x + \triangle x) - f(x)}{\triangle x}.$$

From the definition of $\triangle x$ we see that

$$g(y + \triangle y) = g(y) + \triangle x = x + \triangle x,$$

so

$$y + \triangle y = f(x + \triangle x)$$

and

$$\triangle y = f(x + \triangle x) - y = f(x + \triangle x) - f(x).$$

Therefore

$$g'(d) = 1/f'(c) \approx \frac{\triangle x}{f(x + \triangle x) - f(x)} = \frac{g(y + \triangle y) - g(y)}{\triangle y}.$$

This shows that g is uniformly differentiable at d. ⊣

7B. Derivatives of Trigonometric Functions (§7.1, §7.2)

We first define the trigonometric functions and then use the Inverse Function Theorem 7.3 to show that they are differentiable.

Definition 7.6. *Let $0 \leq \theta \leq \pi/2$ and let $P(x, y)$ be the point at distance θ counter-clockwise around the unit circle starting from $(1, 0)$. We then define*

$$x = \cos\,\theta, \quad y = \sin\,\theta.$$

Lemma 7.7. *For $0 \leq \theta \leq \pi/2$, the functions $y = sin\,\theta$ and $x = \cos\,\theta$ are differentiable, and*

$$\frac{d(\sin\,\theta)}{d\theta} = \cos\,\theta, \quad \frac{d(\cos\,\theta)}{d\theta} = -\sin\,\theta.$$

Proof. We repeat the argument used for Theorem 7.4 but for the special case $x = \sqrt{1 - y^2}, \quad 0 \leq y < 1$. We have

$$\frac{dx}{dy} = \frac{-y}{\sqrt{1 - y^2}},$$

$$\sqrt{1 + \left(\frac{dx}{dy}\right)^2} = \sqrt{1 + \frac{y^2}{1 - y^2}} = \frac{1}{\sqrt{1 - y^2}}.$$

Let θ be the arc length

$$\theta = \int_0^y \frac{1}{\sqrt{1 - u^2}}\,du.$$

By definition, $x = \cos\,\theta$ and $y = \sin\,\theta$. For $0 \leq y < 1$, θ is increasing and has the continuous derivative

$$\frac{d\theta}{dy} = \frac{1}{\sqrt{1 - y^2}}.$$

By the Inverse Function Theorem 7.3, $y = \sin\,\theta$ is differentiable and

$$\frac{dy}{d\theta} = \frac{1}{d\theta/dy} = \sqrt{1 - y^2} = x = \cos\,\theta.$$

Similarly,

$$\frac{dx}{d\theta} = \frac{dx}{dy}\frac{dy}{d\theta} = -y = -\sin\,\theta.$$

⊣

We extend the sine and cosine functions from the interval $[0, \pi/2]$ to the whole real line by defining

$$\begin{aligned} \sin(\theta + \pi/2) &= \cos\,\theta, \\ \sin(\theta + \pi) &= -\sin\,\theta. \end{aligned}$$

One can easily check that the sine and cosine functions have period 2π. Given the derivative formulas for $0 \le \theta < \pi/2$ in Lemma 7.7, we leave the proof of the formulas for arbitrary θ to the reader.

Theorem 7.8.

$$\frac{d(\sin\,\theta)}{d\theta} = \cos\,\theta, \qquad \frac{d(\cos\,\theta)}{d\theta} = -\sin\,\theta.$$

The other trigonometric functions are defined by

$$\tan\,\theta = \frac{\sin\,\theta}{\cos\,\theta}, \quad \cot\,\theta = \frac{\cos\,\theta}{\sin\theta},$$

$$\sec\,\theta = \frac{1}{\cos\,\theta}, \quad \csc\,\theta = \frac{1}{\sin\theta}.$$

The inverse trigonometric functions are obtained by restricting the trigonometric functions to either $[-\pi/2, \pi/2]$ or $[0, \pi]$ and then taking the inverse functions. For example,

$$\begin{aligned} &\text{arc}\ \sin x \text{ is the inverse of } \sin\theta, && -\pi/2 \le \theta \le \pi/2, \\ &\text{arc}\ \tan x \text{ is the inverse of } \tan\theta, && -\pi/2 \le \theta \le \pi/2, \\ &\text{arc}\ \sec x \text{ is the inverse of } \sec\theta, && 0 \le \theta \le \pi. \end{aligned}$$

The Inverse Function Theorem 7.3 shows that the inverse trigonometric functions are differentiable except at the endpoints and leads in the usual way to the formulas for the derivatives:

$$\begin{aligned} \frac{d(\text{arc}\ \sin x)}{dx} &= \frac{1}{\sqrt{1-x^2}}, && |x| < 1, \\ \frac{d(\text{arc}\ \tan x)}{dx} &= \frac{1}{1+x^2}, && \text{all } x, \\ \frac{d(\text{arc}\ \sec x)}{dx} &= \frac{1}{|x|\sqrt{x^2-1}}, && |x| > 1. \end{aligned}$$

7C. Area in Polar Coordinates (§7.9)

We assume that the reader is familiar with the polar coordinate system. The Infinite Sum Theorem 6.1 can be used to obtain a formula for area in polar coordinates.

DEFINITION 7.9. *By a* **basic polar region** *we mean a set of points with polar coordinates of the form*

$$\{(r,\theta)\colon a \le \theta \le b,\ f(\theta) \le r \le g(\theta)\},$$

where $b \le a + 2\pi$, f *and* g *are continuous on* $[a,b]$, *and* $0 \le f(\theta) \le g(\theta)$.

Thus the image of a basic polar region under the mapping

$$(r,\theta) \mapsto (r\cos\theta, r\sin\theta)$$

is a basic rectangular region. If the functions f and g are constants the region is called a **polar rectangle**. The simplest kind of polar rectangle is a circular sector

$$\{(r,\theta)\colon a \le \theta \le b,\ 0 \le r \le c\}.$$

Our starting point for polar areas is the following formula for the area of a circular sector:

$$A = \frac{1}{2}\,c^2(b-a).$$

This formula comes from the formula $A = \frac{1}{2}\,rs$ for a circular sector where r is the radius of the circle and $s = (b-a)r$ is the length of the arc. Consider a polar region of the simple form

$$D = \{(r,\theta)\colon a \le \theta \le b,\ 0 \le r \le g(\theta)\}.$$

DEFINITION 7.10. *Let* $g(\theta)$ *be a nonnegative continuous real function for* $a \le \theta \le b$, *where* $b \le a + 2\pi$. *By a* **polar area function** *for* g *we mean a function* $A(u,v)$ *defined for* $u, v \in [a,b]$ *with the following two properties.*
(i) $A(u,w) = A(u,v) + A(v,w)$.
(ii) If g *has minimum value* m *and maximum value* M *on* $[u,v]$, *then*

$$\frac{1}{2}\,m^2(v-u) \le A(u,v) \le \frac{1}{2}\,M^2(v-u).$$

Condition (ii) says that $A(u,v)$ is between the areas of the inscribed and circumscribed circular sectors for $u \le \theta \le v$.

THEOREM 7.11. *The unique polar area function for* g *is the definite integral*

$$A(a,b) = \int_a^b \frac{1}{2}\,g(\theta)^2\,d\theta.$$

PROOF. $A(u,v)$ is a polar area function for g because

$$\int_u^v \frac{1}{2}\,m^2\,d\theta \le \int_u^v \frac{1}{2}\,g(\theta)^2\,d\theta \le \int_u^v \frac{1}{2}\,M^2\,d\theta,$$

$$\int_u^v \frac{1}{2}\,m^2\,d\theta = \frac{1}{2}\,m^2(v-u), \qquad \int_u^v \frac{1}{2}\,M^2\,d\theta = \frac{1}{2}\,M^2(v-u).$$

To prove uniqueness let $B(u,v)$ be a polar area function for g. Let $[\theta, \theta + \triangle\theta]^*$ be an infinitesimal subinterval of $[a,b]^*$. On $[\theta, \theta + \triangle\theta]^*$, g has a minimum value m and a maximum value M. By the continuity of g, $m \approx g(\theta) \approx M$. Applying the Transfer Axiom to property (ii) we have

$$\frac{1}{2}\, m^2 \triangle\theta \le \triangle B \le \frac{1}{2}\, M^2 \triangle\theta,$$

$$\frac{1}{2}\, m^2 \le \frac{\triangle B}{\triangle\theta} \le \frac{1}{2}\, M^2,$$

$$\frac{\triangle B}{\triangle\theta} \approx \frac{1}{2}\, g(\theta)^2,$$

$$\triangle B \approx \frac{1}{2}\, g(\theta)^2 \triangle\theta \text{ (compared to } \triangle\theta) .$$

By (i), B has the Addition Property. Therefore by the Infinite Sum Theorem 6.1,

$$B(a,b) = \int_a^b \frac{1}{2}\, g(\theta)^2\, d\theta.$$

$\dashv$

The area formula for an arbitrary basic polar region, which is justified in a similar way, is

$$A = \int_a^b \frac{1}{2}\, (g(\theta)^2 - f(\theta)^2)\, d\theta.$$

CHAPTER 8

EXPONENTIAL FUNCTIONS

The exponential and logarithmic functions have been introduced in a variety of ways in calculus texts. Our approach here (and in *Elementary Calculus*) is to define a^x as the unique continuous function of x which has the value $a^{m/n} = \sqrt[n]{a^m}$ when $x = m/n$ is rational. In this chapter we will use hyperreal numbers to prove a general result on uniquely extending continuous functions, and then apply the result to the case of the exponential functions.

8A. Extending Continuous Functions

Let us recall the hyperreal characterizations of the closure of a set of reals and of uniform continuity. By Corollary 1.29, the closure of a set $Y \subseteq \mathbb{R}$ is the set

$$\overline{Y} = \{\mathrm{st}(y) \colon y \text{ is finite and } y \in Y^*\}.$$

By Definition 3.12, a real function f is uniformly continuous on a set $Y \subseteq \mathbb{R}$ if and only if for all $x, y \in Y^*$, $x \approx y$ implies $f(x) \approx f(y)$.

We will see in the next section that the exponential function a^x is not uniformly continuous on the set $\mathbb{Q}$ of all rationals, but is uniformly continuous on every bounded subset of $\mathbb{Q}$.

PROPOSITION 8.1. *A real function f is uniformly continuous on every bounded subset of Y if and only if for every finite $x, y \in Y^*$, $x \approx y$ implies $f(x) \approx f(y)$.*

PROOF. This follows at once from the fact that x, y are finite elements of Y^* if and only if $x, y \in (Y \cap [a, b])^*$ for some interval $[a, b]$. ⊣

THEOREM 8.2. *Let f be a real function which is uniformly continuous on every bounded subset of its domain Y. Then f has a unique extension g whose domain is the closure $\overline{Y}$ of Y and which is continuous on $\overline{Y}$.*

PROOF. Let g be the real function on $\overline{Y}$ defined by

$$g(\mathrm{st}(x)) = \mathrm{st}(f(x)) \text{ for all finite } x \in Y^*.$$

This definition is unambiguous because if $\mathrm{st}(x) = \mathrm{st}(y)$ then $\mathrm{st}(f(x)) = \mathrm{st}(f(y))$. g extends f, because if $r \in Y$ then

$$g(r) = g(\mathrm{st}(r)) = \mathrm{st}(f(r)) = f(r).$$

We show that g is continuous at each point $c \in \overline{Y}$. Let X be the bounded set

$$X = Y \cap [c-1, c+1].$$

We use the ε, δ condition for uniform continuity, Theorem 5.7. Consider a real $\varepsilon > 0$. Since f is uniformly continuous on X, there is a real $\delta \in (0,1)$ such that whenever $x, y \in X$ and $|x-y| < \delta$, $|f(x)-f(y)| < \varepsilon/2$. By Transfer, whenever $x, y \in X^*$ and $|x-y| < \delta$, $|f(x)-f(y)| < \varepsilon/2$. Now let $b \in \overline{Y}$ and $|b-c| < \delta$. We have $b = \mathrm{st}(x), c = \mathrm{st}(y)$ for some $x, y \in Y^*$. Since $|b-c| < \delta < 1$ we have

$$x, y \in X^*, \quad |x-y| < \delta.$$

Therefore

$$\begin{gathered} |f(x)-f(y)| < \varepsilon/2 \\ |\mathrm{st}(f(x)) - \mathrm{st}(f(y))| \le \varepsilon/2 < \varepsilon \\ |g(b)-g(c)| < \varepsilon. \end{gathered}$$

Thus the ε, δ condition for continuity holds for g at c. The function g is unique because for any continuous extension h of f to $\overline{Y}$ and any point $c \in \overline{Y}$ we have

$$h(c) = h(\mathrm{st}(x)) = \mathrm{st}(h(x)) = \mathrm{st}(f(x)) = g(c)$$

where $x \in Y^*$ and $c = \mathrm{st}(x)$, $\dashv$

8B. The Functions a^x and $\log_b x$ (§8.1, §8.2)

In *Elementary Calculus*, hyperreal numbers were used to define and obtain the basic properties of the exponential function. The natural extension of the set $\mathbb{Q}$ of rationals is called the set $\mathbb{Q}^*$ of **hyperrationals**. Here are some properties of $\mathbb{Q}^*$ which follow at once from the Transfer Axiom.

PROPOSITION 8.3. *(i) $\mathbb{Q}^*$ is a subfield of $\mathbb{R}^*$, that is, $\mathbb{Q}^*$ is closed under addition, subtraction, multiplication, and division by nonzero numbers.*

(ii) $\mathbb{Q}^*$ is dense in $\mathbb{R}^*$, that is, if $b < c$ in $\mathbb{R}^*$ then there exists $q \in \mathbb{Q}^*$ with $b < q < c$.

(iii) For every $x \in \mathbb{R}^*$ there is a $q \in \mathbb{Q}^*$ such that $x \approx q$. (This follows from (ii)).

The closure of the set $\mathbb{Q}$ of rationals is the whole real line $\mathbb{R}$. One way to see this is by applying (iii) to $x \in \mathbb{R}$. We will use the following properties of rational exponents.

LEMMA 8.4. *Let a, b be positive real numbers and q, r be rational. Then:*

(i) $1^q = 1$

(ii) $a^{q+r} = a^q a^r, \quad a^{q-r} = a^q/a^r$

(iii) $a^{qr} = (a^q)^r$

(iv) $a^q b^q = (ab)^q$

(v) $a < b$ *and* $q > 0$ *imply* $a^q < b^q$

(vi) $1 < a$ *and* $q < r$ *imply* $a^q < a^r$

(vii) $q \geq 1$ *implies* $(a+1)^q \geq aq+1$.

LEMMA 8.5. *Let* $a > 0$. *The function* $f(x) = a^x$, $x \in \mathbb{Q}$, *is uniformly continuous on every bounded subset of* $\mathbb{Q}$.

PROOF. Assume first that $a \geq 1$. Suppose $q, r \in \mathbb{Q}^*$, $q \approx r$, $q < r$. We must show that $a^q \approx a^r$. Let $b = a^{(q-r)} - 1$. By Lemma 8.4 (vi) and Transfer, $b \geq 0$. Moreover, by Lemma 8.4 (vii) and Transfer,

$$a = (b+1)^{1/(r-q)} \geq \frac{b}{r-q} + 1 \geq 1,$$

so $b/(r-q)$ is finite and hence $b \approx 0$. Thus $a^{r-q} \approx 1$. For some integer n, $n \leq q < n+1$, and by Lemma 8.4 (vi). $a^n < a^q < a^{n+1}$. Therefore a^q is finite and $a^r \approx a^q$.

Now assume $0 < a < 1$. Then $a^{-1} > 1$, so by the preceding paragraph,

$$a^r = \left(a^{-1}\right)^{-r} \approx \left(a^{-1}\right)^{-q} = a^q.$$

⊣

By Theorem 8.2 we may make the following definition.

DEFINITION 8.6. *Let* a *be a positive real number. The* **exponential function with base** a, a^x, *is the unique extension of the function* a^q, $q \in \mathbb{Q}$ *which is continuous on the whole real line.*

THEOREM 8.7. *The exponent rules (i)—(vii) of Lemma 8.4 hold for arbitrary real numbers* q *and* r.

PROOF. All the rules except (iii) are easily proved by using Lemma 8.4 and the fact that there are hyperrational numbers $q_1 \approx q$ and $r_1 \approx r$. We prove (iii) for the case $1 < a$ and $q, r > 0$. Choose hyperrational numbers q_1, q_2 such that

$$q_1 \approx q_2, \quad q_1 < q < q_2.$$

Do the same for r. We have

$$q_1 r_1 < qr < q_2 r_2.$$

By (vi) and Transfer,

$$a^{q_1 r_1} < a^{qr} < a^{q_2 r_2},$$

$$a^{q_1} < a^q < a^{q_2},$$

$$\left(a^{q_1}\right)^{r_1} < \left(a^q\right)^r < \left(a^{q_2}\right)^{r_2}.$$

Using (iii) for hyperrational exponents,

$$a^{q_1 r_1} = \left(a^{q_1}\right)^{r_1}, \quad a^{q_2 r_2} = \left(a^{q_2}\right)^{r_2}.$$

But $q_1 r_1 \approx qr \approx q_2 r_2$, so by the continuity of the function a^x, $a^{q_1 r_1} \approx a^{qr} \approx a^{q_2 r_2}$. It follows that $a^{qr} \approx (a^q)^r$, and since both numbers are real they must be equal. ⊣

Definition 8.8. *Let* $0 < a$ *and* $a \neq 1$. *The* **logarithmic function with base** a *is the inverse of the exponential function with base* a,

$$x = \log_a y \text{ if and only if } y = a^x.$$

By Theorem 7.2, $\log_a y$ is continuous and has domain $(0, \infty)$. The familiar rules for logarithms follow from the corresponding rules for exponents and will be used freely below.

8C. Derivatives of Exponential Functions (§8.3)

In this section we introduce the number e, and differentiate the functions e^x and $\ln x$. The next lemma uses the **geometric series formula**

$$(1 + b + b^2 + \cdots + b^n) = \frac{b^{n+1} - 1}{b - 1}, \quad b \neq 1,$$

which is easily proved by multiplying both sides by $b - 1$.

Lemma 8.9. *The limit* $\lim_{x \to \infty} \left(1 + \frac{1}{x}\right)^x$ *exists.*

Proof. Let b be a real number greater than 1. The function $y = b^t$ is continuous, so it has an integral $c = \int_0^1 b^t \, dt$. y^t is positive for all t, so $c > 0$. Let H be positive infinite. We will prove that

$$\left(1 + \frac{1}{H}\right)^H \approx b^{c/(b-1)}.$$

We work with the logarithm

$$\log_b \left[\left(1 + \frac{1}{H}\right)^H \right] = H \log_b \left(1 + \frac{1}{H}\right).$$

Let

$$\triangle t = \log_b \left(1 + \frac{1}{H}\right).$$

$\triangle t$ is positive infinitesimal, because

$$\triangle t \approx \log_b 1 = 0.$$

Moreover,

$$b^{\triangle t} = 1 + \frac{1}{H}, \quad H = \frac{1}{b^{\triangle t} - 1},$$

$$H \log_b \left(1 + \frac{1}{H}\right) = \frac{\triangle t}{b^{\triangle t} - 1}.$$

We wish to estimate the Riemann sum $\sum_0^1 b^t \triangle t$. Any real solution of

$$\triangle u > 0, \quad n \in \mathbb{Z}, \quad n \triangle u < 1 \leq (n+1) \triangle u \tag{53}$$

is a solution of

$$\left(1 + b^{\triangle u} + \cdots + b^{(n-1)\triangle u}\right)\triangle u \le \sum_0^1 b^u \triangle u$$

$$\le \left(1 + b^{\triangle u} + \cdots + b^{n\triangle u}\right)\triangle u.$$

By the geometric series formula, this simplifies to

$$\frac{b^{n\triangle u} - 1}{b^{\triangle u} - 1}\triangle u \le \sum_0^1 b^u \triangle u \le \frac{b^{(n+1)\triangle u} - 1}{b^{\triangle u} - 1}\triangle u. \tag{54}$$

Let K be the hyperinteger with $K\triangle t < 1 < (K+1)\triangle t$. By Transfer,

$$\frac{b^{K\triangle t} - 1}{b^{\triangle t} - 1}\triangle t \le \sum_0^1 b^t \triangle t \le \frac{b^{(K+1)b^{c/(b-1)}\triangle t} - 1}{b^{\triangle t} - 1}\triangle t.$$

Then

$$H \log_b\left(1 + \frac{1}{H}\right)\left(b^{K\triangle t} - 1\right) \le \sum_0^1 b^t \triangle t$$

$$\le H \log_b\left(1 + \frac{1}{H}\right)\left(b^{(K+1)\triangle t} - 1\right),$$

$$\frac{\sum_0^1 b^t \triangle t}{b^{(K+1)\triangle t} - 1} \le H \log_b\left(1 + \frac{1}{H}\right) \le \frac{\sum_0^1 b^t \triangle t}{b^{K\triangle t} - 1}.$$

Since $K\triangle t \approx 1 \approx (K+1)\triangle t$, we conclude that

$$\frac{c}{b-1} \approx H \log_b\left(1 + \frac{1}{H}\right), \quad c = \int_0^1 b^t \, dt$$

and

$$b^{c/(b-1)} \approx \left(1 + \frac{1}{H}\right)^H.$$

Since this holds for all positive infinite H,

$$b^{c/(b-1)} = \lim_{x\to\infty}\left(1 + \frac{1}{x}\right)^x.$$

⊣

Definition 8.10.

$$e = \lim_{x\to\infty}\left(1 + \frac{1}{x}\right)^x, \quad \ln x = \log_e x.$$

Theorem 8.11. *The exponential function e^x is differentiable, and*

$$\frac{d\,(e^x)}{dx} = e^x.$$

PROOF. Let t be finite and $\triangle t$ be positive infinitesimal. We show that

$$\frac{e^{t+\triangle t}-e^t}{\triangle t} \approx e^t.$$

We have

$$\frac{e^{t+\triangle t}-e^t}{\triangle t} = e^t\left(\frac{e^{\triangle t}-1}{\triangle t}\right).$$

Let $b = \frac{e^{\triangle t}-1}{\triangle t}$. Then $e^{\triangle t} = 1 + b\triangle t$. Since e^t is continuous, $e^{\triangle t} \approx 1$, and $b\triangle t$ is positive infinitesimal. Thus $H = 1/(b\triangle t)$ is positive infinite. Also,

$$e \approx \left(1+\frac{1}{H}\right)^H = (1+b\triangle t)^{1/b\triangle t} = \left(e^{\triangle t}\right)^{1/b\triangle t} = e^{1/b}.$$

Therefore $b \approx 1$, and

$$\frac{e^{t+\triangle t}-e^t}{\triangle t} = be^t \approx e^t. \tag{55}$$

Now let x be real and $\triangle t$ be positive infinitesimal. Then by (55),

$$\frac{e^{x+\triangle t}-e^x}{\triangle t} \approx e^x, \quad \frac{e^x-e^{x-\triangle t}}{\triangle t} \approx e^{x-\triangle t} \approx e^x.$$

Therefore

$$\frac{d\left(e^x\right)}{dx} = e^x.$$

⊣

In general, $a^x = e^{x \ln a}$, so

$$\frac{d\left(a^x\right)}{dx} = \frac{d\left(e^{x \ln a}\right)}{dx} = (\ln a)e^{x \ln a} = (\ln a)a^x.$$

This shows that e is uniquely characterized by the equation $d\left(e^x\right)/dx = e^x$, because when $a \neq e$, $\ln a \neq 1$.

COROLLARY 8.12. *$y = \ln x$ is differentiable for all $x > 0$, and*

$$\frac{d(\ln x)}{dx} = \frac{1}{x}.$$

PROOF. Since $y = \ln x$, $x = e^y$. By the Inverse Function Theorem 7.3, dy/dx exists and

$$\frac{dy}{dx} = \frac{1}{dx/dy} = \frac{1}{e^y} = \frac{1}{x}.$$

⊣

CHAPTER 9

INFINITE SERIES

Throughout this chapter we let H and K denote positive infinite hyperintegers. The set of natural numbers (nonnegative integers) will be denoted by $\mathbb{N}$, and the set of positive integers by $\mathbb{N}^+$.

9A. Sequences (§9.1)

By an **infinite sequence** $\langle a_n \rangle$ we mean a function from either $\mathbb{N}$ or $\mathbb{N}^+$ into the real numbers.

DEFINITION 9.1. *An infinite sequence $\langle a_n \rangle$* **converges** *to a real number L if $a_H \approx L$ for every H, in symbols,* $\lim_{n\to\infty} a_n = L$. *$\langle a_n \rangle$* **diverges to** ∞ *if a_H is positive infinite for all H, in symbols,* $\lim_{n\to\infty} a_n = \infty$.

We begin with some computations of limits.

THEOREM 9.2. *The following sequences diverge to ∞.*

$$\lim_{n\to\infty} \frac{n!}{b^n} = \infty, \quad (b \geq 1)$$
$$\lim_{n\to\infty} \frac{b^n}{n^c} = \infty, \quad (b > 1, c \geq 0)$$
$$\lim_{n\to\infty} \frac{n^c}{\ln(n)} = \infty, \quad (c > 0)$$
$$\lim_{n\to\infty} \ln(n) = \infty.$$

PROOF. We show that each of

$$\frac{H!}{b^H}, \quad \frac{b^H}{H^c}, \quad \frac{H^c}{\ln(H)}, \quad \ln(H)$$

is positive infinite. The natural logarithm $\ln(H)$ is positive infinite because for each real r, $e^r < H$ and hence $r < \ln(H)$. In the other three cases we show that the logarithm of the quotient is positive infinite.

$H!/b^H$: For an integer $m > b$ we have

$$\ln\left(\frac{H!}{b^H}\right) = \ln(1) + \cdots + \ln(m-1) + \ln(m) + \cdots + \ln(H) - H\ln(b)$$

$$> (H - m)\ln(m) - H\ln(b) = H(\ln m - \ln b) - m\ln m.$$

Since $\ln m > \ln b$, $\ln\left(H!/b^H\right)$ is positive infinite.

b^H/H^c: We have

$$\ln\left(\frac{b^H}{H^c}\right) = H\ln b - c\ln H = H\left(\ln(b) - c\frac{\ln(H)}{H}\right).$$

By l'Hospital's Rule, $\lim_{x\to\infty}(\ln(x)/x) = 0$, so $\ln(H)/H \approx 0$ and $\ln\left(b^H/H^c\right)$ is positive infinite.

$H^c/\ln H$: Putting $K = \ln(H)$,

$$\ln\left(\frac{H^c}{\ln(H)}\right) = c\ln(H) - \ln(\ln(H)) = K\left(c - \frac{\ln(K)}{K}\right),$$

so $\ln\left(H^c/\ln(H)\right)$ is positive infinite. ⊣

Here are three equivalence theorems for limits of sequences.

Theorem 9.3. *Given a sequence $\langle a_n\rangle$ and a real number L, the following are equivalent.*

(i) $\lim_{n\to\infty} a_n = L$.

(ii) There is an H such that for all $K \ge H$, $a_K \approx L$.

(iii) The ε, M condition: For every real $\varepsilon > 0$ there is a positive integer M such that for all $n \ge M$, $|a_n - L| < \varepsilon$.

The proof is similar to Theorem 5.1, the equivalence theorem for limits of functions.

Theorem 9.4. *The following are equivalent.*

(i) $\lim_{n\to\infty} a_n = \infty$.

(ii) There is an H such that for all $K \ge H$, a_K is positive infinite.

(iii) For every real B there is a positive integer M such that for all $n \ge M$, $a_n > B$.

Theorem 9.5. *The following are equivalent.*

(i) The sequence $\langle a_n\rangle$ converges.

(ii) Hyperreal Cauchy Condition: For all H and K, $a_H \approx a_K$.

(iii) Real Cauchy Condition: For every real $\varepsilon > 0$ there is a positive integer M such that for all $m, n \ge M$, $|a_m - a_n| < \varepsilon$.

Proof. Assume (i), say $\lim_{n\to\infty} a_n = L$. Then for all H and K, $a_H \approx L \approx a_K$, so (ii) holds. Now assume (ii). Suppose the Real Cauchy Condition (iii) fails for some real $\varepsilon > 0$. Then every $M \in \mathbb{N}^+$ is a partial real solution of

$$m \in \mathbb{N}^+, \quad n \in \mathbb{N}^+, \quad M \le m, \quad M \le n, \quad |a_m - a_n| \ge \varepsilon.$$

Let J be positive infinite. By the Partial Solution Theorem 1.20 there exist $H, K \ge J$ such that $|a_H = a_K| \ge \varepsilon$, contradicting (ii).

Assume the Real Cauchy Condition (iii). There is a positive integer M_1 such that for all integers $m, n \ge M_1$, $|a_m - a_n| < 1$. By Transfer. for all integers $m \ge M_1$, $|a_m - a_H| < 1$. Therefore a_H is finite. Let $L = st\,(a_H)$. We show that the sequence converges to L. Let $\varepsilon > 0$ be real and let M be the corresponding positive integer. Using Transfer again, for all hyperreal $n \geq M$ we have $|a_n - a_H| < \varepsilon$. Thus for all K, $|a_K - a_H| < \varepsilon$. Since this holds for each ε, we have $a_K \approx a_H \approx L$ for every K, whence $\lim_{n\to\infty} a_n = L$, ⊣

We now give hyperreal proofs of the Bolzano-Weierstrass Theorem and the countable Heine-Borel Theorem. A **subsequence** of $\langle a_n \rangle$ is a sequence $\langle a_{f(n)} \rangle$ where f is an increasing function from $\mathbb{N}$ into $\mathbb{N}$.

LEMMA 9.6. *(i) If* $\operatorname{st}(H) = L$ *then* $\langle a_n \rangle$ *has a subsequence converging to* L. *(ii) If* a_H *is positive infinite then* $\langle a_n \rangle$ *has a subsequence diverging to* ∞.

PROOF. (i) For each $n \in \mathbb{N}^+$ and each H we have

$$H \in \mathbb{N}^*, \quad H \geq n, \quad |a_H - L| \leq \frac{1}{n}.$$

By the Partial Solution Theorem, each $n \in \mathbb{N}^+$ is a partial real solution of

$$m \in \mathbb{N}^+, \quad m \geq n, \quad |a_m - L| < \frac{1}{n}.$$

Define $f(0) = 0$, and for each $n > 0$ define $f(n)$ to be the first positive integer m such that

$$m > f(n-1), \quad |a_m - L| < \frac{1}{n}.$$

Then f is increasing and

$$|a_{f(n)} - L| < \frac{1}{n}.$$

By the ε, δ condition for limits (Theorem 5.1), the subsequence $\langle a_{f(n)} \rangle$ converges to L.

(ii) The proof is similar to (i). ⊣

$\langle a_n \rangle$ is a **bounded sequence** if its range $\{a_n : n \in \mathbb{N}\}$ is a bounded set. Thus by Theorem 1.31, $\langle a_n \rangle$ if bounded if and only if a_H is finite for every $H \in \mathbb{N}^*$.

THEOREM 9.7. *(Bolzano-Weierstrass Theorem) Every bounded sequence has a convergent subsequence.*

PROOF. Let $\langle a_n \rangle$ be bounded and choose an infinite H. Then a_H is finite, and by Lemma 9.6 $\langle a_n \rangle$ has a subsequence converging to $\operatorname{st}(a_H)$. ⊣

THEOREM 9.8. *(Countable Heine-Borel Theorem) Let*

$$X_0 \supseteq X_1 \supseteq \cdots \supseteq X_n \supseteq \cdots$$

be a decreasing chain of nonempty closed bounded sets of real numbers. Then

$$\bigcap_{n=0}^{\infty} X_n \neq \emptyset.$$

Proof. Let $P(x,n)$ be the binary relation $x \in X_n$ on the reals. Then for each natural number n,

$$X_n^* = \{x \in \mathbb{R}^* \colon P^*(x,n)\}.$$

Choose an infinite H and define

$$X_H^* = \{x \in \mathbb{R}^* \colon P^*(x,H)\}.$$

By hypothesis, each X_n is nonempty, so each real solution of $n \in \mathbb{N}$ is a partial solution of

$$n \in \mathbb{N}, \quad P(x,n).$$

By the Partial Solution Theorem, there is a hyperreal x such that $P^*(x,H)$, so $x \in X_H^*$. Using Transfer we see that $X_H^* \subseteq X_n^*$ for all $n \leq H$, and hence for all $n \in \mathbb{N}$. Therefore $x \in X_n^*$ for all $n \in \mathbb{N}$. Since X_n is closed and bounded, x is finite and $\mathrm{st}(x) \in X_n$. Thus $\mathrm{st}(x) \in \bigcap_{n=0}^{\infty} X_n$. ⊣

Theorem 9.9. *An increasing sequence $\langle a_n \rangle$ either diverges to ∞ or converges.*

Proof. Suppose the sequence does not diverge to ∞. Then some a_H is not positive infinite. Since $\langle a_n \rangle$ is increasing, it follows from the Transfer Axiom that the natural extension of $\langle a_n \rangle$ is increasing, that is, $a_m < a_n$ whenever $m < n$ in $\mathbb{N}^*$. Then $a_H > a_0$, so a_H is not negative infinite. Therefore a_H is finite. By Lemma 9.6, $\langle a_n \rangle$ has a subsequence $\langle a_{f(n)} \rangle$ which converges to $L = \mathrm{st}(a_H)$. For every natural number $n \geq f(0)$ there is an m such that $f(m) \leq n < f(m+1)$. Take an infinite K. By the Partial Solution Theorem there is an $M \in \mathbb{N}^*$ such that $f(M) \leq K \leq f(M+1)$. M must be positive infinite, so $a_{f(M)} \approx L \approx a_{f(M+1)}$. But since $\langle a_n \rangle$ is increasing, $a_{f(M)} \leq a_K \leq a_{f(M+1)}$. Therefore $a_K \approx L$, so $\langle a_n \rangle$ converges to L. ⊣

9B. Series (§9.2 – §9.6)

Given an infinite sequence

$$\langle a_n \rangle = a_0, a_1, \dots, a_n, \dots,$$

The **partial sum sequence** $\langle S_n \rangle$ is defined by

$$S_n = a_0 + a_1 + \cdots + a_n = \sum_{k=0}^{n} a_k.$$

By the Function Axiom, the natural extension of the function S_n has a value S_H for each positive infinite hyperinteger H, which we call the **infinite partial sum**

$$S_H = a_0 + a_1 + \cdots + a_H = \sum_{k=0}^{H} a_k.$$

Definition 9.10. *The* **infinite series** $\sum_{n=0}^{\infty} a_n$ *is said to* **converge** *to* S*, in symbols*

$$S = a_0 + a_1 + \cdots + a_n + \cdots = \sum_{n=0}^{\infty} a_n,$$

if the partial sum sequence converges to S*,*

$$S = \lim_{n\to\infty} S_n = \lim_{n\to\infty}(a_0 + a_1 + \cdots + a_n).$$

That is, for every positive infinite hyperinteger H*,*

$$S \approx a_0 + a_1 + \cdots + a_H.$$

The series is said to **diverge** *if the partial sum sequence diverges.*

Proposition 9.11. *(Geometric Series) For each* $b \in (-1, 1)$*,*

$$\sum_{n=0}^{\infty} b^n = \frac{1}{1-b}.$$

Proof. By the geometric series formula and Transfer, for each infinite H we have

$$(1 + b + b^2 + \cdots + b^H) = \frac{b^{H+1} - 1}{b - 1} \approx \frac{1}{1-b}.$$

⊣

The equivalence theorems for limits of sequences in the preceding section automatically give equivalence theorems for sums of series. In particular, the Hyperreal Cauchy Condition in Theorem 9.5 has the following consequence, where $a_H + \cdots + a_K$ means $\sum_{n=0}^{K} a_n - \sum_{n=0}^{H} a_n$.

Proposition 9.12. *(i)* $\sum_{n=0}^{\infty} a_n$ *converges if and only if for all infinite* $H \leq K$*,*

$$a_H + \cdots + a_K \approx 0.$$

(ii) If $\sum_{n=0}^{\infty} a_n$ *converges, then* $\lim_{n\to\infty} a_n = 0$.

Proof. (i) is the Hyperreal Cauchy Condition stated in series notation. (ii) follows by setting $K = H$, so that $a_H \approx 0$. ⊣

The usual converge tests for infinite series are developed in *Elementary Calculus* in the standard way. The following hyperreal form of the Limit Comparison Test is simpler to state and use than the standard result.

Theorem 9.13. *(Limit Comparison Test) Let* $\sum_{n=0}^{\infty} a_n$ *and* $\sum_{n=0}^{\infty} b_n$ *be positive term series and* c *be a positive real number. Suppose that* $a_K \leq c\, b_K$ *for all infinite* K*. If* $\sum_{n=0}^{\infty} b_n$ *converges then* $\sum_{n=0}^{\infty} a_n$ *converges.*

Proof. Let $H \le K$ be infinite. By the Hyperreal Cauchy Condition 9.12,

$$b_H + \cdots + b_K \approx 0.$$

Hence

$$0 \le a_H + \cdots + a_K \le c(b_H + \cdots + b_K) \approx 0,$$

so $a_H + \cdots + a_K \approx 0$ and $\sum_{n=0}^{\infty} a_n$ converges. ⊣

As an illustration of the use of the Limit Comparison Test we give a proof of a basic result on power series. A series $\sum_{n=0}^{\infty} a_n$ is **absolutely convergent** if the series $\sum_{n=0}^{\infty} |a_n|$ is convergent.

Theorem 9.14. *If a power series $\sum_{n=0}^{\infty} a_n\, x^n$ is convergent when $x = u$, then it is absolutely convergent whenever $|x| < |u|$.*

Proof. Let $|v| < |u|$ and $b = |v|/|u|$. Then $0 \le b < 1$, so the geometric series $\sum_{n=0}^{\infty} b^n$ converges. We assume that $\sum_{n=0}^{\infty} a_n\, u^n$ converges, so $\lim_{n\to\infty} a_n u^n = 0$ by Proposition 9.12. Then for positive infinite H, $a_H\, u^H \approx 0$. Thus

$$\left|a_H\, v^H\right| = \left|a_H\, u^H\right| b^H \le b^H,$$

so by the Limit Comparison Test 9.13, $\sum_{n=0}^{\infty} |a_n\, v^n|$ converges. ⊣

9C. Taylor's Formula and Higher Differentials (§9.10)

In *Elementary Calculus* the standard proof of Taylor's Formula is given. Let f be a real function and $c, x \in \mathbb{R}$.

Theorem 9.15. *(Taylor's Formula) If $f^{(n+1)}$ exists between c and x, then*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(c)}{k!}\,(x-c)^k + \frac{f^{(n+1)}(t)}{(n+1)!}\,(x-c)^{n+1}$$

for some real t between c and x.

Using the Partial Solution Theorem we obtain the following consequence concerning infinitesimals.

Corollary 9.16. *Suppose c is real and $f^{(n)}$ is continuous at c. Then whenever $x \approx c$ and $\triangle x$ is a nonzero infinitesimal,*

$$f(x + \triangle x) \approx \sum_{k=0}^{n} \frac{f^{(k)}(x)}{k!}\triangle x^k \ (\textit{compared to } \triangle x^n)\,.$$

PROOF. $f^{n)}(t)$ is defined for $t \approx c$. By Theorem 9.15 and the Partial Solution Theorem there is a t between x and $x + \triangle x$ such that

$$\begin{aligned} f(x+\triangle x) &= \sum_{k=0}^{n-1} \frac{f^{(k)}(x)}{k!}\triangle x^k + \frac{f^{(n)}(t)}{n!}\triangle x^n \\ &= \sum_{k=0}^{n} \frac{f^{(k)}(x)}{k!}\triangle x^k + \frac{f^{(n)}(t) - f^{(n)}(x)}{n!}\triangle x^n. \end{aligned}$$

By continuity of $f^{(n)}$ at c, $f^{(n)}(t) - f^{(n)}(x) \approx 0$, and the result follows. ⊣

By modifying the proof of Taylor's Formula, we show that Corollary 9.16 holds at $x = c$ assuming only that $f^{(n)}(c)$ exists.

THEOREM 9.17. *Suppose c is real, $\triangle x$ is nonzero infinitesimal, and $f^{(n)}(c)$ exists. Then*

$$f(c+\triangle x) \approx \sum_{k=0}^{n} \frac{f^{(k)}(c)}{k!}\triangle x^k \ (\textit{compared to } \triangle x^n).$$

PROOF. Since $f^{(n)}(c)$ exists, $f^{(n-1)}(x)$ exists for all x in some real neighborhood of c. Let

$$\begin{aligned} F(x) &= f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(c)}{k!}(x-c)^k, \\ G(x) &= (x-c)^n. \end{aligned}$$

Then for $m < n$, $F^{(m)}(c) = 0$ and $G^{(m)}(c) = 0$. Using the Generalized Mean Value Theorem 5.11 $n-1$ times we see that

$$\frac{F(x)}{G(x)} = \frac{F^{(n-1)}(t)}{G^{(n-1)}(t)}$$

for some t between c and x. Note that the $(n-1)$-st derivative of $(t-c)^{(n-1)}$ is $(n-1)!$, and the $(n-1)$-st derivative of $(t-c)^n$ is $n!(t-c)$. From this, we make the computations

$$F^{(n-1)}(t) = f^{(n-1)}(t) - f^{(n-1)}(c) - f^{(n)}(c)(t-c),$$

$$G^{(n-1)}(t) = n!(t-c),$$

and we have in turn

$$\begin{aligned} \frac{F(x)}{G(x)} &= \frac{f^{(n-1)}(t) - f^{(n-1)}(c)}{n!(t-c)} - \frac{f^{(n)}(c)}{n!}, \\ F(x) &= \left(\frac{f^{(n-1)}(t) - f^{(n-1)}(c)}{n!(t-c)} - \frac{f^{(n)}(c)}{n!}\right)(x-c)^n, \\ f(x) &= \sum_{k=0}^{n-1} \frac{f^{(k)}(c)}{k!}(x-c)^k + \frac{f^{(n)}(c)}{n!}(x-c)^n + F(x), \end{aligned}$$

$$f(x)=\sum_{k=0}^{n-1}\frac{f^{(k)}(c)}{k!}(x-c)^k+\frac{f^{(n-1)}(t)-f^{(n-1)}(c)}{n!(t-c)}(x-c)^n.$$

By the Partial Solution Theorem there is a hyperreal t_1 between c and $c+\triangle x$ such that

$$f(c+\triangle x)=\sum_{k=0}^{n-1}\frac{f^{k)}(c)}{k!}\triangle x^k+\frac{f^{(n-1)}(t_1)-f^{(n-1)}(c)}{n!(t_1-c)}\triangle x^n.$$

Since $f^{(n)}(c)$ exists, we have

$$f^{(n)}(c)\approx\frac{f^{(n-1)}(t_1)-f^{(n-1)}(c)}{t_1-c}.$$

Therefore

$$f(c+\triangle x)=\sum_{k=0}^{n}\frac{f^{(k)}(c)}{k!}\triangle x^k+\varepsilon\triangle x^n$$

for some $\varepsilon\approx 0$, and the desired result follows. ⊣

When $n=1$, Theorem 9.17 reduces to the Increment Theorem,

$$f(c)+\triangle x)\approx f(c)+f'(c)\triangle x \text{ (compared to } \triangle x).$$

Another natural way to generalize the Increment Theorem involves the notion of an n-th increment.

DEFINITION 9.18. *Let $y=f(x)$ be defined on an interval I. The* **n-th increment**

$$\triangle^n y=\triangle^n f(x,\triangle x)$$

is defined by induction on n as follows.

$$\triangle' y=\triangle' f(x,\triangle x)=f(x+\triangle x)-f(x),$$
$$\triangle^{n+1}y=\triangle^{n+1}f(x,x+\triangle x)=\triangle^n f(x+\triangle,\triangle x)-\triangle^n f(x,\triangle x).$$

Thus $\triangle^n f(x,\triangle x)$ is defined whenever $x\in I$ and $x+n\triangle x\in I$. For $n=2$ we have

$$\begin{aligned}\triangle^2 y&=[f(x+2\triangle x)-f(x+\triangle x)]-[f(x+\triangle x)-f(x)]\\&=f(x+2\triangle x)-2f(x+\triangle x)+f(x).\end{aligned}$$

For $n=3$,

$$\triangle^3 y=f(x+3\triangle x)-3f(x+2\triangle x)+3f(x+\triangle x)-f(x).$$

In general, it can be shown by induction that

$$\triangle^n y=\sum_{k=0}^{n}(-1)^{n-k}\binom{n}{k}f(x+k\triangle x)$$

where $\binom{n}{k}=\frac{n!}{k!(n-k)!}$ is the binomial coefficient. The following theorem gives a connection between the n-th increment and the n-th differential $d^n y=f^{(n)}(x)dx^n$.

Theorem 9.19. *Suppose $y = f(x)$ is a real function, c is a real number, and $\triangle x$ is a nonzero infinitesimal.*

(i) If $f^{(n)}$ is continuous at c, then at every hyperreal point $x \approx c$ we have

$$\frac{\triangle^n y}{\triangle x^n} \approx \frac{d^n y}{dx^n},$$

that is,

$$\triangle^n f(x, \triangle x) \approx f^{(n)}(x)\triangle x^n \text{ (compared to } \triangle x^n).$$

(ii) If $f^{(n)}$ exists at c, then at $x = c$ we have

$$\frac{\triangle^n y}{\triangle x^n} \approx \frac{d^n y}{dx^n},$$

that is,

$$\triangle^n f(c, \triangle x) \approx f^{(n)}(c)\triangle x^n \text{ (compared to } \triangle x^n).$$

Proof. We give the proof for the case $\triangle x > 0$. We first prove a generalization of the Mean Value Theorem 3.30.

(a) Let u and $\triangle u$ be real and suppose $g^{(m)}(t)$ exists for $u \le t \le u + m\triangle u$. Then $\triangle^m g(u, \triangle u) = g^{(m)}(t_1)\triangle u^m$ for some $t_1 \in (u, u + m\triangle u)$.

The proof is by induction on m. For $m = 1$ it is just the Mean Value Theorem,

$$g(u + \triangle u) - g(u) = g'(t)\triangle u.$$

Assume (a) for $m - 1$ and let $h(t) = g(t + \triangle u) - g(t)$. Then

$$h^{(m-1)}(t) = g^{(m-1)}(t + \triangle u) - g^{(m-1)}(t)$$

exists for $u \le t \le u + (m-1)\triangle u$, so for some $t_0 \in (u, u + (m-1)\triangle u)$,

$$\triangle^{m-1} h(u, \triangle u) = h^{(m-1)}(t_0)\triangle u^{m-1}.$$

Moreover,

$$\triangle^{m-1} h(u, \triangle u) = \triangle^m g(u, \triangle u).$$

By the Mean Value Theorem,

$$h^{(m-1)}(t_0) = g^{(m-1)}(t_0 + \triangle u) - g^{(m-1)}(t_0) = g^{(m)}(t_1)\triangle u$$

for some $t_1 \in (t_0, t_0 + \triangle u)$, and

$$\triangle^m g(u, \triangle u) = g^{(m)}(t_1)\triangle u.$$

This proves (a) for m, and completes the induction.

We now prove (i). Since $f^{(n)}$ is continuous at c, $f^{(n)}(t)$ exists for $x \le t \le x + \triangle x$. By (a) and the Partial Solution Theorem,

$$\frac{\triangle^n f(x, \triangle x)}{\triangle x^n} = f^{(n)}(t_1)$$

for some $t_1 \in (x, x + n\triangle x)$. Then $t_1 \approx x \approx c$, and since $f^{(n)}$ is continuous at c,

$$\frac{\triangle^n f(x,\triangle x)}{\triangle x^n} \approx f^{(n)}(x).$$

To prove (ii) we let $m = n - 1$ and $g(t) = f(t + \triangle x) - f(t)$. By (a) and the Partial Solution Theorem there is a $t_1 \in (c, c + (n-1)\triangle x)$ such that

$$\triangle^{n-1} g(c, \triangle x) = g^{(n-1)}(t_1)\triangle x^{n-1},$$

that is,

$$\triangle^n f(c.\triangle x) = \frac{f^{(n-1)}(t_1) + \triangle x) - f^{(n-1)}(t_1)}{\triangle x}\triangle x^n.$$

Let $t_2 = t_1 + \triangle x$. Then $t_1 \approx c$, $t_2 \approx c$, so

$$\begin{aligned}
\frac{\triangle^n f(c,\triangle x)}{\triangle x^n} &= \frac{f^{(n-1)}(t_2) - f^{(n-1)}(t_1)}{\triangle x} \\
&= \frac{f^{(n-1)}(t_2) - f^{(n-1)}(c)}{t_2 - c}\,\frac{t_2 - c}{\triangle x} + \frac{f^{(n-1)}(c) - f^{(n-1)}(t_1)}{c - t_1}\,\frac{c - t_1}{\triangle x} \\
&\approx f^{(n)}(c)\, st\left(\frac{t_2 - c}{\triangle x}\right) + f^{(n)}(c)\, st\left(\frac{c - t_1}{\triangle x}\right) = f^{(n)}(c).
\end{aligned}$$

Thus

$$\frac{\triangle^n f(c,\triangle x)}{\triangle x^n} \approx f^{(n)}(c).$$

$\dashv$

CHAPTER 10

VECTORS

10A. Hyperreal Vectors (§10.8)

Chapter 10 of *Elementary Calculus* deals mostly with two and three dimensional vectors over the reals. Only the last section of the chapter, §10.8, concerns hyperreal vectors.

Let n be a fixed natural number. The simplest way to define a real vector in n dimensions is as an n-tuple of real numbers, $\langle r_1, \ldots, r_n\rangle$. In *Elementary Calculus* we used the following more geometric definition. Given two points

$$P = \langle p_1, \ldots, p_n\rangle, \quad Q = \langle q_1, \ldots, q_n\rangle$$

in $\mathbb{R}^n$, the ordered pair $\overrightarrow{PQ}$ is called the **directed line segment** from P to Q. The **components** of $\overrightarrow{PQ}$ are the terms in the n-tuple

$$\langle q_1 - p_1, \ldots, q_n - p_n\rangle.$$

Two directed line segments are said to be **equivalent** if they have the same components (hence the same length and direction), and a **vector $\mathbf{A}$** is an equivalence class of directed line segments. If $\mathbf{A}$ is the equivalence class of $\overrightarrow{PQ}$, the components of $\overrightarrow{PQ}$ are called the components of $\mathbf{A}$, denoted by $\langle a_1, \ldots, a_n\rangle$, and $\mathbf{A}$ is called the **vector from P to Q**.

The **basis vectors** are the vectors $\mathbf{j}_1, \ldots, \mathbf{j}_n$, where $\mathbf{j}_m$ is the vector with m-th component 1 and all other components 0. For real vectors the **vector sum $\mathbf{A} + \mathbf{B}$**, **vector difference $\mathbf{A} - \mathbf{B}$**, and the **scalar multiple $c\mathbf{A}$** are defined in the usual way in terms of components. Thus the vector $\mathbf{A}$ with components $\langle a_1, \ldots, a_n\rangle$ is equal to the sum

$$\mathbf{A} = a_1\mathbf{j}_1 + \cdots + a_n\mathbf{j}_n.$$

The **inner product** and **length** are defined by

$$\mathbf{A} \cdot \mathbf{B} = a_1 b_1 + \cdots + a_n b_n,$$

$$|\mathbf{A}| = \sqrt{\mathbf{A} \cdot \mathbf{A}} = \sqrt{a_1^2 + \cdots + a_n^2}.$$

The **zero vector** in n dimensions is denoted by $\mathbf{0}$.

In two dimensions, the basis vectors are denoted by $\mathbf{i}, \mathbf{j}$, so the vector $\mathbf{A}$ with components $\langle a, b\rangle$ is the sum $\mathbf{A} = a\mathbf{i} + b\mathbf{j}$.

The **hyperreal vectors** in n dimensions are defined in a similar way. By the Function and Transfer Axioms, the hyperreal vectors in n dimensions form an inner product space over the field $\mathbb{R}^*$ of hyperreal numbers. That is, the usual algebraic rules for vector sums and differences, scalar multiples, and inner products hold for hyperreal vectors. Hyperreal vectors $\mathbf{A}$ will be classified by the behavior of the length $|\mathbf{A}|$ and the unit vector $\frac{1}{|\mathbf{A}|}\mathbf{A}$.

DEFINITION 10.1. *A hyperreal vector* $\mathbf{A}$ *is said to be* **infinitesimal**, **finite**, *or* **infinite** *if its length* $|\mathbf{A}|$ *is infinitesimal, finite, or infinite respectively.* $\mathbf{A}$ *is* **infinitely close** *to* $\mathbf{B}$, *in symbols* $\mathbf{A} \approx \mathbf{B}$, *if and only if* $\mathbf{B}-\mathbf{A}$ *is infinitesimal. The* **monad** *and* **galaxy** *of* $\mathbf{A}$ *are defined by*

$$\operatorname{monad}(\mathbf{A}) = \{\mathbf{B}\colon \mathbf{B} \approx \mathbf{A}\},$$

$$\operatorname{galaxy}(\mathbf{A}) = \{\mathbf{B}\colon \mathbf{B} - \mathbf{A} \textit{ is finite}\}.$$

Thus $\operatorname{monad}(\mathbf{0})$ *is the set of infinitesimal vectors, and* $\operatorname{galaxy}(\mathbf{0})$ *is the set of finite vectors.*

The next proposition follows easily from the inequalities

$$|a_m| \le \sqrt{a_1^2 + \cdots a_m^2} \le |a_1| + \cdots + |a_m|.$$

PROPOSITION 10.2. *(i)* $\mathbf{A}$ *is infinitesimal if and only if all its components are infinitesimal.*
(ii) $\mathbf{A}$ *is finite if and only if all its components are finite.*
(iii) $\mathbf{A}$ *is infinite if and only if at least one of its components is infinite.*
(iv) $\mathbf{A} \approx \mathbf{B}$ *if and only if* $a_1 \approx b_1, \ldots, a_n \approx b_n$.

PROPOSITION 10.3. *The sets* $\operatorname{monad}(\mathbf{0})$ *and* $\operatorname{galaxy}(\mathbf{0})$ *are closed under vector sums, vector differences, and finite scalar multiples.*

This follows from Proposition 10.2 and the fact that the sets monad(0) and galaxy(0) of hyperreal numbers are closed under sums, differences, and finite multiples.

DEFINITION 10.4. *Given a finite hyperreal vector* $\mathbf{A}$, *the* **standard part** $\operatorname{st}(\mathbf{A})$ *is defined as the real vector*

$$\operatorname{st}(\mathbf{A}) = \operatorname{st}(a_1)\mathbf{j}_1 + \cdots + \operatorname{st}(a_n)\mathbf{j}_n.$$

Thus $\operatorname{st}(\mathbf{A})$ *is the unique real vector infinitely close to* $\mathbf{A}$.

The next result also follows from Proposition 10.2.

Proposition 10.5. *The mapping st preserves sums, differences, finite scalar multiples, and lengths. That is, if* $\mathbf{A}, \mathbf{B}$ *are finite vectors and* c *is a finite scalar,*

$$\begin{aligned}
\mathrm{st}(\mathbf{A} + \mathbf{B}) &= \mathrm{st}(\mathbf{A}) + \mathrm{st}(\mathbf{B})\\
\mathrm{st}(\mathbf{A} - \mathbf{B}) &= \mathrm{st}(\mathbf{A}) - \mathrm{st}(\mathbf{B})\\
\mathrm{st}(c\mathbf{A}) &= \mathrm{st}(c)\mathrm{st}(\mathbf{A})\\
\mathrm{st}(\mathbf{A} \cdot \mathbf{B}) &= \mathrm{st}(\mathbf{A}) \cdot \mathrm{st}(\mathbf{B})\\
\mathrm{st}(|\mathbf{A}|) &= |\mathrm{st}(\mathbf{A})|.
\end{aligned}$$

Definition 10.6. *A hyperreal vector* $\mathbf{A}$ *has* **real length** *if* $|\mathbf{A}|$ *is real. A* **unit vector** *is a hyperreal vector of length* 1*. If* $\mathbf{A} \neq \mathbf{0}$*, the* **unit vector of** $\mathbf{A}$ *is the unit vector* $\mathbf{U} = \frac{1}{|\mathbf{A}|}\mathbf{A}$*.* $\mathbf{A}$ *has* **real direction** *if* $\mathbf{A} \neq \mathbf{0}$ *and the unit vector of* $\mathbf{A}$ *is a real vector.*

Proposition 10.7. *A nonzero vector* $\mathbf{A}$ *is real if and only if* $\mathbf{A}$ *has both real length and real direction.*

Proof. $\mathbf{A}$ has real length and direction if and only if both $|\mathbf{A}|$ and $\mathbf{U} = \frac{1}{|\mathbf{A}|}\mathbf{A}$ are real, which holds if and only if $\mathbf{A} = |\mathbf{A}|\mathbf{U}$ is real. $\dashv$

Definition 10.8. *Two nonzero vectors* $\mathbf{A}$ *and* $\mathbf{B}$ *are said to be* **parallel** *if their unit vectors* $\mathbf{U}$ *and* $\mathbf{V}$ *are equal or opposite,* $\mathbf{U} = \mathbf{V}$ *or* $\mathbf{U} = -\mathbf{V}$*.*

We introduce a weaker notion for hyperreal vectors. Two nonzero hyperreal vectors $\mathbf{A}$ *and* $\mathbf{B}$ *are said to be* **almost parallel** *if their unit vectors* $\mathbf{U}$ *and* $\mathbf{V}$ *are such that either* $\mathbf{U} \approx \mathbf{V}$ *or* $\mathbf{U} \approx -\mathbf{V}$*.*

Proposition 10.9. *Every nonzero hyperreal vector* $\mathbf{A}$ *is almost parallel to some real vector.*

Proof. Let $\mathbf{U}$ be the unit vector of $\mathbf{A}$. Then $\mathbf{U}$ has finite length 1, so the real vector $\mathbf{V} = \mathrm{st}(\mathbf{U})$ exists. $\mathbf{V}$ is its own unit vector because $|\mathbf{V}| = |\mathrm{st}(\mathbf{U})| = \mathrm{st}(|\mathbf{U}|) = 1$. Finally, $\mathbf{U} \approx \mathbf{V}$, so $\mathbf{A}$ is almost parallel to $\mathbf{V}$. $\dashv$

The notion of almost parallel can be generalized to finite sequences of vectors.

Definition 10.10. *A* k*-tuple of real vectors* $\mathbf{A}_1, \dots, \mathbf{A}_k$ *is* **linearly dependent over** $\mathbb{R}$ *if there exist real numbers* $c_1, \dots, c_k$*, not all zero, such that*

$$c_1\mathbf{A}_1 + \cdots + c_k\mathbf{A}_k = \mathbf{0}. \tag{56}$$

Similarly, a k*-tuple* $\mathbf{A}_1, \dots, \mathbf{A}_k$ *of hyperreal vectors is* **linearly dependent over** $\mathbb{R}^*$ *if there exist hyperreal numbers* $c_1, \dots, c_k$*, not all zero, such that such that (56) holds. Finally, a* k*-tuple of hyperreal vectors* $\mathbf{A}_1, \dots, \mathbf{A}_k$ *is* **almost linearly dependent over** $\mathbb{R}$ *if either some* $\mathbf{A}_m = \mathbf{0}$ *or there exist real numbers* $c_1, \dots, c_k$*, not all zero, such that*

$$c_1\mathbf{U}_1 + \cdots + c_k\mathbf{U}_k \approx \mathbf{0},$$

where $\mathbf{U}_m$ *is the unit vector of* $\mathbf{A}_m$*.*

Thus nonzero hyperreal vectors $\mathbf{A}_1, \dots, \mathbf{A}_k$ are almost linearly dependent over $\mathbb{R}$ if and only if the standard parts of the unit vectors, $\mathrm{st}(\mathbf{U}_1), \dots, \mathrm{st}(\mathbf{U}_k)$, are linearly dependent over $\mathbb{R}$. For pairs of nonzero vectors, linearly dependent means parallel, and almost linearly dependent means almost parallel.

Theorem 10.11. *(i) If $\mathbf{A}_1, \dots, \mathbf{A}_k$ are linearly dependent over $\mathbb{R}^*$ then they are almost linearly dependent over $\mathbb{R}$.*

(ii) For real vectors, linear dependence over $\mathbb{R}$, linear dependence over $\mathbb{R}^*$, and almost linear dependence over $\mathbb{R}$ are equivalent.

Proof. (i) Let $c_1\mathbf{A}_1 + \cdots + c_k\mathbf{A}_k = \mathbf{0}$ where the hyperreal numbers c_k are not all zero. We may assume that all the $\mathbf{A}_m$ are nonzero since otherwise the vectors are trivially almost linearly dependent over $\mathbb{R}$. Then

$$c_1|\mathbf{A}_1|\mathbf{U}_1 + \cdots + c_k|\mathbf{A}_k|\mathbf{U}_k = \mathbf{0}.$$

Let $b = |c_m||\mathbf{A}_m|$ be the largest of the k hyperreal numbers $|c_1||\mathbf{A}_1|, \dots, |c_k||\mathbf{A}_k|$, and let $d_\ell = c_\ell|\mathbf{A}_\ell|/b$ for $\ell = 1, \dots, k$. Then

$$d_1\mathbf{U}_1 + \cdots + d_k\mathbf{U}_k = \mathbf{0},$$

each d_ℓ is finite, and $|d_m| = 1$. Taking standard parts we see that

$$\mathrm{st}(d_1)\mathrm{st}(\mathbf{U}_1) + \cdots + \mathrm{st}(d_k)\mathrm{st}(\mathbf{U}_k) = \mathbf{0}$$

and $\mathrm{st}(d_m) \neq 0$. Therefore the standard parts of the unit vectors are linearly dependent over $\mathbb{R}$, and hence $\mathbf{A}_1, \dots, \mathbf{A}_k$ are almost linearly dependent over $\mathbb{R}$. This proves (i).

(ii) By the Partial Solution Theorem, linear dependence over $\mathbb{R}$ and over $\mathbb{R}^*$ are equivalent for real vectors. Moreover, real vectors are linearly dependent over $\mathbb{R}$ if and only if the standard parts of their unit vectors are linearly dependent over $\mathbb{R}$. This proves (ii). ⊣

10B. Vector Functions (§10.6)

An n-dimensional **real vector function F** maps a subset of $\mathbb{R}$ into the set of n dimensional real vectors. The components of a real vector function $\mathbf{F}$ are real functions $\langle f_1, \dots, f_n\rangle$ with the same domain as $\mathbf{F}$. In symbols,

$$\mathbf{F}(t) = f_1(t)\mathbf{j}_1 + \cdots + f_n(t)\mathbf{j}_n.$$

The natural extension of $\mathbf{F}$ is the hyperreal vector function $\mathbf{F}^*$ whose components are the natural extensions $\langle f_1^*, \dots, f_n^*\rangle$.

We conclude this chapter with definitions and equivalence theorems for limits, continuity, and derivatives of real vector functions of one variable.

Definition 10.12. *Let $\mathbf{F}$ be a real vector function, $\mathbf{A}$ be a real vector, and c be a real number.*

$$\lim_{t\to c}\mathbf{F}(t) = \mathbf{A}$$

means that whenever $t \approx c$ *but* $t \neq c$, $\mathbf{F}(t) \approx \mathbf{A}$. $\mathbf{F}$ *is* **continuous** *at* c *if whenever* $t \approx c$, $\mathbf{F}(t) \approx \mathbf{F}(c)$, *that is,*

$$\lim_{t \to c} \mathbf{F}(t) = \mathbf{F}(c).$$

Theorem 10.13. *The following are equivalent.*

(i) $\lim_{t \to c} \mathbf{F}(t) = \mathbf{A}$.

(ii) For $m = 1, \ldots, n$, $\lim_{t \to c} f_m(t) = a_m$.

(iii) The ε, δ *condition. For every real* $\varepsilon > 0$ *there is a real* $\delta > 0$ *such that whenever* $0 < |t - c| < \delta$, $|\mathbf{F}(t) - \mathbf{A}| < \varepsilon$.

Proof. The equivalence of (i) and (ii) follows from Proposition 10.2, and the equivalence of (ii) and (iii) follows from Theorem 5.1, the equivalence theorem for limits of functions. ⊣

Definition 10.14. *Let* $\mathbf{F}$ *be a real function,* $\mathbf{S}$ *be a real vector, and* c *be a real number.* $\mathbf{F}$ *has* **vector derivative** $\mathbf{S}$ *at* c, *in symbols* $\mathbf{F}'(c) = \mathbf{S}$, *if for every nonzero infinitesimal* $\triangle t$ *we have*

$$\frac{\mathbf{F}(c + \triangle t) - \mathbf{F}(c)}{\triangle t} \approx \mathbf{S}.$$

Thus

$$\mathbf{F}'(c) = \lim_{\triangle t \to 0} \frac{\mathbf{F}(c + \triangle t) - \mathbf{F}(c)}{\triangle t}.$$

Theorem 10.13 leads to ε, δ conditions for the vector derivative. In particular, it follows that $\mathbf{F}'(c)$ exists if and only if $f_1'(c), \ldots, f_n'(c)$ all exist, and

$$\mathbf{F}'(c) = f_1'(c)\mathbf{j}_1 + \cdots + f_n'(c)\mathbf{j}_n.$$

Vector increments and vector differentials are defined as vector dependent variables as follows. We are given a real vector function $\mathbf{X} = \mathbf{F}(t)$ where t is a scalar independent variable and $\mathbf{X}$ a vector dependent variable. We introduce a new scalar independent variable $\triangle t$ and a new vector dependent variable $\triangle \mathbf{X}$, called the **vector increment** of $\mathbf{X}$, with the equation

$$\triangle \mathbf{X} = \mathbf{F}(t + \triangle t) - \mathbf{F}(t).$$

The **vector differential** of $\mathbf{X}$ is a second vector dependent variable whose values are given by the equation

$$d\mathbf{X} = \mathbf{F}'(t)\triangle t.$$

Thus $d\mathbf{X}$ exists when $\mathbf{F}'(t)$ exists, and putting $dt = \triangle t$ we have

$$\frac{d\mathbf{X}}{dt} = \mathbf{F}'(t), \quad \frac{\triangle \mathbf{X}}{\triangle t} \approx \frac{d\mathbf{X}}{dt}.$$

CHAPTER 11

PARTIAL DIFFERENTIATION

11A. Continuity in Two Variables (§11.1, §11.2)

For simplicity we concentrate on real functions of two variables. However, all the notions and results readily extend to n variables.

Recall from Definition 1.32 that the **distance** between two hyperreal points $P(x_1, y_1)$ and $Q(x_2, y_2)$ is

$$|P - Q| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2},$$

that P is **infinitely close** to Q, $P \approx Q$, if $|P - Q| \approx 0$, and that the **monad** of P is the set of all points Q infinitely close to P. We have $P(x_1, y_1) = Q(x_2, y_2)$ if and only if $x_1 = x_2$and $y_1 = y_2$.

Given a real point P and a real number $\delta > 0$, the **real neighborhood** $N_\delta(P)$ is defined as the set of all real points Q such that $|Q - P| < \delta$. The **interior** of a set $D \subseteq \mathbb{R}^2$ is the set of all points P such that some real neighborhood $N_\delta(P)$ is contained in D. $D \subseteq \mathbb{R}^2$ is an **open set** if it is equal to its interior. A point P belongs to the **boundary** of D if every real neighborhood of P meets both D and its complement. It follows that the interior of D and the boundary of D are disjoint, so each open set is disjoint from its boundary. D is a **closed set** if D contains its boundary.

Throughout this chapter we let f be a real function of two variables. The graph of $z = f(x, y)$ is a surface in (x, y, z) space.

The following result was proved for one variable in Theorem 1.28 and Corollary 1.30, and the proof for two variables is similar.

THEOREM 11.1. *(i) Let $Y \subseteq \mathbb{R}^2$ be a set of real points and let $P \in \mathbb{R}^2$ be a real point. If Y^* contains the monad of P, then Y contains some real neighborhood of P.*

(ii) Let $P \in \mathbb{R}^2$. If $f(x, y)$ is defined for all $(x, y) \approx P$, then f is defined on some real neighborhood of P.

DEFINITION 11.2. *Let $(a, b) \in \mathbb{R}^2$ and $L \in \mathbb{R}$. The limit*

$$\lim_{(x,y)\to(a,b)} f(x, y) = L$$

means that whenever $(x, y) \approx (a, b)$ *but* $(x, y) \neq (a, b)$*,* $f(x, y) \approx L$*. Infinite limits are defined in a similar way.*

The function f *is* **continuous** *at* (a, b) *if* $f(a, b)$ *is defined, and* $(x, y) \approx (a, b)$ *implies* $f(x, y) \approx f(a, b)$*.*

COROLLARY 11.3. *f is continuous at (a, b) if and only if f is defined at (a, b) and*

$$\lim_{(x,y)\to(a,b)} f(x, y) = f(a, b).$$

Theorem 1.12 on the standard part function shows that the functions $x+y$, $x-y$, and xy are everywhere continuous. For example, if $(x, y) \approx (a, b) \in \mathbb{R}^2$ then

$$x + y \approx \mathrm{st}(x + y) = \mathrm{st}(x) + \mathrm{st}(y) = a + b.$$

It follows from Theorem 11.1 that any function which is continuous at (a, b) is defined on some real neighborhood of (a, b). As in the one variable case, we have the following equivalence theorem.

THEOREM 11.4. *The following are equivalent.*
(i) $\lim_{(x,y)\to(a,b)} f(x, y) = L$,
(ii) For every real $\varepsilon > 0$ there is a real $\delta > 0$ such that for all $(x, y) \neq (a, b)$ in $N_\delta(a, b)$, $|f(x, y) - L| < \varepsilon$.
(iii) There is a hyperreal $\delta > 0$ such that whenever

$$0 < |(x, y) - (a, b)| < \delta,$$

we have $f(x, y) \approx L$*.*

PROPOSITION 11.5. *Compositions of continuous functions are continuous. That is, if $f(x, y)$ and $g(x, y)$ are continuous at $(x, y) = (a, b)$, and if $h(u, v)$ is continuous at $(u, v) = (f(a, b), g(a, b))$, then the composition*

$$k(x, y) = h(f(x, y), g(x, y))$$

is continuous at $(x, y) = (a, b)$*.*

The proof is the same as in the one variable case, Proposition 3.11.

11B. Partial Derivatives (§11.3, §11.4)

DEFINITION 11.6. *Given a real function $f(x, y)$ and a real point (a, b), the* **partial derivatives** *are defined by*

$$f_x(a, b) = g'(a) \text{ where } g(x) = f(x, b),$$

$$f_y(a, b) = h'(b) \text{ where } h(y) = f(a, y).$$

If $z = f(x, y)$ we use the notation

$$\frac{\partial z}{\partial x} = \frac{\partial f}{\partial x}(a, b) = f_x(a, b), \quad \frac{\partial z}{\partial y} = \frac{\partial f}{\partial y}(a, b) = f_y(a, b).$$

The mere existence of the partial derivatives of $f(x,y)$ tells us nothing about the behavior of f off of the lines $x=a, y=b$. We introduce three stronger notions of differentiability which will be used later on in this chapter.

DEFINITION 11.7. *Let $(a,b)\in\mathbb{R}^2$ and suppose both partial derivatives $f_x(a,,b)$ and $f_y(a,b)$ exist.*
(i) f is **smooth** *at (a,b) if both f_x and f_y are continuous at (a,b).*
(ii) f is **differentiable** *at (a,b) if for any nonzero infinitesimal point $(\triangle x, \triangle y)$,*

$$f(a+\triangle x, b+\triangle y) - f(a,b) \approx f_x(a,b)\triangle x + f_y(a,b)\triangle y$$
$$\left(\text{compared to } \sqrt{\triangle x^2+\triangle y^2}\right).$$

(iii) f is **uniformly differentiable** *at (a,b) if for any hyperreal point $(x,y)\approx(a,b)$ and nonzero infinitesimal point $(\triangle x,\triangle y)$,*

$$f(x+\triangle x, y+\triangle y) - f(x,y) \approx f_x(a,b)\triangle x + f_y(a,b)\triangle y$$
$$\left(\text{compared to } \sqrt{\triangle x^2+\triangle y^2}\right).$$

Differentiability and uniform differentiability correspond to the one variable notions. They are equivalent to real ε,δ conditions which are left to the reader.

When $z=f(x,y)$ we introduce two new dependent variables, the **increment** $\triangle z$ given by

$$\triangle z = f(x+\triangle x, y+\triangle y) - f(x,y)$$

and the **total differential** dz given by

$$dz = f_x(x,y)\triangle x + f_y(x,y)\triangle y = \frac{\partial z}{\partial x}\triangle x + \frac{\partial z}{\partial y}\triangle y.$$

Both $\triangle z$ and dz depend on the four independent variables $x, y, \triangle x, \triangle y$. Using dependent variable notation, f is differentiable at (a,b) if and only if for every nonzero infinitesimal $(\triangle x,\triangle y)$,

$$\triangle z \approx dz \left(\text{compared to } \sqrt{\triangle x^2+\triangle y^2}\right).$$

If f is differentiable at (a,b), the **tangent plane** of f at (a,b) is defined as the plane with the equation

$$z - f(a,b) = f_x(a,b)(x-a) + f_y(a,b)(y-b).$$

Putting $\triangle x = x-a$ and $\triangle y = y-b$, we see that

$$\begin{aligned}\triangle z &= \text{ change in } z \text{ along the surface,}\\ dz &= \text{ change in } z \text{ along the tangent plane.}\end{aligned}$$

By definition, if f is differentiable at (a,b) then whenever (x,y) is infinitely close to but not equal to (a,b), the tangent plane is infinitely close to the surface compared to $\sqrt{\triangle x^2+\triangle y^2}$.

The following result shows the relationship between the three notions of differentiability. The implication (i) $\Rightarrow$ (iii) was called the Increment Theorem in *Elementary Calculus*. The one-variable form of the implication (i) $\Rightarrow$ (ii) was given by Theorem 3.37 (i).

THEOREM 11.8. *Let (a, b) be a real point. Each condition below implies the next, (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv).*

(i) f is smooth at (a, b).
(ii) f is uniformly differentiable at (a, b).
(iii) f is differentiable at (a, b).
(iv) f is continuous at (a, b).

PROOF. (i) $\Rightarrow$ (ii): Let $(x, y) \approx (a, b)$ and let $(\triangle x, \triangle y)$ be nonzero infinitesimal. Then

$$f(x+\triangle x, y+\triangle y) - f(x, y) =$$
$$[f(x+\triangle x, y+\triangle y) - f(x+\triangle x, y)] + [f(x+\triangle x, y) - f(x, y)].$$

Since f_x and f_y are continuous at (a, b) they are defined everywhere in the monad of (a, b). Using the Hyperreal Mean Value Theorem 3.34 for the one variable function $g(u) = f(x+\triangle x, u)$, we have

$$f(x+\triangle x, y+\triangle y) - f(x+\triangle x, y) = f_y(x+\triangle x, u)\triangle y$$

for some u between y and $y + \triangle y$. Similarly,

$$f(x+\triangle x, y) - f(x, y) = f_x(t, y)\triangle x$$

for some t between x and $x + \triangle x$. Since

$$f_y(x+\triangle x, u) \approx f_y(a, b), \quad f_x(t, y) \approx f_x(a, b)$$

and

$$\frac{\triangle x}{\sqrt{\triangle x^2 + \triangle y^2}}, \quad \frac{\triangle y}{\sqrt{\triangle x^2 + \triangle y^2}}$$

are finite, we have

$$f(x+\triangle x, y+\triangle y) - f(x, y) \approx f_x(a, b)\triangle x + f_y(a, b)\triangle y$$
$$\left(\text{compared to } \sqrt{\triangle x^2 + \triangle y^2}\right).$$

Thus (i) implies (ii). Condition (ii) trivially implies (iii).

(iii) $\Rightarrow$ (iv): Let $(\triangle x, \triangle y) \approx (0, 0)$. By (iii),

$$f(a+\triangle x, b+\triangle y) - f(a, b) \approx f_x(a, b)\triangle x + f_y(a, b)\triangle y \approx 0.$$

Therefore f is continuous at (a, b). ⊣

The next two theorems are the two-variable analogues of Theorems 3.37 (ii) and 3.39. The proofs are left to the reader.

THEOREM 11.9. *If f is uniformly differentiable at every point of an open set $Y \subseteq \mathbb{R}^2$, then f is smooth at every point of Y.*

Theorem 11.10. *If $f(x, y)$ is uniformly differentiable at a real point (a, b) then f is continuous on some real neighborhood of (a, b).*

11C. Chain Rule and Implicit Functions (§11.5, §11.6)

The Chain Rule for functions of two variables holds when all functions involved are differentiable in the sense of the preceding section.

In *Elementary Calculus* the theory was simplified by concentrating on smooth functions. Some theorems were stated under the hypothesis that f is smooth when only differentiability or uniform differentiability was actually needed. In this section we give detailed proofs of the basic results and keep track of the kind of differentiability which is actually needed in each case.

Theorem 11.11. *(Chain Rule for Two Variables) Suppose $x = f(s, t)$ and $y = g(s, t)$ are differentiable at the real point (s_0, t_0) and $z = h(x, y)$ is differentiable at the real point*

$$(x_0, y_0) = (f(s_0, t_0), g(s_0, t_0)).$$

Then the composition

$$z = H(s, t) = h(f(s, t), g(s, t))$$

is differentiable at (s_0, t_0), and its partial derivatives are

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s}, \quad \frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}.$$

Proof. Let $(\triangle s, \triangle t)$ be a nonzero infinitesimal point and let $\triangle x, \triangle y$, and $\triangle z$ be the corresponding increments in x, y, and z. Let $\delta = \sqrt{\triangle s^2 + \triangle t^2}$. Then at the point (s_0, t_0),

$$\triangle x \approx \frac{\partial x}{\partial s}\triangle s + \frac{\partial x}{\partial t}\triangle t \text{ (compared to } \delta),$$

$$\triangle y \approx \frac{\partial y}{\partial s}\triangle s + \frac{\partial y}{\partial t}\triangle t \text{ (compared to } \delta).$$

It follows that $\triangle x \approx 0, \triangle y \approx 0$. Also, $\triangle x/\delta, \triangle y/\delta$, and hence $\sqrt{\triangle x^2 + \triangle y^2}/\delta$ are finite. Therefore

$$\triangle z \approx \frac{\partial z}{\partial x}\triangle x + \frac{\partial z}{\partial y}\triangle y \text{ (compared to } \delta).$$

When $\triangle t = 0$ we have $\delta = \triangle s$ and

$$\frac{\triangle z}{\triangle s} \approx \frac{\partial z}{\partial x}\frac{\triangle x}{\triangle s} + \frac{\partial z}{\partial y}\frac{\triangle y}{\triangle s} \approx \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s},$$

so

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s}.$$

The formula for $\partial z/\partial t$ is obtained in a similar way when $\triangle s = 0$. Finally, H is differentiable at (s_0, t_0) because

$$\frac{\triangle z}{\delta} \approx \frac{\partial z}{\partial x}\frac{\triangle x}{\delta} + \frac{\partial z}{\partial y}\frac{\triangle y}{\delta}$$
$$\approx \frac{\partial z}{\partial x}\left(\frac{\partial x}{\partial s}\triangle s + \frac{\partial x}{\partial t}\triangle t\right)\frac{1}{\delta} + \frac{\partial z}{\partial y}\left(\frac{\partial y}{\partial s}\triangle s + \frac{\partial y}{\partial t}\triangle t\right)\frac{1}{\delta},$$

so

$$\triangle z \approx \frac{\partial z}{\partial s}\triangle s + \frac{\partial z}{\partial t}\triangle t \text{ (compared to } \delta).$$

⊣

We now turn to the Implicit Function Theorem, first for two variables and then for three variables.

DEFINITION 11.12. *An* **implicit function** *of a real curve $F(x, y) = 0$ at a real point (a, b) is a real function $y = g(x)$ such that:*

(i) $g(a) = b$;
(ii) The domain of g is a real neighborhood of a;
(iii) The graph of g is a subset of the graph of the equation $F(x, y) = 0$.

An **implicit function** *of a real surface $F(x, y, z) = 0$ at a real point (a, b, c) is a real function $z = h(x, y)$ such that:*

(iv) $h(a, b) = c$;
(v) The domain of h is a real neighborhood of (a, b);
(vi) The graph of h is a subset of the graph of $F(x, y, z) = 0$.

The notion of uniform differentiability will be useful for the Implicit Function Theorem.

THEOREM 11.13. *(Two Variable Implicit Function Theorem) Suppose that at the real point (a, b), $z = F(x, y)$ is uniformly differentiable, $F(a, b) = 0$, and $F_y(a, b) \neq 0$. Then the curve $F(x, y) = 0$ has an implicit function at (a, b). Moreover, for every implicit function $g(x)$ at (a, b), g is uniformly differentiable at a and*

$$g'(a) = -\frac{F_x(a, b)}{F_y(a, b)}.$$

PROOF. Since $F_y(a, b) \neq 0$, it follows from the ε, δ condition for uniform differentiability at (a, b) (the two variable form of Theorem 5.4) that there is a real neighborhood $N_\delta(a, b)$ with the following property. For any two distinct points (x, y) and $(x, y + \triangle y)$ of $N_\delta(a, b)$, $F(x, y) \neq F(x, y + \triangle y)$. We may therefore define a real function g by

$$g(x) = y \text{ iff } (x, y) \in N_\delta(a, b) \text{ and } F(x, y) = 0.$$

Obviously $g(a) = b$. We show first that g is continuous at a. Let $x_1 \approx a$. Since F is continuous at (a, b), $F(x_1, b) \approx 0$. Since F is uniformly differentiable at (a, b), for every nonzero infinitesimal $\triangle y$ we have

$$\frac{F(x_1, b + \triangle y) - F(x_1, b)}{\triangle y} \approx F_y(a, b) \neq 0.$$

Hence there is a $\triangle y \approx 0$ such that 0 is between $F(x_1, b)$ and $F(x_1, b + \triangle y)$. Since F is continuous on $N_\delta(a, b)$, it follows from the Hyperreal Intermediate Value Theorem 3.32 that there is a hyperreal point $y_1 \approx b$ with $F(x_1, y_1) = 0$. Every real solution of

$$F(x, y) = 0, \quad |(x, y) - (a, b)| < \delta$$

is a real solution of $y = g(x)$. By Transfer, $y_1 = g(x_1)$. Thus g is defined in the monad of a and is continuous at a. By Theorem 11.1 (or Corollary 1.30), g is defined on some real neighborhood of a. By definition, the graph of g is a subset of the graph of $F(x, y) = 0$.

It remains to show that g is uniformly differentiable at a and has the required derivative. Let $x \approx a$ and let $\triangle x$ be nonzero infinitesimal. Let $y = g(x)$, $\triangle y = g(x + \triangle x) - g(x)$, and $\triangle s = \sqrt{\triangle x^2 + \triangle y^2}$. Since g is continuous at s, $\triangle y \approx 0$. Since F is uniformly differentiable at (a, b),

$$\begin{aligned} 0 &= F(x + \triangle x, y + \triangle y) - F(x, y) \\ &\approx F_x(a, b)\triangle x + F_y(a, b)\triangle y \text{ (compared to } \triangle x\text{)}. \end{aligned}$$

Since $F_y(a, b) \neq 0$,

$$\frac{\triangle y}{\triangle s} \approx -\frac{F_x(a, b)}{F_y(a, b)} \frac{\triangle x}{\triangle s}. \tag{57}$$

We cannot have $\triangle x/\triangle s \approx 0$, because it leads to the contradiction

$$1 \approx \frac{\triangle y}{\triangle s} \approx -\frac{F_x(a, b)}{F_y(a, b)} \cdot 0 = 0.$$

Therefore we may divide both sides of (57) by $\triangle x/\triangle s$, and get

$$\frac{\triangle y}{\triangle x} = \frac{\triangle y/\triangle s}{\triangle x/\triangle s} \approx -\frac{F_x(a, b)}{F_y(a, b)}.$$

We conclude that g is uniformly differentiable at a and

$$g'(a) = -\frac{F_x(a, b)}{F_y(a, b)}.$$

⊣

We now turn to the Implicit Function Theorem in three variables. It is proved in exactly the same way as the two variable theorem.

Theorem 11.14. *(Three variable Implicit Function Theorem) Suppose that at the point (a, b, c), $w = F(x, y, z)$ is uniformly differentiable, $F(a, b, c) = 0$, and $F_z(a, b, c) \neq 0$. Then the curve $F(x, y, z) = 0$ has an implicit function at (a, b). Moreover, every implicit function $h(x, y)$ is uniformly differentiable at (a, b), has partial derivatives*

$$h_x(a, b) = -\frac{F_x(a, b, c)}{F_z(a, b, c)}, \quad h_y(a, b) = -\frac{F_y(a, b, c)}{F_z(a, b, c)}$$

and has the tangent plane

$$F_x(a, b, c)(x - a) + F_y(a, b, c)(y - b) + F_z(a, b, c)(z - c) = 0.$$

11D. Maxima and Minima (§11.7)

We discuss maxima and minima for real functions of two variables.

Definition 11.15. *Let $f(x, y)$ be a real function with domain D. f has a* **maximum** *at a point $(a, b) \in D$ if $f(a, b) \geq f(x, y)$ for all $(x, y) \in D$. The value $f(a, b)$ is called the* **maximum value** *of f. A* **minimum** *of f and the* **minimum value** *are defined analogously.*

The following two results are proved exactly as in the one variable case.

Proposition 11.16. *Suppose a real function $f(x, y)$ with domain D has a maximum at a real point (a, b). Then the natural extension of f has a maximum at (a, b), that is, $f(a, b) \geq f(x, y)$ for all hyperreal points $(x, y) \in D^*$.*

Theorem 11.17. *Given a real function $f(x, y)$ and a real point (a, b), the following are equivalent.*
(i) f has a local maximum at (a, b), that is, the restriction of f to some real neighborhood of (a, b) has a maximum at (a, b).
(ii) Whenever $(x, y) \approx (a, b)$, $f(a, b) \geq f(x, y)$.

By Corollary 1.29 and Theorem 1.31 for two variables, a set of points $D \subseteq \mathbb{R}^2$ is closed and bounded if and only if

(i) Every point $(x, y) \in D^*$ is finite, and
(ii) Whenever $(x, y) \in D^*$, $(\mathrm{st}(x), \mathrm{st}(y)) \in D$.

We use this fact to prove the Extreme Value Theorem for two variables.

Theorem 11.18. *(Extreme Value Theorem) Suppose the domain of $f(x, y)$ is a closed and bounded set D, and $f(x, y)$ is continuous on D. Then f has a maximum and a minimum.*

Proof. Let Y be the range of f,

$$Y = \{f(x, y) \colon (x, y) \in D\}.$$

By Proposition 1.27, f^* has domain D^* and range Y^*. For each $(x_1, y_1) \in D^*$, we have $(\mathrm{st}(x_1), \mathrm{st}(y_1)) \in D$, and by the continuity of f,

$$f(x_1, y_1) \approx f(\mathrm{st}(x_1), \mathrm{st}(y_1)).$$

Therefore $f(x_1, y_1)$ is finite, so every element of Y^* is finite. By Theorem 1.31, the set Y is bounded, so $Y \subseteq [A, B]$ for some real A and B. Consider a positive integer n, and partition the interval $[A, B]$ into n equal subintervals of length $\delta = (B - A)/n$. Let $k = k(n)$ be the largest integer such that Y meets the kth subinterval. Then there is a point $(g(n), h(n))$ such that

$$(g(n), h(n)) \in D, \quad A + (k-1)\delta \leq f(g(n), h(n)),$$

but for all $(x, y) \in D$,

$$f(x, y) \leq A + k\delta.$$

Now let n_1 be a positive infinite hyperinteger and let $\delta_1 = (B - A)/n_1$ and $k_1 = k(n_1)$. Using Transfer we see that

$$(g(n_1), h(n_1)) \in D^*, \quad A + (k_1 - 1)\delta_1 \leq f(g(n_1), h(n_1)).$$

By another use of Transfer, for every point $(x_1, y_1) \in D^*$ we have

$$f(x_1, y_1) \leq A + k_1\delta_1.$$

Since D is closed and bounded, $(g(n_1), h(n_1))$ is finite, and has a standard part

$$(a, b) = (\mathrm{st}(g(n_1)), \mathrm{st}(h(n_1))) \in D.$$

By continuity of f,

$$f(a, b) = \mathrm{st}(f(g(n_1), h(n_1))).$$

It follows that for any point $(x, y) \in D$,

$$f(x, y) \leq \mathrm{st}(A + k_1\delta_1) = \mathrm{st}(A + (k_1 - 1)\delta_1) \leq f(a, b),$$

so f has a maximum at (a, b). $\dashv$

The above proof differs from the proof of the one variable Extreme Value Theorem 3.28 because this time we partitioned the range of f instead of the domain of f. The one variable proof can be generalized to the present case by using a rectangular grid to partition the domain of f.

Here is a hyperreal form of the Extreme Value Theorem.

Theorem 11.19. *(Hyperreal Extreme Value Theorem) Suppose f is continuous on its domain D, and E is a hyperreal closed rectangle which is contained in D^*. Then f^* has a maximum and minimum on E.*

Proof. By the Extreme Value Theorem 11.18, for every real closed rectangle E with sides $[s, t]$ and $[u, v]$ there is a point $x = g(s, t, u, v), y = h(s, t, u, v)$ such that $(x, y) \in E$ and either $(x, y) \notin D$ or f has a maximum on E at (x, y). Then every real solution of

$$s \leq x_1 \leq t,\ u \leq y_1 \leq v,\ x = g(s, t, u, v),\ y = h(s, t, u, v),\ (x, y) \in D \tag{58}$$

is a solution of

$$s \le x \le t,\ u \le y \le v,\ f(x,y) \ge f(x_1,y_1). \tag{59}$$

By Transfer, every hyperreal solution of (58) is a solution of (59). For any hyperreal closed rectangle $E \subseteq D^*$ with sides $[s,t]^*$ and $[u,v]^*$, f^* is defined everywhere in E. It follows that f^* has a maximum on E at

$$x = g(s,t,u,v), y = h(s,t,u,v).$$

The proof that f^* has a minimum is similar. ⊣

The above theorem can easily be generalized. For instance, later in this section we will introduce the notion of a basic closed hyperreal region, and a proof similar to the above works when E is such a region.

DEFINITION 11.20. *For simplicity, for the rest of this chapter we only consider functions f such that f is differentiable at every interior point of its domain D. An interior point (a,b) of D is said to be a* **critical point** *of f if both partial derivatives of f are equal to zero,*

$$f_x(a,b) = 0, \quad f_y(a,b) = 0.$$

THEOREM 11.21. *(Critical Point Theorem) Suppose f is continuous on its domain D and f is differentiable at every interior point of D. If f has a maximum or minimum at a real point (a,b), then (a,b) is either a boundary point of D or a critical point of f.*

PROOF. Suppose f has a maximum at an interior point (a,b) of D. Then the one variable function $g(x) = f(x,b)$ has a maximum at the interior point a. By the one variable Critical Point Theorem 3.29, $g'(a) = f_x(a,b) = 0$. Similarly, $f_y(a,b) = 0$. ⊣

As in the one variable case, the Extreme Value Theorem and Critical Point Theorem lead to a method for finding the maxima and minima of a function with a bounded closed domain. We will see that the method can often be applied for other domains as well. In *Elementary Calculus* we concentrated on **basic closed regions**, that is, sets D of the form

$$D = \{(x,y) \colon a \le x \le b,\ g(x) \le y \le h(x)\}$$

where g and h are continuous and $g(x) \le h(x)$ for x in $[a,b]$. The following lemma ties the present approach to the simpler treatment in *Elementary Calculus.*

LEMMA 11.22. *Let D be a basic closed region in the plane.*
(i) D is a bounded closed set.
(ii) The boundary of D is the set of all points of D on one of the four curves

$$x = a, \quad x = b, \quad y = g(x), \quad y = h(x).$$

PROOF. (i) Since g and h are continuous, they have maxima and minima on $[a, b]$, so D is bounded. Let $(x_1, y_1) \in D^*$ and let $x = \mathrm{st}(x_1), y = \mathrm{st}(y_1)$. Then

$$a \le x_1 \le b, \quad g(x_1) \le y_1 \le h(x_1).$$

By the continuity of g and h,

$$g(x) = \mathrm{st}(g(x_1)), \quad h(x) = \mathrm{st}(h(x_1)).$$

Therefore

$$a \le x \le b, \quad g(x) \le y \le h(x),$$

so $(x, y) \in D$.

(ii) Obviously any point on one of the four curves is on the boundary of D. For example, if $a \le x \le b$ and $g(x) = y$ then for any real $\varepsilon > 0$, $y - \varepsilon/2 < g(x)$ and hence $(x, y - \varepsilon/2) \notin D$, so $N_\varepsilon(x, y)$ is not contained in D and (x, y) is on the boundary of D.

Suppose $(x, y) \in D$ but (x, y) is not on one of the four curves. Then

$$a < x < b, \quad g(x) < y < h(x).$$

For any $(x_1, y_1) \approx (x, y)$ we have $a < x_1 < b$. Also, $g(x_1) < y_1 < h(x_1)$ because if, for example, $y_1 \le g(x_1)$, then

$$y = \mathrm{st}(y_1) \le \mathrm{st}(g(x_1)) = g(st(x_1)) = g(x),$$

contradicting the assumption that $g(x) < y$. Thus $(x_1, y_1) \in D^*$, so D^* contains the monad of (x, y). It follows by Theorem 1.28 that (x, y) belongs to the interior of D. ⊣

In many applications, a function f is continuous on a basic closed region D and is differentiable and has only finitely many critical points on the interior of D. The maximum of f can usually be found as follows. First, evaluate f at each of its critical points. Second, on each of the four boundary curves of D find the maxima of f by eliminating one variable and using the one variable method. Finally, the largest of the values of f at the critical points and on the boundary curves must be the maximum value of f.

We now present a method for finding maxima and minima on open regions. A **basic open region** in the plane is a set $D \subseteq \mathbb{R}^2$ of one of the forms

$$\begin{aligned}
&\{(x, y) \colon x \in I,\, g(x) < y < h(x)\},\\
&\{(x, y) \colon x \in I,\, -\infty < y < h(x)\},\\
&\{(x, y) \colon x \in I,\, g(x) < y < \infty\},\\
&\{(x, y) \colon x \in I,\, -\infty < y < \infty\}
\end{aligned}$$

where I is an open interval, g and h are continuous real functions on I, and $g(x) < h(x)$ for all $x \in I$.

A **basic closed hyperreal region** in the plane is a set $E \subseteq \mathbb{R}^*$ of hyperreal points of the form

$$E = \{(x, y) \colon a \le x \le b,\, g(x) + A \le y \le h(x) + B\}$$

where a, b, A, B are hyperreal numbers and g, h are real functions on an interval I such that $a, b \in I^*$.

A **hyperreal cover** of a basic open region D is a basic closed hyperreal region E obtained as follows. Let ε be a positive infinitesimal and H be a positive infinite hyperreal number. If

$$D = \{(x, y) \in \mathbb{R}^2 \colon a < x < b,\ g(x) < y < h(x)\},$$

then

$$E = \{(x, y) \in (\mathbb{R}^*)^2 \colon a + \varepsilon \leq x \leq b - \varepsilon,\ g(x) + \varepsilon \leq y \leq h(x) - \varepsilon\}.$$

If

$$D = \{(x, y) \in \mathbb{R}^2 \colon a < x < \infty,\ g(x) < y < \infty\},$$

then

$$E = \{(x, y) \in (\mathbb{R}^*)^2 \colon a + \varepsilon \leq x \leq H,\ g(x) + \varepsilon \leq y \leq H\}.$$

The other cases are similar.

Note that every basic open region D has a hyperreal cover. In fact, there is one hyperreal cover for each positive infinitesimal $\varepsilon > 0$ and positive infinite H.

Lemma 11.23. *If E is a hyperreal cover of a basic open region D, then $D \subseteq E \subseteq D^*$.*

This lemma follows easily from the definition.

The standard calculus course often gives the Second Derivative Test for local maxima, minima, and saddle points, but no practical test for global maxima and minima. The following method for finding global maxima and minima for open regions is presented in *Elementary Calculus*. It is an application of hyperreal numbers to an elementary problem about real functions. For a function f with an open domain D, no point of D is on the boundary. Thus by the Critical Point Theorem, if f has a maximum, it must occur at a critical point. But it is often hard to determine whether or not f has a maximum. The next result is a practical criterion for the existence of a maximum of f.

By a **maximum critical point** of $f(x, y)$ we mean a critical point (x_0, y_0) such that for every other critical point (x_1, y_1), $f(x_0, y_0) \geq f(x_1, y_1)$. Obviously, if $f(x, y)$ has at least one but only finitely many critical points, it has a maximum critical point.

Theorem 11.24. *Suppose a real function $f(x, y)$ is defined on a basic open region D, has partial derivatives at every point of D, and has a maximum critical point (c, d). Let E be a hyperreal cover of D. Then*

(i) If $f(c, d) \geq f(x, y)$ for all (x, y) on the boundary of E, then f has a maximum in D at (c, d).

(ii) If $f(c, d) < f(x, y)$ for some (x, y) on the boundary of E, then f has no maximum in D.

A similar result holds for minima.

Proof. We give the proof in the case that D has the form

$$D = \{(x, y)\colon a < x < \infty,\ g(x) < y < \infty\}.$$

The hyperreal cover E has the form

$$E = \{(x, y)\colon a + \varepsilon \le x \le H,\ g(x) + \varepsilon \le y \le H\}$$

where ε is positive infinitesimal and H is positive infinite. We first prove (ii). Suppose f has a maximum in D. To prove (ii) it suffices to show that $f(c, d) \ge f(x, y)$ for all (x, y) on the boundary of E. Since D is open, no point of D is on the boundary of D. Hence by the Critical Point Theorem 11.21, the maximum must occur at the maximum critical point (c, d). By Proposition 11.16, $f(c, d) \ge f(x, y)$ for all $(x, y) \in D^*$. Since $E \subseteq D^*$, it follows that $f(c, d) \ge f(x, y)$ for all (x, y) on the boundary of E. This proves (ii).

The proof of (i) uses the Partial Solution Theorem. For each positive real ε_0 and H_0, let $E(\varepsilon_0, H_0)$ be the basic closed region

$$E(\varepsilon_0, H_0) = \{(x, y)\colon a + \varepsilon_0 \le x \le H_0,\ g(x) + \varepsilon_0 \le y \le H_0\}.$$

Then

$$E(\varepsilon_0, H_0) \subseteq D \subseteq E \subseteq D^*.$$

Let $\partial E(\varepsilon_0, H_0)$ be the boundary of $E(\varepsilon_0, H_0)$. The relations $(x, y) \in E(\varepsilon_0, H_0)$ and $(x, y) \in \partial E(\varepsilon_0, H_0)$ can both be expressed by systems of formulas. By the Extreme Value and Critical Point Theorems, if $(c, d) \in E(\varepsilon_0, H_0)$ then in $E(\varepsilon_0, H_0)$, f has a maximum either at (c, d) or on $\partial E(\varepsilon_0, H_0)$. Thus every real solution of

$$(c, d) \in E(\varepsilon_0, H_0), \quad (s, t) \in E(\varepsilon_0, H_0), \quad f(c, d) < f(s, t) \tag{60}$$

is a partial real solution of

$$(s, t) = (s, t), \quad (x, y) \in \partial E(\varepsilon_0, H_0), \quad f(c, d) < f(x, y). \tag{61}$$

Suppose f does not have a maximum in D at (c, d). To prove (i) we must show that $f(c, d) < f(u, v)$ for some (u, v) on the boundary of E. We have $f(c, d) < f(s, t)$ for some real point $(s, t) \in D$. Since $D \subseteq E = E(\varepsilon, H)$, (s, t, ε, H) is a hyperreal solution of the system of formulas (60). By the Partial Solution Theorem, there is a hyperreal point (u, v) such that (61) holds for $(s, t, \varepsilon, H, u, v)$. Thus (u, v) is on the boundary of E and $f(c, d) < f(u, v)$. This proves (i). ⊣

11E. Second Partial Derivatives (§11.8)

Given a function $f(x, y)$ of two variables, one can consider four second partial derivatives, the pure second partials

$$\frac{\partial^2 f}{\partial x^2} = f_{xx} = (f_x)_x \text{ and } \frac{\partial^2 f}{\partial y^2} = f_{yy} = (f_y)_y$$

and the mixed second partials

$$\frac{\partial^2 f}{\partial x \partial y} = f_{xy} = (f_x)_y \text{ and } \frac{\partial^2 f}{\partial y \partial x} = f_{yx} = (f_y)_x.$$

We give a hyperreal proof of the equality of the mixed second partial derivatives. The hypothesis we need is the uniform differentiability of both first partial derivatives. (In *Elementary Calculus* we used the stronger hypothesis that both first partial derivatives are smooth). The proof uses the following form of the Hyperreal Mean Value Theorem for two variables.

Theorem 11.25. *(Hyperreal Mean Value Theorem) Suppose $f(x, y)$ is a real function whose partial derivative $f_x(x, y)$ exists on an open rectangle D. Then for every pair of hyperreal points (x, y) and $(x + \triangle x, y)$ in D^* with $\triangle x \neq 0$ there is a point t between x and $x + \triangle x$ such that*

$$\frac{f(x + \triangle x, y) - f(x, y)}{\triangle x} = f_x(t, y).$$

A similar result holds for the other partial derivative f_y.

Proof. By the real Mean Value Theorem 3.30 in one variable, every real solution of

$$(x, y) \in D, \quad (x, x + \triangle x, y) \in D, \quad \triangle x > 0$$

is a partial real solution of

$$\frac{f(x, x + \triangle x, y) - f(x, y)}{\triangle x} = f_x(t, y), \quad x < t < x + \triangle x.$$

The result now follows from the Partial Solution Theorem. $\dashv$

Theorem 11.26. *Suppose both partial derivatives f_x and f_y of a real function $f(x, y)$ are uniformly differentiable at a real point (a, b). Then the mixed second partial derivatives are equal at (a, b),*

$$\frac{\partial^2 f}{\partial x \partial y}(a, b) = \frac{\partial^2 f}{\partial y \partial x}(a, b).$$

Proof. Let $\triangle x$ and $\triangle y$ be nonzero infinitesimals. Since f_x and f_y are uniformly differentiable at (a, b), they exist and are continuous on a real neighborhood of (a, b), by Theorem 11.10. Let

$$\delta = [f(a + \triangle x, b + \triangle y) - f(a + \triangle x, b)] - [f(a, b + \triangle x) - f(a, b)].$$

By the Hyperreal Mean Value Theorem 11.25 applied to the function

$$g(x, \triangle y) = f(x, b + \triangle y) - f(x, b),$$

we have

$$\begin{aligned} \frac{\delta}{\triangle y} &= \frac{f(a + \triangle x, b + \triangle y) - f(a + \triangle x, b)}{\triangle y} - \frac{f(a, b + \triangle x) - f(a, b)}{\triangle y} \\ &= f_y(a + \triangle x, u) - f_y(a, u) \end{aligned}$$

for some u between b and $b + \triangle y$. By uniform differentiability of f_x and f_y,

$$\frac{\delta}{\triangle x \triangle y} = \frac{f_x(t, b + \triangle y) - f_x(t, b)}{\triangle y} \approx \frac{\partial^2 f}{\partial y \partial x}(a, b)$$

and

$$\frac{\delta}{\triangle x \triangle y} = \frac{f_y(a + \triangle x, u) - f_y(a, u)}{\triangle x} \approx \frac{\partial^2 f}{\partial x \partial y}(a, b).$$

Therefore

$$\frac{\partial^2 f}{\partial y \partial x}(a, b) = \frac{\partial^2 f}{\partial x \partial y}(a, b).$$

$\dashv$

CHAPTER 12

MULTIPLE INTEGRATION

Double and triple integrals are developed in Chapter 12 of *Elementary Calculus.* To simplify our notation we will concentrate entirely on double integrals here. However, everything we do in this chapter cn readily be generalized to triple integrals.

Permanent Assumption *We assume throughout this Chapter that* $D_0 \subseteq \mathbb{R}^2$ *is an open set in the plane, and that* $f(x, y)$ *is a real function which is continuous on* D_0*.*

12A. Double Integrals (§12.1, §12.2)

We will define the double integral of a continuous real function $f(x, y)$ and state the basic results. Most of the proofs are similar to corresponding proofs for the single integral in Chapter 4 and will be omitted here.

We will consider basic closed subregions $D \subseteq D_0$. The analogue for f of an area function for a real function of one variable is a volume function.

Definition 12.1. *A* **volume function** *for* f *is a function* $B(D)$ *from the set of basic closed regions* $D \subseteq D_0$ *into the real numbers which has the following properties:*

Addition Property*: If* $D = D_1 \cup D_2$ *where*

$$D_1 = \{(x, y) \in D \colon x \le c\},$$

$$D_2 = \{(x, y) \in D \colon x \ge c\},$$

then $B(D) = B(D_1) + B(D_2)$*.*

If $D = D_1 \cup D_2$ *where*

$$D_1 = \{(x, y) \in D \colon y \le k(x)\},$$
$$D_2 = \{(x, y) \in D \colon y \ge k(x)\},$$

where $k(x)$ *is continuous on* $\mathbb{R}$*, then* $B(D) = B(D_1) + B(D_2)$*.*

Cylinder Property: *If $f(x,y)$ has minimum value m and maximum value M on D, and A is the area of D, then*

$$mA \leq B(D) \leq MA.$$

The Addition Property states that if we divide D into two basic closed regions D_1 and D_2 with either a vertical line $x = c$ or a continuous curve $y = k(x)$, then $B(D) = B(D_1) + B(D_2)$. The Cylinder Property states that $B(D)$ is between the volume of the cylinder above D with the minimum height m and the volume of the cylinder above D with the maximum height M. The intuitive notion of the volume of a solid above the region D between the horizontal plane $z = 0$ and the surface $z = f(x,y)$ has the Addition and Cylinder properties. It will turn out that the double integral is the unique volume function for f. Our first step in defining the double integral is the finite double Riemann sum.

Let

$$D = \{(x,y)\colon a \leq x \leq b,\, g(x) \leq y \leq h(x)\}$$

be a basic closed region included in D_0, and let

$$B_1 = \text{ minimum value of } g(x),\ a \leq x \leq b,$$

$$B_2 = \text{ minimum value of } h(x),\ a \leq x \leq b.$$

The rectangle

$$[a,b] \times [B_1, B_2]$$

is called the **circumscribed rectangle** of D. Let $\triangle x$ and $\triangle y$ be positive real numbers and divide the intervals $[a,b]$ and $[B_1, B_2]$ into subintervals of length $\triangle x$ and $\triangle y$ respectively,

$$x_0 = a,\ x_1 = a + \triangle x,\ \ldots,\ x_m = a + m\triangle x$$

where $x_m < b \leq x_m + \triangle x$,

$$y_0 = B_1,\ y_1 = B_1 + \triangle y,\ \ldots,\ y_n = B_1 + n\triangle y$$

where $y_n < B_2 \leq y_n + \triangle y$. If $\triangle x$ does not evenly divide $b - a$, the last subinterval $[x_m, b]$ will have length less than $\triangle x$ and will be covered by the interval $[x_m, x_m + \triangle x]$. Thus the circumscribed rectangle $[a,b] \times [B_1, B_2]$ of D is covered by a grid of $\triangle x$ by $\triangle y$ rectangles.

Definition 12.2. *The* **finite double Riemann sum** *of f over D is defined as*

$$\sum\sum_D f(x,y)\triangle x\triangle y = \sum_{k=0}^{m}\sum_{\ell=0}^{n}\{f(x_k, y_\ell)\triangle x\triangle y\colon (x_k, y_\ell) \in D\}.$$

Thus $\sum\sum_D f(x,y)\triangle x\triangle y$ is the sum of the volumes of the rectangles of base $\triangle x\triangle y$ and height $f(x_k, y_\ell)$ such that the point (x_k, y_ℓ) belongs to D. For a given region D, the finite double Riemann sum $\sum\sum_D f(x,y)\triangle x\triangle y$ is a real function of two variables $\triangle x$ and $\triangle y$, and is defined for all $\triangle x > 0, \triangle y > 0$. By the Function Axiom, its natural extension is defined for all hyperreal $\triangle x > 0, \triangle y > 0$.

Definition 12.3. *If dx and dy are positive infinitesimals, the hyperreal number*

$$\sum\sum_{D} f(x,y)\,dx\,dy$$

is called the **infinite Riemann sum** *of f over D with respect to dx and dy.*

Lemma 12.4. *For any basic closed region D and positive infinitesimals dx and dy, the infinite Riemann sum $\sum\sum_D f(x,y)\,dx\,dy$ is a finite hyperreal number.*

Definition 12.5. *Let dx and dy be positive infinitesimals. The* **double integral** *of f over D with respect to dx and dy is the standard part of the infinite Riemann sum,*

$$\iint_D f(x,y)\,dx\,dy = \mathrm{st}\left(\sum\sum_{D} f(x,y)\,dx\,dy\right).$$

Theorem 12.6. *The value of the double integral of f over a basic closed region D does not depend on dx or dy. That is, if dx, d_1x, dy, d_1y are positive infinitesimals, then*

$$\iint_D f(x,y)\,dx\,dy = \iint_D f(x,y)\,d_1x\,d_1y.$$

Hereafter we will write $dA = dx\,dy$ and denote the double integral by

$$\iint_D f(x,y)\,dx\,dy = \iint_D f(x,y)\,dA.$$

Theorem 12.7. *The double integral is the unique volume function for f.*

This theorem justifies the definition of the **volume** V of the solid over the region D above the plane $z = 0$ and below the surface $z = f(x,y)$ as

$$V = \iint_D f(x,y)\,dA.$$

It follows from the Cylinder Property that the area of a basic closed region D is equal to the definite integral

$$A = \iint_D 1\,dA.$$

The next theorem provides the basic tool for computing double integrals.

Theorem 12.8. *(Iterated Integral Theorem) Let D be the basic closed region*

$$D = \{(x,y) \colon a \le x \le b,\ g(x) \le y \le h(x)\}.$$

Then the double integral of f over D is equal to the iterated integral

$$\iint_D f(x,y)\,dA = \int_a^b \int_{g(x)}^{h(x)} f(x,y)\,dy\,dx.$$

PROOF. Let
$$B(D) = \int_a^b \int_{g(x)}^{h(x)} f(x,y)\,dy\,dx.$$
Using properties of the single integral, a simple computation shows that $B(D)$ is a volume function for f. By Theorem 12.7, the double integral is the only volume function for f, so it must be equal to $B(D)$. ⊣

COROLLARY 12.9. *Suppose D is a basic closed region in both the (x,y) plane and the (y,x) plane, that is,*
$$D = \{(x,y)\colon a \le x \le b,\ g(x) \le y \le h(x)\},$$
$$D = \{(x,y)\colon c \le y \le d,\ k(y) \le x \le \ell(y)\}.$$
Then the two iterated integrals are equal,
$$\int_a^b \int_{g(x)}^{h(x)} f(x,y)\,dy\,dx = \int_c^d \int_{k(y)}^{\ell(y)} f(x,y)\,dx\,dy.$$

PROOF. The definition of the double Riemann sum is symmetric between x and y, so the double integral in the sense of the (x,y) plane equals the double integral in the sense of the (y,x) plane. By the Iterated Integral Theorem 12.8, each of the iterated integrals is equal to the double integral
$$\iint_D f(x,y)\,dA.$$
⊣

12B. Infinite Sum Theorem for Two Variables (§12.3)

As in the case of single integrals, the Infinite Sum Theorem can be used for applications of the double integral.

DEFINITION 12.10. *An* **element of area** *is a hyperreal rectangle $\triangle D \subseteq (D_0)^*$ whose sides are infinitesimal and parallel to the x and y axes. Given an element of area $\triangle D$ we let*
$$(x,y) = \text{ lower left corner of } \triangle D,$$
$$\triangle x, \triangle y = \text{ dimensions of } \triangle D,$$
$$\triangle A = \triangle x \triangle y = \text{ area of } \triangle D.$$

THEOREM 12.11. *(Infinite Sum Theorem) Let $f(x,y)$ be a real function which is continuous on an open set D_0 and let $B(D)$ be a function from basic closed regions $D \subseteq D_0$ to real numbers. Let $\triangle x$ and $\triangle y$ be positive infinitesimals. Assume that:*

(i) B has the Addition Property;
(ii) $B(D) \ge 0$ for all D;

(iii) For every element of area $\triangle D \subseteq (D_0)^*$ with dimensions $\triangle x$ and $\triangle y$,

$$B(\triangle D) \approx f(x,y)\triangle x \triangle y \text{ (compared to } \triangle x \triangle y\text{)}.$$

Then for every basic closed region $D \subseteq D_0$,

$$B(D) = \iint_D f(x,y)\, dA.$$

In *Elementary Calculus* we stated a weaker form of the Infinite Sum Theorem, where (iii) is assumed for all elements of area rather than only for elements of area with fixed dimensions $\triangle x$ and $\triangle y$. The proof there was given only in the case that D is a rectangle. Here we give the proof in general. The proof in the general case is much longer than the proof in the case that D is a rectangle, and uses several lemmas.

Lemma 12.12. *Let D be a closed and bounded subset of the open set D_0. There is a positive real number ε such that the ε-neighborhood of every point of D is included in D_0.*

Proof. Suppose that for each real $\varepsilon > 0$ there are points $P \in D$ and $Q \notin D_0$ such that $|P - Q| < \varepsilon$. Then by the Partial Solution Theorem there are hyperreal points $P_1 \in D^*, Q_1 \notin (D_0)^*$ with $P_1 \approx Q_1$. Since D is closed and bounded, by Corollary 1.29 and Theorem 1.31, P_1 has a standard part P_0, and $P_0 \in D$. Thus $P_0 \in D_0$ and $Q_1 \approx P_0$. Since D_0 is open, by Theorem 1.28 we have $Q_1 \in (D_0)^*$. This contradiction completes the proof. ⊣

Lemma 12.13. *Suppose f is a real function which is continuous on D_0. For every basic closed region $D \subseteq D_0$ there is a basic open region D_1 such that $D \subseteq D_1 \subseteq D_0$ and f is bounded on D_1.*

Proof. Let

$$D = \{(x,y) \colon a \le x \le b,\ g(x) \le y \le h(x)\}.$$

By Lemma 12.12 there is a real $\varepsilon > 0$ such that the ε-neighborhood of any point of D is included in D_0. Therefore the basic closed region

$$D_2 = \{(x,y) \colon a - \varepsilon \le x \le b + \varepsilon,\ g(x) - \varepsilon \le y \le h(y) + \varepsilon\}$$

is included in D_0. By the Extreme Value Theorem 11.18, f has a maximum and a minimum value in D_2, so f is bounded in D_2. Let D_1 be the interior of D_2,

$$D_1 = \{(x,y) \colon a - \varepsilon < x < b + \varepsilon,\ g(x) - \varepsilon < y < h(y) + \varepsilon\}.$$

Then D_1 is a basic open region, $D \subseteq D_1 \subseteq D_0$, and f is bounded on D_1. ⊣

Lemma 12.14. *Suppose the function $B(D)$ on basic closed regions $D \subseteq D_0$ has the Addition Property and $B(D) \ge 0$ for all D. Then whenever $D_1 \subseteq D_2 \subseteq D_0$ we have $B(D_1) \le B(D_2)$.*

Proof. By extending the left and right boundaries of the subregion D_1, which are vertical line segments, to the lower and upper boundary curves of D_2, we get a partition of D_2 into five basic closed regions

$$D_2 = D_1 \cup E_1 \cup E_2 \cup E_3 \cup E_4$$

which meet only on boundaries. By the Addition Property,

$$B(D_2) = B(D_1) + B(E_1) + B(E_2) + B(E_3) + B(E_4).$$

Since each term on the right is ≥ 0, $B(D_1) \leq B(D_2)$. ⊣

Lemma 12.15. *Let D be a basic closed region. For every real $\varepsilon > 0$ there is a real $\delta > 0$ such that for every partition of the circumscribed rectangle of D into a grid of subrectangles of length and width less than δ, the total area of the subrectangles which meet the boundary of D is less than ε.*

Proof. Let

$$D = \{(x, y) \colon a \leq x \leq b,\ g(x) \leq y \leq h(x)\}$$

and let $[a, b] \times [B_1, B_2]$ be the circumscribed rectangle of D. Let ε_1 be a positive real number. Since $g(x)$ and $h(x)$ are continuous, they are uniformly continuous on $[a, b]$ by Theorem 3.15. Hence there is a real $\delta > 0$ such that $\delta < \varepsilon_1$ and whenever $x_1, x_2 \in [a, b]$ with $|x_1 - x_2| < \delta$, we have

$$|g(x_2) - g(x_1)| < \varepsilon_1, \quad |h(x_2) - h(x_1)| < \varepsilon_1.$$

Let $0 < \triangle x < \delta$, $0 < \triangle y < \delta$ and partition the circumscribed rectangle into a grid of $\triangle x$ by $\triangle y$ subrectangles. Each vertical boundary of D is covered by at most two columns of subrectangles which have total area less than $2\varepsilon_1(B_2 - B_1)$, Since $g(x)$ and $h(x)$ change by less than ε_1 over an interval of length $\triangle x$, each of the upper and lower boundary curves of D is covered by a set of subrectangles of total area less than $2\varepsilon_1(b - a)$. Therefore the set of subrectangles which meet the boundary of D has total area less than

$$4\varepsilon_1[(B_2 - B_1) + (b - a)].$$

Thus a value of δ corresponding to

$$\varepsilon_1 = \frac{\varepsilon}{4\varepsilon_1[(B_2 - B_1) + (b - a)]}$$

has the required property. ⊣

Proof of the Infinite Sum Theorem. Let D be the basic closed region

$$D = \{(x, y) \colon a \leq x \leq b,\ g(x) \leq y \leq h(x)\}$$

and suppose $D \subseteq D_0$. We wish to show that $B(D) = \iint_D f(x, y)\, dA$. By Lemma 12.13, we may assume without loss of generality that f is bounded on D_0. We have $f(x, y) \geq 0$ on D_0 because of hypotheses (i) and (iii). Thus for some real M, $0 \leq f(x, y) \leq M$ on D_0. By Lemma 12.12 there is a real $\varepsilon > 0$ such that the ε-neighborhood of every point in D is included in D_0. Let

$[a, b] \times [B_1, B_2]$ be the circumscribed rectangle of D. Consider real numbers $\triangle X, \triangle Y$ such that

$$0 < \triangle X < \varepsilon/2, \quad 0 < \triangle Y < \varepsilon/2.$$

Partition the circumscribed rectangle into a grid of $\triangle X$ by $\triangle Y$ subrectangles. (If $\triangle X$ does not evenly divide $b-a$. the right column of subrectangles will be smaller, and similarly for $\triangle Y$). Let G be the set of all subrectangles and

$$G_1 = \{E \in G \colon E \text{ in included in the interior of } D\},$$

$$G_2 = \{E \in G \colon E \text{ meets the boundary of } D\}.$$

Since $\triangle X + \triangle Y < \varepsilon$, we have $E \subseteq D_0$ whenever $E \in G_1 \cup G_2$. We wish to estimate the difference between $B(D)$ and the finite Riemann sum $\sum\sum_D f(x,y)\triangle X \triangle Y$. Let $p(\triangle X, \triangle Y)$ be the total area of the subrectangles which meet the boundary of D, that is,

$$p(\triangle X, \triangle Y) = \text{ area of } \bigcup G_2.$$

For each $E \in G_1 \cup G_2$ let

$$F(E) = f(x,y)\triangle X \triangle Y$$

where (x, y) is the lower left corner of E, and let $q(\triangle X, \triangle Y)$ be the maximum of the values

$$\frac{|B(E) - F(E)|}{\triangle X \triangle Y}, \quad E \in G_1 \cup G_2.$$

Since $0 \le f(x,y) \le M$ we have

$$0 \le F(E) \le M\triangle X\triangle Y,$$
$$0 \le B(E) \le (M + q(\triangle X, \triangle Y))\triangle X \triangle Y.$$

By Lemma 12.14,

$$\sum_{E \in G_2} B(D \cap E) \le \sum_{E \in G_2} B(E).$$

It follows that

$$\begin{aligned} &\left| B(D) - \sum\sum_D f(x,y)\triangle X \triangle Y \right| \\ &\le \left| \sum_{E \in G_1} (B(E) - F(E)) \right| + \sum_{E \in G_2} B(D \cap E) + \sum_{E \in G_2} F(E) \\ &\le \sum_{E \in G_1} |B(E) - F(E)| + \sum_{E \in G_2} (B(E) + F(E)) \\ &\le (b-a)(B_2 - B_1)q(\triangle X, \triangle Y) + p(\triangle X, \triangle Y)(2M + q(\triangle X, \triangle Y)). \end{aligned}$$

Now consider the given positive infinitesimals $\triangle x, \triangle y$. By Transfer,

$$\left| B(D) - \sum\sum_D f(x,y)\triangle x \triangle y \right|$$

$$\leq (b-a)(B_2 - B_1)q(\triangle x, \triangle y) + p(\triangle x, \triangle y)(2M + q(\triangle x, \triangle y)).$$

By Lemma 12.15 and Transfer, $0 < p(\triangle x, \triangle y) < \varepsilon_0$ for every positive real ε_0, so $p(\triangle x, \triangle y) \approx 0$. By the Partial Solution Theorem there is a $\triangle x$ by $\triangle y$ element of area $\triangle D$ such that $\triangle D$ meets D and

$$\frac{|B(\triangle D) - f(x,y)\triangle x \triangle y|}{\triangle x \triangle y} = q(\triangle x, \triangle y).$$

Therefore $\triangle D \subseteq (D_0)^*$ and by the hypothesis (iii), $q(\triangle x, \triangle y)$ is infinitesimal. We conclude that

$$\left| B(D) - \sum \sum_D f(x,y)\triangle x \triangle y \right| \approx 0,$$

whence

$$B(D) = \iint_D f(x,y)\, dx\, dy.$$

$\dashv$

In *Elementary Calculus* the Infinite Sum Theorem for two variables was used to justify integration formulas for the volume between two surfaces, for mass, first moments, and moments of inertia of plane objects, and for the area of a surface. Another application is given in the next section.

12C. Change of Variables in Double Integrals (§12.5)

In *Elementary Calculus* the Infinite Sum Theorem is used to prove the formula for a double integral over a region

$$D = \{(r,\theta) \colon \alpha \leq \theta \leq \beta,\ a(\theta) \leq r \leq b(\theta)\}$$

in polar coordinates:

$$\iint_D f(x,y)\, dA = \int_\alpha^\beta \int_{a(\theta)}^{b(\theta)} f(r\cos\theta, r\sin\theta) r\, dr\, d\theta.$$

Here we will use the Infinite Sum Theorem to obtain the general formula for a double integral with a change of variables. For comparison we first state the simpler result for single integrals (§4.7 in *Elementary Calculus*).

Theorem 12.16. *Suppose $f(u)$ is continuous on an open interval I, g maps an open interval J into I, and g is continuously differentiable on J. Then for all a, b in J,*

$$\int_{g(a)}^{g(b)} f(u)\, du = \int_a^b f(g(t))g'(t)\, dt.$$

This is proved in *Elementary Calculus* using the Chain Rule and the Fundamental Theorem of Calculus.

In the analogous theorem for two variables, the mapping g is replaced by a smooth transformation from an open region in the (s,t) plane into the (x,y) plane, and the derivative of g is replaced by the Jacobian matrix of the transformation. Our elementary treatment of double integrals over basic closed regions is not adequate for integration by change of variables. The difficulty is that the image of a basic closed region under a smooth transformation is not necessarily a basic closed region. To give a natural treatment we need the classical theory of Jordan content. Once this theory is developed, the Infinite Sum Theorem can be used to obtain the change of variables formula. Our main purpose in this section is to show how the Infinite Sum Theorem is used. To prepare the way we sketch a hyperreal form of Jordan content.

Let D be a bounded set in the plane. The **circumscribed rectangle** E of D is the intersection of all closed rectangles containing D. Given positive real numbers $\triangle x$ and $\triangle y$, partition E into a grid of $\triangle x$ by $\triangle y$ subrectangles. Let $\underline{C}(\triangle x, \triangle y)$ be the sum of the areas of the subrectangles which are entirely within D, and let $\overline{C}(\triangle x, \triangle y)$ be the sum of the areas of the subrectangles which meet D. By the Function Axiom, the natural extensions $\underline{C}(dx, dy)$ and $\overline{C}(dx, dy)$ are defined for all positive infinitesimal dx and dy. We say that D has **Jordan content** $A(D)$ if for all positive infinitesimals dx and dy,

$$A(D) = \mathrm{st}(\underline{C}(dx, dy)) = \mathrm{st}(\overline{C}(dx, dy)).$$

Intuitively, the total area of the infinitesimal subrectangles included in D^* has standard part $A(D)$. Also, the total area of the infinitesimal subrectangles which meet D^* has standard part $A(D)$. The Jordan content of D is also called the **area** of D.

For the rest of this section, suppose that D is a set whose closure is contained in D_0, and D has Jordan content $A(D)$.

The double Riemann sum and double integral are defined exactly as in the case of a basic closed region. That is, for positive infinitesimal dx and dy, we partition the circumscribed rectangle E of D into a dx by dy grid and define

$$\sum\sum_D f(x,y)\,dx\,dy = \sum\sum\{f(x_K, x_L)\,dx\,dy \colon (x_K, x_L) \in D^*\},$$

$$\iint_D f(x,y)\,dx\,dy = \mathrm{st}\left(\sum\sum_D f(x,y)\,dx\,dy\right).$$

The basic properties of double integrals are readily generalized to integrals over sets which have Jordan content, and take the following form.

Theorem 12.17. *The double integral $\iint_D f(x,y)\,dx\,dy$ does not depend on the infinitesimals dx and dy.*

THEOREM 12.18. *(Addition Property) If $D = D_1 \cup D_2$ where $D_1 \cap D_2$ has Jordan content zero, and D_1, D_2 have Jordan content, then*

$$\iint_D f(x,y)\,dx\,dy = \iint_{D_1} f(x,y)\,dx\,dy + \iint_{D2} f(x,y)\,dx\,dy.$$

THEOREM 12.19. *(Cylinder Property) Suppose $m \le f(x,y) \le M$ for $(x,y) \in D$. Then*

$$m \cdot A(D) \le \iint_D f(x,y)\,dx\,dy \le M \cdot A(D).$$

Taking $f(x,y) = 1$, we see at once that

$$A(D) = \iint_D 1\,dx\,dy.$$

It follows that if D is a basic closed region, then D has Jordan content equal to the area of D as defined in Chapter 6.

THEOREM 12.20. *(Sum Rule) If f and g are continuous on D_0, then*

$$\iint_D (f(x,y) + g(x,y))\,dx\,dy = \iint_D f(x,y)\,dx\,dy + \iint_D g(x,y)\,dx\,dy.$$

This follows from the analogous formula for the double Riemann sum. We now introduce the notion of a smooth transformation.

DEFINITION 12.21. *Let D_0 be an open set in the (s,t) plane. A* **smooth transformation** *T on D_0 is a mapping*

$$T(s,t) = (g(s,t), h(s,t)) = (x,y)$$

of D_0 into the (x,y) plane such that each of the functions $x = g(s,t)$ and $y = h(s,t)$ has continuous partial derivatives on D_0.

The **image** *of a subset $D \subseteq D_0$ under T is the set*

$$T(D) = \{T(s,t) \colon (s,t) \in D\}.$$

The **Jacobian** *of a smooth transformation T, denoted by $J(T)$ or $\frac{\partial(x,y)}{\partial(s,t)}$, is the function*

$$\frac{\partial(x,y)}{\partial(s,t)} = \begin{vmatrix} \frac{\partial x}{\partial s} & \frac{\partial x}{\partial t} \\ \frac{\partial y}{\partial s} & \frac{\partial y}{\partial t} \end{vmatrix} = \frac{\partial x}{\partial s}\frac{\partial y}{\partial t} - \frac{\partial x}{\partial t}\frac{\partial y}{\partial s}.$$

Thus $\frac{\partial(x,y)}{\partial(s,t)}$ is a continuous function with domain D_0. For example, the polar coordinate transformation

$$x = r\cos\theta, \quad y = r\sin\theta$$

is smooth and its Jacobian is

$$\frac{\partial(x,y)}{\partial(s,t)} = \begin{vmatrix} cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r.$$

We will use the following two Lemmas which can be found in many advanced calculus books (for example, the book Buck [B]).

Lemma 12.22. *Let T be a smooth transformation which is one to one and has nonzero Jacobian on an open set D_0. Let D be a closed bounded subset of D_0 which has Jordan content. Then*
(i) $T(D_0)$ is an open set.
(ii) $T(D)$ is a closed bounded set which has Jordan content.
(iii) T maps the interior of D onto the interior of $T(D)$, and maps the boundary of D onto the boundary of $T(D)$.

Lemma 12.23. *Suppose T is a smooth transformation with nonzero Jacobian on an open set D_0, and the partial derivatives $\partial x/\partial s, \partial x/\partial t, \partial y/\partial s, \partial y \partial t$ are bounded on D_0. Then for every real $\varepsilon > 0$ there is a real $\delta > 0$ such that for every square $\triangle D \subseteq D_0$ with side $\triangle s < \delta$, the area of the image of $\triangle D$ is within $\varepsilon \triangle s^2$ of the Jacobian times the area of $\triangle D$, that is,*

$$|A(T(\triangle D)) - |J(T)|\triangle s^2| < \varepsilon \triangle s^2.$$

The number $|J(T)\triangle s^2|$ is the area of the parallelogram with vertex $T(s,t)$ and sides

$$\triangle s \left(\frac{\partial x}{\partial s}\mathbf{i} + \frac{\partial y}{\partial s}\mathbf{j} \right), \quad \triangle s \left(\frac{\partial x}{\partial t}\mathbf{i} + \frac{\partial y}{\partial t}\mathbf{j} \right).$$

The lemma is proved by showing that the boundaries of $T(\triangle D)$ are close to the boundaries of the parallelogram.

Theorem 12.24. *(Change of Variables) Let T be a smooth transformation which is one to one and has nonzero Jacobian on an open set D_0. Then for every continuous function $h(x,y)$ on $T(D_0)$ and every basic closed region $D \subseteq D_0$,*

$$\iint_{T(D)} h(x,y)\,dx\,dy = \iint_D h(T(s,t))|J(T)|\,ds\,dt.$$

For example, if D is a basic closed region

$$D = \{(r,\theta) \colon \alpha \le \theta \le \beta,\ a(\theta) \le r \le b(\theta)\}$$

where $\alpha < \beta < \alpha + 2\pi$ and $0 < a(\theta) \le b(\theta)$, then the polar coordinate transformation T satisfies the hypotheses of the theorem, and we obtain the formula

$$\iint_{T(D)} h(x,y)\,dx\,dy = \iint_D h(r\cos\theta, r\sin\theta) r\,dr\,d\theta.$$

We required $\beta < \alpha + 2\pi$ to make T one to one, and $0 < a(\theta)$ to make $r = J(T) \ne 0$ in D. But by passing to a limit we see that the polar integration formula is valid even for

$$\alpha \le \beta \le \alpha + 2\pi, \quad 0 \le a(\theta) \le b(\theta).$$

Proof of Theorem 12.24. Let D be a basic closed region contained in D_0. By Lemma 12.13 we may assume that the partial derivatives of x and y

are bounded on D_0. We may also assume that $h(T(s,t))$ is bounded on D_0, so that $h(x,y)$ is bounded on $T(D_0)$. We first give the proof in the case that $h(x,y) \geq 0$ on $T(D_0)$. Let $B(D_1)$ be defined for basic closed regions $D_1 \subseteq D_0$ by

$$B(D_1) = \iint_{T(D_1)} h(x,y)\,dx\,dy.$$

We want to prove that

$$B(D) = \iint_D h(T(s,t))|J(T)|\,ds\,dt. \tag{62}$$

To do this we will use the Infinite Sum Theorem 12.11. By the Cylinder Property, $B(D_1) \geq 0$ for all D_1. If $D_1 = D_2 \cup D_3$ where D_2 and D_3 have disjoint interiors, then $T(D_1) = T(D_2) \cup T(D_3)$, and by Lemma 12.22, $T(D_2)$ and $T(D_3)$ have disjoint interiors. Thus by the Addition Property for double integrals, $B(D_1) = B(D_2) + B(D_3)$. We have now shown that hypotheses (i) and (ii) of the Infinite Sum Theorem hold for $B(D)$. Hypothesis (iii) says that for at least one pair of positive infinitesimals $\triangle s, \triangle t$, for every element of area $\triangle D \subseteq (D_0)^*$ with lower left corner (s,t) and dimensions $\triangle s$ and $\triangle t$ we have

$$B(\triangle D) \approx h(T(s,t))|J(T)|\triangle s\triangle t \text{ (compared to } \triangle s\triangle t)\,.$$

Since we only need this for some pair of positive infinitesimals, it suffices to prove it for a pair of positive infinitesimals $\triangle s, \triangle t$ with $\triangle t = \triangle s$.

Suppose $\triangle s$ is positive infinitesimal, $\triangle D$ is an infinitesimal square of side $\triangle s$ with lower left corner (s,t), and $\triangle D$ is contained in D^*. By the Hyperreal Extreme Value Theorem 11.19, $h(T(u,v))$ has a minimum value m and a maximum value M for $(u,v) \in \triangle D$. Therefore $h(x,y)$ has minimum value m and maximum value M for $(x,y) \in T(\triangle D)$. Since T is continuous, every point of $T(\triangle D)$ is infinitely close to $T(s,t)$. By the continuity of f, m and M are infinitely close to $h(T(s,t))$. Then by the Cylinder Property and Transfer Axiom,

$$m \cdot A(T(\triangle D)) \leq B(\triangle D) \leq M \cdot A(T(\triangle D)),$$

whence

$$B(\triangle D) \approx h(T(s,t))A(T(\triangle D)) \text{ (compared to } A(T(\triangle D)))\,.$$

Therefore by Lemma 12.23,

$$\frac{B(\triangle D)}{\triangle s^2} = \frac{B(\triangle D)}{A(T(\triangle D))}\frac{A(T(\triangle D))}{\triangle s^2} \approx h(T(s,t))|J(T)|,$$

and

$$B(\triangle D) \approx h(T(s,t))|J(T)|\triangle s^2 \left(\text{compared to } \triangle s^2\right).$$

Hence by Infinite Sum Theorem,

$$B(D) = \iint_D h(T(s,t))|J(T)|\,ds\,dt.$$

Finally, we consider the general case where f is not necessarily ≥ 0 in D_0. Since f is bounded on D_0 there is a real number $c > 0$ such that $h(x, y) \geq -c$ on D_0. then $h(x, y) + c \geq 0$ on D_0, so by the previous case,

$$\iint_{T(D)} (h(x, y) + c)\, dx\, dy = \iint_D (h(T(s, t)) + c)|J(T)|\, ds\, dt,$$

$$\iint_{T(D)} c\, dx\, dy = \iint_D c|J(T)|\, ds\, dt.$$

Using the Sum Rule for double integrals, Theorem 12.20,

$$\begin{aligned}
\iint_{T(D)} h(x, y)\, dx\, dy &= \iint_{T(D)} (h(x, y) + c)\, dx\, dy - \iint_{T(D)} c\, dx\, dy \\
&= \iint_D (h(T(s, t)) + c)|J(T)|\, ds\, dt - \iint_D c\, |J(T)|\, ds\, dt \\
&= \iint_D h(T(s, t))|J(T)|\, ds\, dt.
\end{aligned}$$

$\dashv$

CHAPTER 13

VECTOR CALCULUS

13A. Line Integrals (§13.2)

In this section we will give a hyperreal characterization of the line integral. Recall that by Theorem 7.4, every smooth parametric curve has a reparametrization in which the curve length itself is the independent variable. We will take line integrals over parametric curves of this kind.

DEFINITION 13.1. *Let A and B be points in the (x, y) plane. A* **smooth curve** *from A to B is a real parametric curve*

$$C\colon x = g(s), \quad y = h(s), \quad s \in [0, L]$$

where $g(s)$ and $h(s)$ are continuously differentiable functions on the closed interval $[0, L]$,

$$(g(0), h(0)) = A, \quad (g(L), h(L)) = B,$$

*and s is the length of the part of the curve from A to the point $(x, y) = (g(s), h(s))$. We call A the **initial point** of C, and B the **terminal point** of C.*

In *Elementary Calculus* the line integral was defined as an ordinary definite integral as follows.

DEFINITION 13.2. *Let*

$$\mathbf{F}(x, y) = P(x, y)\mathbf{i} + Q(x, y)\mathbf{j}$$

be a continuous real vector valued function on an open rectangle D in the (x, y) plane which contains a smooth curve C. The **line integral** *of $\mathbf{F}$ along C, denoted by*

$$\int_C \mathbf{F} \cdot d\mathbf{S} = \int_C P\,dx + Q dy,$$

is defined as the definite integral

$$\int_0^L \left(P\frac{dx}{ds} + Q\frac{dy}{ds}\right) ds$$

where $x = g(s)$ and $y = h(s)$.

Let n be a positive integer and let $\triangle s = L/n$. For $k = 0, \dots, n-1$ let

$$\begin{aligned} s_k &= k\triangle s, \quad x_k = g(s_k), \quad y_k = h(s_k), \\ \triangle x_k &= x_{k+1} - x_k, \quad \triangle y_k = y_{k+1} - y_k, \\ \triangle \mathbf{S}_k &= \triangle x_k \mathbf{i} + \triangle y_k \mathbf{j}. \end{aligned}$$

The finite Riemann sum along C,

$$\sum_C \mathbf{F} \cdot \triangle \mathbf{S} = \sum_C P\triangle x + Q\triangle y,$$

is defined as the sum

$$\sum_{k=0}^{n} \mathbf{F}(x_k, y_k) \cdot \triangle \mathbf{S}_k = \sum_C P(x_k, y_k)\triangle x_k + Q(x_k, y_k)\triangle y_k.$$

This sum corresponds to a polygonal line connecting $n+1$ points along C. For a given curve C and vector valued function $\mathbf{F}$ it depends only on n, so we may write

$$I(n) = \sum_C \mathbf{F} \cdot \triangle \mathbf{S}.$$

By the Function Axiom, the natural extension $I(H)$ is defined for all positive hyperintegers H. When H is a positive infinite hyperinteger, the value $I(H)$ is called an **infinite Riemann sum along** C and is denoted by

$$I(H) = \sum_C \mathbf{F} \cdot d\mathbf{S} = \sum_C P\,dx + Qdy.$$

THEOREM 13.3. *Let $\mathbf{F}(x, y)$ be a continuous real vector valued function on an open rectangle D containing a smooth curve C. Let H be a positive infinite hyperinteger. Then the line integral of $\mathbf{F}$ along C is equal to the standard part of the infinite Riemann sum of $\mathbf{F}$ along C,*

$$\int_C \mathbf{F} \cdot d\mathbf{S} = \operatorname{st}\left(\sum_C \mathbf{F} \cdot d\mathbf{S}\right),$$

or

$$\int_C P\,dx + Qdy = \operatorname{st}\left(\sum_C P\,dx + Qdy\right).$$

PROOF. First consider a positive integer n and form the finite Riemann sum

$$\sum_C P\,dx + Qdy = \sum_C P(x_k, y_k)\triangle x_k + Q(x_k, y_k)\triangle y_k.$$

Let r be a positive real number. If the Riemann sum differs from the line integral by at least r, then on one of the subintervals the Riemann sum differs from the line integral by at least r/n. That is, every real solution of

$$n \in \mathbb{Z},\ 0 < n,\ \left|\sum_C P\triangle x + Q\triangle y - \int_C P\,dx + Qdy\right| \geq r \tag{63}$$

is a partial real solution of

$$\triangle s = L/n,\ 0 \le s < s + \triangle s \le L,\ x = g(s),\ y = h(s),$$

$$\triangle x = g(s+\triangle s) - g(s),\ \triangle y = h(s+\triangle s) - h(s), \tag{64}$$

$$\left| P(x,y)\triangle x + Q(x,y)\triangle y - \int_s^{s+\triangle s} (Pg' + Qh')\, ds \right| \ge r/n.$$

By the Partial Solution Theorem, every hyperreal solution of (63) is a partial solution of (64). Now form the infinite Riemann sum $\sum_C P\,dx + Qdy$ with respect to H. Assume that

$$\left| \sum_C P\,dx + Qdy - \int_C P\,dx + Qdy \right| \ge r. \tag{65}$$

Let $\triangle s = L/H$. Then there exist $0 \le s_1 < s_1 + \triangle s \le L$ such that, putting

$$x_1 = g(s_1), \quad y_1 = h(s_1),$$

$$\triangle x = g(s_1 + \triangle s) - g(s_1),\ \triangle y = h(s_1 + \triangle s) - h(s_1),$$

we have

$$\left| P(x_1,y_1)\triangle x + Q(s_1,y_1)\triangle y - \int_{s_1}^{s_1+\triangle s} (Pg' + Qh')\, ds \right| \ge \frac{r}{H}.$$

Dividing by $\triangle s$,

$$\left| P(x_1,y_1)\frac{\triangle x}{\triangle s} + Q(x_1,y_1)\frac{\triangle y}{\triangle s} - \frac{\int_{s_1}^{s_1+\triangle s}(Pg' + Qh')\, ds}{\triangle s} \right| \ge \frac{r}{L}. \tag{66}$$

Since $x = g(s), y = h(s)$, and $\int_0^s (Pg' + Qh')\, ds$ are continuously differentiable functions of s, we have

$$\frac{\triangle x}{\triangle s} \approx g'(s_1), \quad \frac{\triangle y}{\triangle s} \approx h'(s_1),$$

$$\frac{\int_{s_1}^{s_1+\triangle s}(Pg' + Qh')\, ds}{\triangle s} \approx P(x_1,y_1)g'(s_1) + Q(x_1,y_1)h'(s_1).$$

But then

$$P(x_1,y_1)\frac{\triangle x}{\triangle s} + Q(x_1,y_1)\frac{\triangle y}{\triangle s} \approx \frac{\int_{s_1}^{s_1+\triangle s}(Pg' + Qh')\, ds}{\triangle s},$$

contradicting (66). We conclude that (65) fails for every positive real r. It follows that

$$\sum_C P\,dx + Qdy \approx \int_C P\,dx + Qdy.$$

$\dashv$

A **piecewise smooth curve** is a parametric curve C which is a finite union of smooth curves $C = C_1 \cup \cdots \cup C_n$ such that the terminal point of C_k is the initial point of C_{k+1} for $k = 1, \ldots, n-1$. The line integral of $\mathbf{F}$ along a piecewise smooth curve C is defined as the sum

$$\int_C \mathbf{F} \cdot d\mathbf{S} = \int_{C_1} \mathbf{F} \cdot d\mathbf{S} + \cdots + \int_{C_n} \mathbf{F} \cdot d\mathbf{S}.$$

13B. Green's Theorem (§13.3, §13.4)

In this section we will prove Green's Theorem for basic closed regions with smooth boundaries, using the Infinite Sum Theorem for one variable. We need two lemmas about single integrals of functions with more than one variable.

Lemma 13.4. *Suppose E is a real open rectangular solid, $f(x, y, z)$ and $g(x, y, z)$ are continuous on E, the hyperreal points (a, y, z) and (b, y, z) belong to E^*, $b - a$ is finite, and*

$$f(x, y, z) \approx g(x, y, z) \text{ for all } x \in [a, b]^*.$$

Then

$$\int_a^b f(x, y, z)\, dx \approx \int_a^b g(x, y, z)\, dx.$$

Proof. For each real $r > 0$ we have

$$f(x, y, z) - r \le g(x, y, z) \le f(x, y, z) + r \text{ for all } x \in [a, b]^*.$$

By the rules of integrals in one variable and Transfer,

$$\int_a^b f(x, y, z)\, dx - r(b-a) \le \int_a^b g(x, y, z)\, dx \le \int_a^b f(x, y, z)\, dx + r(b-a).$$

Therefore

$$\left| \int_a^b f(x, y, z)\, dx - \int_a^b g(x, y, z)\, dx \right| \le r(b-a).$$

Since $b - a$ is finite and this holds for all real $r > 0$, we have

$$\int_a^b f(x, y, z)\, dx \approx \int_a^b g(x, y, z)\, dx.$$

$\dashv$

The following lemma was stated in Section §13.3 of *Elementary Calculus*, and used in the standard proof of the Path Independence Theorem for line integrals. A proof of the lemma was sketched there. We give a complete proof here.

LEMMA 13.5. *Suppose $P(x,y)$ is a smooth function on a real open rectangle D containing the point (a,b). Then whenever $(x,y) \in D$,*

$$\frac{\partial}{\partial x}\int_a^x P(t,y)\,dt = P(x,y)$$

and

$$\frac{\partial}{\partial y}\int_a^x P(t,y)\,dt = \int_a^x \frac{\partial P}{\partial y}(t,y)\,dt.$$

PROOF. The first formula follows at once from the Second Fundamental Theorem of Calculus, Theorem 4.17. We prove the second formula. Let

$$F(x,y) = \int_a^x P(t,y)\,dt, \quad (x,y) \in D.$$

Any real solution of

$$(x,y) \in D, \quad (x, y+\triangle y) \in D, \quad \triangle y \neq 0 \tag{67}$$

is a solution of

$$\frac{F(x,y+\triangle y) - F(x,y)}{\triangle y} = \int_a^x \frac{P(t,y+\triangle y) - P(t,y)}{\triangle y}\,dt. \tag{68}$$

By Transfer, any hyperreal solution of (67) is a solution of (68). Now let $\triangle y$ be positive infinitesimal. By Theorem 11.8, P is uniformly differentiable on D. Therefore, whenever $(x,y) \in D$ and $t \approx x$,

$$P(t,y+\triangle y) - P(t,y) \approx \frac{\partial P}{\partial y}(x,y)\triangle y \quad (\text{compared to } \triangle y),$$

so

$$\frac{P(t,y+\triangle y) - P(t,y)}{\triangle y} \approx \frac{\partial P}{\partial y}(x,y).$$

Since P is smooth on D, $\partial P/\partial y$ is continuous on D. Hence whenever $(x,y) \in D$ and $t \approx x$, the above formula holds with t in place of x,

$$\frac{P(t,y+\triangle y) - P(t,y)}{\triangle y} \approx \frac{\partial P}{\partial y}(t,y).$$

Since D is open, $(x,y+\triangle y) \in D^*$ whenever $(x,y) \in D$. The function

$$\frac{P(x,y+\triangle y) - P(x,y)}{\triangle y}$$

is continuous in the variables $(x,y,\triangle y)$ on $D \times (0,\infty)$. Now fix $(x,y) \in D$. By Lemma 13.4,

$$\int_a^x \frac{P(t,y+\triangle y) - P(t,y)}{\triangle y}\,dt \approx \int_a^x \frac{\partial P}{\partial y}(t,y)\,dt.$$

Then by (68),

$$\frac{F(x,y+\triangle y) - F(x,y)}{\triangle y} \approx \int_a^x \frac{\partial P}{\partial y}(t,y)\,dt.$$

Taking standard parts, we have

$$\frac{\partial F}{\partial y}(x,y) = \int_a^x \frac{\partial P}{\partial y}(t,y)\,dt.$$

$\dashv$

Given a basic closed region

$$D = \{(x,y)\colon a \le x \le b,\ g(x) \le y \le h(x)\},$$

whose upper and lower boundary curves $g(x)$ and $h(x)$ are smooth, the (counter-clockwise) **boundary curve** of D is the piecewise smooth curve

$$\partial D = C_1 \cup C_2 \cup C_3 \cup C_4$$

where

C_1 is the lower boundary curve of D moving from left to right,
C_2 it the right vertical boundary line of D moving upward,
C_3 is the upper boundary curve of D moving from right to left,
C_4 is the left vertical boundary line of D moving downward.

∂D is a closed curve, that is, its initial point and terminal point are the same, both equal to $(a, g(a))$.

The line integral of $P\,dx + Q\,dy$ around the boundary curve ∂D is denoted by

$$\oint_{\partial D} P\,dx + Q\,dy.$$

In *Elementary Calculus*, Green's theorem was stated for basic closed regions but only proved for rectangles. Here we prove the general case.

THEOREM 13.6. *(Green's Theorem) Suppose $P(x,y)$ and $Q(x,y)$ are smooth functions on an open set containing a basic closed region D with smooth upper and lower boundary curves. Then*

$$\oint_{\partial D} P\,dx + Q\,dy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\,dA.$$

PROOF. For $a \le u < v \le b$ let

$$B(u,v) = \oint_{\partial E(u,v)} P\,dx + Q\,dy$$

where

$$E(u,v) = \{(x,y) \in D\colon u \le x \le v\}$$

is the part of D with $x \in [u,v]$. Then B has the Addition Property

$$B(u,v) + B(v,w) = B(u,w)$$

because the right boundary of $E(u,v)$ is the same vertical line segment V as the left boundary of $E(v,w)$, and the upward line integral over V in $B(u,v)$

cancels the downward line integral over V in $B(v.w)$. Our plan is to show that for any infinitesimal subinterval $[x, x + \triangle x]^*$ of $[a, b]^*$,

$$\frac{\triangle B}{\triangle x} \approx P(x, g(x)) - P(x, h(x)) + \int_{g(x)}^{h(x)} \frac{\partial Q}{\partial x}(x, y)\, dy. \tag{69}$$

After (69) is verified, the proof is completed as follows. By (69) and the Infinite Sum Theorem 6.1,

$$B(a, b) = \int_a^b \left[P(x, g(x)) - P(x, h(x)) + \int_{g(x)}^{h(x)} \frac{\partial Q}{\partial x}(x, y)\, dy \right] dx.$$

By the Fundamental Theorem of Calculus 4.14,

$$P(x, g(x)) - P(x, h(x)) = \int_{g(x)}^{h(x)} -\frac{\partial P}{\partial y}\, dy$$

for each real $x \in [a, b]$. Therefore

$$B(a, b) = \int_a^b \int_{g(x)}^{h(x)} \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dy\, dx = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dA,$$

as required.

It remains to prove (69). First let $[x, x + \triangle x]$ be a real subinterval of $[a, b]$, that is,

$$a \le x < x + \triangle x \le b. \tag{70}$$

Then

$$\begin{aligned}
\frac{\triangle B}{\triangle x} = {} & \frac{1}{\triangle x} \int_x^{x+\triangle x} [P(t, g(t)) + Q(t, g(t))g'(t)]\, dt \\
& + \frac{1}{\triangle x} \int_{g(x+\triangle x)}^{h(x+\triangle x)} Q(x + \triangle x, y)\, dy \\
& - \frac{1}{\triangle x} \int_x^{x+\triangle x} [P(t, h(t)) + Q(t, h(t))h'(t)]\, dt \\
& - \frac{1}{\triangle x} \int_{g(x)}^{h(x)} Q(x, y)\, dy.
\end{aligned}$$

We rewrite this equation in the form

$$\frac{\triangle B}{\triangle x} = \frac{1}{\triangle x}\int_x^{x+\triangle x} [P(t,g(t)) - P(t,h(t))]\,dt \tag{71}$$

$$+\frac{1}{\triangle x}\int_x^{x+\triangle x} [Q(t,g(t))g'(t) - Q(t,h(t))h'(t))]\,dt$$

$$+\frac{1}{\triangle x}\int_{g(x)}^{h(x)} [Q(x+\triangle x,y) - Q(x,y)]\,dy$$

$$-\frac{1}{\triangle x}\int_{g(x)}^{g(x+\triangle x)} Q(x+\triangle x,y)\,dy$$

$$+\frac{1}{\triangle x}\int_{h(x)}^{h(x+\triangle x)} Q(x+\triangle x,y)\,dy.$$

By Transfer, every hyperreal solution of (70) is a solution of (71). Now let $[x, x+\triangle x]^*$ be an infinitesimal subinterval of $[a,b]^*$. Then $(x, x+\triangle x)$ is a hyperreal solution of (70), so it is also a solution of (71).

Let us now consider the third line of (71). Let $G(x,t)$ be the function

$$G(x,t) = \int_{g(x)}^{h(x)} Q(t,y)\,dy.$$

Then the third line of (71) is equal to the quotient

$$\frac{G(x,x+\triangle x) - G(x,x)}{\triangle x} = \frac{1}{\triangle x}\int_{g(x)}^{h(x)} [Q(x+\triangle x,y) - Q(x,y)]\,dy. \tag{72}$$

By Lemma 13.5, for x, t in $[a,b]$ we have

$$\frac{\partial}{\partial t} G(x,t) = \frac{\partial}{\partial t}\int_{g(x)}^{h(x)} Q(t,y)\,dy = \int_{g(x)}^{h(x)} \frac{\partial Q}{\partial t}(t,y)\,dy. \tag{73}$$

The right side of (73) is continuous in x and t. Since $g'(x)$ and $h'(x)$ are continuous, one can see from the Second Fundamental Theorem of Calculus that $\frac{\partial}{\partial x} G(x,t)$ is also continuous in x and t. Therefore by Theorem 11.8, $G(x,t)$ is uniformly differentiable on $[a,b]\times[a,b]$, and hence for all $t \approx x$ we have

$$\frac{G(x,x+\triangle x) - G(x,x)}{\triangle x} \approx \frac{\partial}{\partial t} G(x,t). \tag{74}$$

Then by (72)– (74), the third line of (71) has the infinitely close approximation

$$\frac{1}{\triangle x}\int_{g(x)}^{h(x)} [Q(x+\triangle x,y) - Q(x,y)]\,dy \approx \int_{g(x)}^{h(x)} \frac{\partial Q}{\partial x}(x,y)\,dy.$$

It follows from the Hyperreal Mean Value Theorem 3.34 that the sum of the second, fourth, and fifth lines of (71) is infinitesimal. Therefore, from (71) we obtain

$$\frac{\triangle B}{\triangle x} \approx P(x, g(x)) - P(x, h(x)) + \int_{g(x)}^{h(x)} \frac{\partial Q}{\partial x}(x, y)\, dy.$$

This completes the proof of (69). ⊣

CHAPTER 14

DIFFERENTIAL EQUATIONS

In *Elementary Calculus*, most of the material in Chapter 14 deals with ways to find explicit solutions of first and second order linear differential equations, using standard methods. However, infinitesimals are used extensively in Section 14.4, on the approximation, existence, and uniqueness of solutions of general first order differential equations with continuous coefficients. We will expand upon that section here.

We assume throughout this chapter that $f(t,y)$ is a real function which is continuous on $I \times \mathbb{R}$ for some open interval I containing t_0, and that $[t_0, T)$ is a half-open subinterval of I, where T is either a positive real number or ∞. A (first order) **differential equation** is an equation of the form

$$\frac{dy}{dt} = f(t,y).$$

An **initial value problem** is a differential equation together with an **initial value**,

$$\frac{dy}{dt} = f(t,y), \quad y(t_0) = y_0. \tag{75}$$

A **solution** of the initial value problem on $[t_0, T)$ is a function $y(t)$ with domain $[t_0, T)$ which satisfies (75) for all $t \in [t_0, T)$. It is helpful to think of the variable t as time.

Using the Fundamental Theorem of Calculus 4.14 and the Second Fundamental Theorem 4.17, (75) can also be written as an **integral equation**

$$y(t) = y_0 + \int_{t_0}^{t} f(s, y(s))\, ds. \tag{76}$$

All the results we present in this chapter can be generalized without difficulty to finite systems of differential equations, where f and y are vector valued functions of dimension n. We will confine our attention to solutions on intervals $[t_0, T)$, that is, solutions with time $t \geq t_0$. One can easily extend the results to solutions on intervals $(-T, t_0]$ by making a change of variables.

14A. Existence of Solutions (§14.4)

In this section we will use hyperfinite Euler approximations to prove that every initial value problem has at least one solution. At the end of the section we will draw a conclusion about standard Euler approximations of solutions.

Let $\triangle t > 0$ be real. By a **polygonal function** on $[t_0, T)$ with increment $\triangle t$ we will mean a real function $Y(t), t \in [t_0, T)$ such that for each natural number n, the graph of $Y(t)$ for $t_0 + n\triangle t \le t \le t_0 + (n+1)\triangle t$ is a straight line segment. Thus a polygonal function with increment $\triangle t$ is continuous and its graph is a broken line with corners at t_0 plus multiples of $\triangle t$. For a given increment $\triangle t$ and assignment of values at multiples of $\triangle t$, there is exactly one polygonal function.

The (standard) **Euler approximation** with initial value y_0 and increment $\triangle t$ for a differential equation $y'(t) = f(t, y)$ is the polygonal function $Y(t)$ on $[t_0, T)$ with increment $\triangle t$ such that $Y(t_0) = y_0$ and for each $t = t_0 + n\triangle t$, the slope on the interval $[t, t + \triangle t]$ is equal to the value $f(t, Y(t))$. That is, $Y(\cdot)$ is the polygonal function on $[t_0, T)$ with increment $\triangle t$ such that

$$Y(t_0) = y_0,\ Y(t + \triangle t) = Y(t) + f(t, Y(t))\triangle t \text{ for } t = t_0 + n\triangle t.$$

It is defined for each natural number n by induction, starting with

$$Y(t_0) = y_0,\ Y(t_0 + \triangle t) = y_0 + f(t_0, y_0)\triangle t.$$

Thus for each positive integer n, $Y(t_0 + n\triangle t)$ is equal to the finite sum

$$Y(t_0 + n\triangle t) = y_0 + \sum_{k=0}^{n-1} f(t_0 + k\triangle t, Y(t_0 + k\triangle t))\triangle t.$$

Following the notation for Riemann sums, we may write this as

$$Y(s) = y_0 + \sum_{t_0}^{s} f(t, Y(t))\triangle t.$$

For a given function $f(t, y)$, let us denote the Euler approximation on $[t_0, T)$ with initial value y_0 and increment $\triangle t$ by $Y_{y_0,\triangle t}$. $Y_{y_0,\triangle t}(t)$ is a function of three real variables $(y_0, \triangle t, t)$. By the Function Axiom, its natural extension is a hyperreal function of three hyperreal variables. We call the Euler approximation $Y_{z,dt}$ on $[t_0, T)^*$ with infinitesimal increment dt and an initial value $z \approx y_0$ a **hyperfinite Euler approximation** of the initial value problem (75), and write it as an infinite sum

$$Y(s) = z + \sum_{t_0}^{s} f(t, Y(t))\, dt.$$

For each z and dt, $Y_{z,dt}(\cdot)$ is a hyperreal function of t with domain $[t_0, T)^*$.

In *Elementary Calculus*, we only considered hyperfinite Euler approximations with the standard initial value y_0. Here we will give a fuller treatment, and it will be useful to allow hyperreal initial values as well. One advantage of doing this is that one can sometimes get a different solution of the initial value problem by making an infinitesimal change in the initial value in an Euler approximation, as in Section 14C of this chapter.

Given a hyperreal function $Y(\cdot)$ with domain $[t_0, T)^*$, a real function $y(\cdot)$ is said to be the **standard part** of $Y(\cdot)$ on $[t_0, T)$ if whenever $t \in [t_0, T)^*$, $s \in [t_0, T)$, and $t \approx s$ we have $Y(t) \approx y(s)$. That is, the standard part of $Y(t)$ exists and depends only on the standard part of t, and we have $y(\mathrm{st}(t)) = \mathrm{st}(Y(t))$.

The next theorem shows that as long as a hyperfinite Euler approximation is finite, its standard part exists and is a solution of the initial value problem on some interval $[t_0, T)$. The special case of the theorem with $z = y_0$ was already proved in *Elementary Calculus.*

THEOREM 14.1. *Let $Y(t)$ be the hyperfinite Euler approximation of the initial value problem*

$$\frac{dy}{dt} = f(t, y), \quad y(t_0) = y_0$$

with initial value z and positive infinitesimal increment dt. Suppose $Y(t)$ is finite whenever $\mathrm{st}(t) \in [t_0, T)$. Then the standard part of $Y(\cdot)$ exists and is a solution of the initial value problem on the interval $[t_0, T)$.

PROOF. We will show that the standard part of $Y(\cdot)$ on $[t_0, T)$ exists and is a solution of the of the equivalent integral equation (76) on $[t_0, T)$. It is enough to show that for each positive real number $c \in (t_0, T)$, the standard part of $Y(\cdot)$ exists and is a solution of (76) on $[t_0, c)$. Take an arbitrary real $c \in (t_0, T)$. For each real $s \in [t_0, c]$ let $y(s) = \mathrm{st}(Y(s))$. $y(\cdot)$ is a real function which has a natural extension $y^*(\cdot)$ to $[t_0, c]^*$. The main steps are to prove the following for each $t \in [t_0, c]^*$:

$$\mathrm{st}(Y(t)) = \mathrm{st}(y(t)) = y(\mathrm{st}(t)) \tag{77}$$

and

$$\mathrm{st}(Y(t)) = y_0 + \mathrm{st}\left(\sum_{t_0}^{t} f(s, y(s))\, dt\right). \tag{78}$$

It follows from (77) that $y(\cdot)$ is the standard part of $Y(\cdot)$ on $[t_0, c]$. The right side of (78) is the infinite Riemann sum of $f(s, y(s))$, and it follows that

$$y(t) = y_0 + \int_{t_0}^{t} f(s, y(s))\, dt$$

for all $t \in [t_0, c]$.

Proof of (77): By the Extreme Value Theorem, for each real initial value u and positive real increment $\triangle t$, the continuous function $|f(t, Y_{u,\triangle t}(t))|$ has a maximum $N_{u,\triangle t} \in [t_0, c]$, and a maximum value $M_{u,\triangle t}$. $Y_{u,\triangle t}(t)$ never changes by more than $\triangle t\, M_{u,\triangle t}$ over a subinterval of $[t_0, c]$ of length $\triangle t$, so whenever $s, t \in [t_0, c]$ we have

$$|Y_{u,\triangle t}(t) - Y_{u,\triangle t}(s)| \leq M_{u,\triangle t}|t - s|.$$

By Transfer, $|f(t, Y(t))|$ has a maximum at $N_{z,dt}$ in $[t_0, c]^*$ and a maximum value $M_{z,dt}$, and whenever $s, t \in [t_0, c]^*$ we have

$$|Y(t) - Y(s)| \leq M_{z,dt}|t - s|. \tag{79}$$

Since $Y(t)$ is finite whenever $\mathrm{st}(t) \in [t_0, T)$, it is finite whenever $t \in [t_0, c]^*$. Therefore the maximum value $M_{z,dt} = |f(N_{z,dt}, Y(N_{z,dt}))|$ is finite. Thus whenever $s = \mathrm{st}(t) \in (t_0, c]$,

$$\mathrm{st}(Y(t)) = st(Y(s)) = y(s).$$

Taking standard parts in the inequality (79), we see that for all real $s, t \in [t_0, c]$,

$$|y(t) - y(s)| \leq m|t - s| \text{ where } m = \mathrm{st}(M_{z,dt}).$$

By Transfer, for all hyperreal $s, t \in [t_0, c]^*$,

$$|y(t) - y(s)| \leq m|t - s|.$$

Then whenever $t \in [t_0, c]^*$ and $s = \mathrm{st}(t)$,

$$\mathrm{st}(Y(t)) = y(s) = \mathrm{st}(y(s)) = \mathrm{st}(y(t)).$$

This proves (77).

By (77), the function $y(\cdot)$ is continuous on $[t_0, c]$. By the Extreme Value Theorem, for each positive real $\triangle t$ and initial value u the difference

$$|f(t, Y_{u,\triangle t}(t)) - f(t, y(t))|$$

has a maximum at some point $J_{u,\triangle t} \in [t_0, c]$ and a maximum value $K_{u,\triangle t}$. By Transfer,

$$|f(t, Y_{z,dt}(t)) - f(t, y(t))|$$

has a maximum at $J_{z,dt}$ in $[t_0, c]^*$ and a maximum value $K_{z,dt}$. By (77) and the continuity of f, $K_{z,dt}$ is infinitesimal. For each positive real $\triangle t$ and real u, and each $t \in [t_0, c]$, we have

$$\left| Y_{u,\triangle t}(t) - \left(u + \sum_{t_0}^{t} f(s, y(s))\triangle t \right) \right| \leq K_{u,\triangle t}(t - t_0).$$

By Transfer, for each $t \in [t_0, c]^*$ we have

$$\left| Y_{z,dt}(t) - \left(z + \sum_{t_0}^{t} f(s, y(s))\, dt \right) \right| \leq K_{z,dt}(t - r).$$

Since $K_{z,dt}$ is infinitesimal, this proves (78). $\dashv$

To apply the above theorem, we need a convenient criterion for a hyperfinite Euler approximation to be finite. The next theorem gives one.

Theorem 14.2. *We are given an initial value problem*

$$\frac{dy}{dt} = f(t,y), \quad y(t_0) = y_0.$$

Let M and ε be positive reals.

(i) Suppose

$$|f(t,y)| \le M \text{ whenever } t \in [t_0,T) \text{ and } |y - y_0| \le M(t-t_0) + \varepsilon.$$

Then for every hyperfinite Euler approximation $Y(\cdot)$ with initial value $z \approx y_0$, the standard part of $Y(\cdot)$ exists and is a solution of the initial value problem on the interval $[t_0,T)$.

(ii) Suppose

$$|f(t,y)| \le M \text{ whenever } t \in [t_0,T) \text{ and } |y - y_0| \le M(t-t_0).$$

Then for every hyperfinite Euler approximation $Y(\cdot)$ with initial value y_0, the standard part of $Y(\cdot)$ exists and is a solution of the initial value problem on the interval $[t_0,T)$.

Proof. (i) We show that $Y(t)$ is finite whenever $\mathrm{st}(t) \in [t_0,T)$. and then use Theorem 14.1. Suppose c, u are real numbers such that $t_0 < c < T$, u is within ε of y_0, and $\triangle t > 0$. Let $U = Y_{u,\triangle t}$. We show by induction on n that

(80) if $t = t_0 + n\triangle t \le c$ then $|U(t) - u| \le M(t - t_0)$.

This is true for $n = 0$ because $U(t_0) = u$. Assume that (80) is true for n and let $s = t_0 + n\triangle t$, $t = s + \triangle t$. Then $t = t_0 + (n+1)\triangle t$. Suppose $t \le c$. Then $s \le c$, so by the induction hypothesis we have $|U(s) - u| \le M(s - t_0)$. But u is within ε of y_0, so

$$|U(s) - y_0| \le M(s - t_0) + \varepsilon,$$

and thus $|f(s, U(s))| \le M$. By definition,

$$U(t) = U(s + \triangle t) = U(s) + f(s, U(s))\triangle t.$$

Therefore

$$|U(t) - u| \le |U(s) - u| + |f(s, U(s))|\triangle t \le M(s - t_0) + M\triangle t = M(t - t_0)$$

as required. This completes the induction. Since U is a polygonal function, it follows that

$$|U(t) - u| \le M(t - t_0) \text{ for all } t \in [t_0, c].$$

Since $z \approx y_0$, z is within ε of y_0. By Transfer, $|Y(t) - z| \le M(t - t_0)$ and hence $Y(t)$ is finite whenever $t \in [t_0, c]^*$. Since this holds for all $c \in (t_0, T)$, $Y(t)$ is finite whenever $\mathrm{st}(t) \in [t_0, T)$, as required.

The proof of (ii) is similar, but with $\varepsilon = 0$ and $z = y_0$. ⊣

Corollary 14.3. *(Peano Existence Theorem) For some positive real c, the initial value problem*

$$\frac{dy}{dt} = f(t,y), \quad y(0) = y_0$$

has a solution on $[t_0, c)$.

Proof. Choose real $a \in (t_0, T)$ and $b > 0$. By the Extreme Value Theorem, $|f(t, y)|$ has a maximum value M on the closed rectangle $[t_0, a] \times [y_0 - b, y_0 + b]$. Let $c = \min(a, b/M)$. Then

$$|f(t, y)| \leq M \text{ whenever } t \in [t_0, c] \text{ and } |y - y_0| \leq M(t - t_0)$$

because when $t \in [t_0, c]$, $t \in [t_0, a]$ and $M(t - t_0) \leq Mc \leq b$. By Theorem 14.2, the standard part of a hyperfinite Euler approximation with initial value y_0 is a solution on $[t_0, c]$. ⊣

Corollary 14.4. *If $f(t, y)$ is bounded on $[t_0, T) \times \mathbb{R}$ then for every y_0 the initial value problem has a solution on $[t_0, T)$.*

If $y(t)$ is a solution of the initial value problem and is the standard part of some hyperfinite Euler approximation on $[t_0, T)$, we say that $y(t)$ is **Euler approximable** on $[t_0, T)$. Theorem 14.2 shows that every initial value problem has an Euler approximable solution on some subinterval $[t_0, c)$ of $[t_0, T)$. The preceding two corollaries also give Euler approximable solutions.

We now show that any Euler approximable solution of an initial value problem has the property that it is within ε of a standard Euler approximation for each real $\varepsilon > 0$.

Proposition 14.5. *Given an initial value problem*

$$\frac{dy}{dt} = f(t, y), \quad y(0) = y_0,$$

let $y(\cdot)$ be an Euler approximable solution on $[t_0, T)$. Then $y(\cdot)$ has the following standard property: for each real $\varepsilon > 0$ and $c < T$ there are real numbers $u, \triangle t$ such that u is within ε of y_0, $\triangle t \in (0, \varepsilon)$, and the Euler approximation $Y_{u,\triangle t}(t)$ is within ε of $y(t)$ for all $t \in [t_0, c]$.

Proof. Since $y(\cdot)$ is Euler approximable, there is a hyperreal value $z \approx y_0$ and infinitesimal $dt > 0$ such that $y(\cdot)$ is the standard part of the hyperfinite Euler approximation $Y_{z,dt}(\cdot)$ on $[t_0, T)$. Suppose that the property fails for some $\varepsilon > 0$ and $c < T$. Then every real solution of

$$u = u, \quad 0 < \triangle t < \varepsilon$$

is a partial real solution of

$$t \in [t_0, c], \quad |Y_{u,\triangle t}(t) - y(t)| > \varepsilon.$$

By the Partial Solution Theorem, for every hyperreal $z \approx y_0$ and infinitesimal $dt > 0$, there exists $t \in [t_0, c]^*$ such that

$$|Y_{z,dt}(t) - y^*(t)| > \varepsilon.$$

But $y(\cdot)$ is continuous on $[t_0, c]$, so $y(\mathrm{st}(t)) \approx \mathrm{st}(y^*(t))$, and hence

$$|Y_{z,dt}(t) - y(\mathrm{st}(t))| > \varepsilon.$$

This contradicts the fact that $y(\cdot)$ is the standard part of $Y_{z,dt}(\cdot)$ on $[t_0, T)$, and completes the proof. ⊣

The above result also has a converse which we will not prove here. It says that any solution of the initial value problem which satisfies the property in the theorem is Euler approximable.

In the next section we will see that if the function f is nice then the solution of the initial value problem is unique and Euler approximable. In the last section we will examine in detail an example of an initial value problem where there are many solutions, but every solution is Euler approximable.

14B. Uniqueness of Solutions (§14.4)

Definition 14.6. *We say that the function f has* **Lipschitz bound** L *(in y) on a rectangle E if for all points (t, y) and (t, z) in E,*

$$|f(t, y) - f(t, z)| \leq L|y - z|.$$

We say that f is **locally Lipschitz** *on a rectangle E if f has a finite Lipschitz bound on each bounded closed rectangle $D \subseteq E$.*

Proposition 14.7. *If the partial derivative $f_y(t, y)$ exists and is continuous on an open rectangle E, then f is locally Lipschitz on E.*

Proof. Let D be a bounded closed rectangle contained in E. By the Extreme Value Theorem, $|f_y(t, y)|$ has a finite bound L on D. Take (t, y) and (t, z) in D, with $y < z$. By the Mean Value Theorem 3.30, there is a point $u \in (y, z)$ such that

$$\frac{f(t, z) - f(t, y)}{z - y} = f_y(t, u).$$

The point (t, u) also belongs to D. Therefore $|f_y(t, u)| \leq L$, and hence

$$|f(t, z) - f(t, y)| \leq L|z - y|,$$

and f has Lipschitz bound L on D. ⊣

Lemma 14.8. *Suppose $L > 0$, $g(t)$ is continuous on $[t_0, c]$, and for all $t \in [t_0, c]$,*

$$0 \leq g(t) \leq L \int_{t_0}^{t} g(s)\, ds.$$

Then $g(t) = 0$ for all $t \in [t_0, c]$.

Proof. Let $h(t) = \int_{t_0}^{t} g(s)\, ds$. Then $h(t_0) = 0$, and by the Second Fundamental Theorem of Calculus, $h'(t) = g(t)$ for $t \in (t_0, c)$. By hypothesis, $g(t) - Lh(t) \leq 0$, so $h'(t) - Lh(t) \leq 0$. By the Product Rule for derivatives,

$$\frac{d}{dt}\left(e^{-Lt}h(t)\right) = e^{-Lt}h'(t) - Le^{-Lt}h(t) = e^{-Lt}(h'(t) - Lh(t)).$$

Thus $e^{-Lt}h(t)$ is continuous on $[t_0, c]$ and has derivative ≤ 0 on (t_0, c). By Corollary 3.31, $0 = e^{-Lt_0}h(t_0) \geq e^{-Lt}h(t) \geq 0$ and hence $h(t) = 0$ for all $t \in [t_0, c]$. Therefore $g(t) = 0$ for all $t \in [t_0, c]$. ⊣

THEOREM 14.9. *(Uniqueness Theorem) Suppose f is locally Lipschitz on $[t_0, T) \times \mathbb{R}$. Then the initial value problem*

$$\frac{dy}{dt} = f(t, y), \quad y(t_0) = y_0$$

has at most one solution on $[t_0, T)$. It also has at most one solution on $[t_0, S) \times \mathbb{R}$ where $0 < S \leq T$.

PROOF. Since f is locally Lipschitz on $[t_0, T) \times \mathbb{R}$, it is also locally Lipschitz on $[t_0, S) \times \mathbb{R}$. Let $x(t)$ and $y(t)$ be two solutions and let $t_0 < c < S$. By the Extreme Value Theorem, $|x(t)|$ and $|y(t)|$ have maximum values for $t \in [t_0, c]$, so there is a closed rectangle $D \subseteq [t_0, S) \times \mathbb{R}$ which contains the graphs of $x(t)$ and $y(t)$. By hypothesis, f has a finite Lipschitz bound L on D. Let $g(t) = |x(t) - y(t)|$. Then g is continuous on $[t_0, c]$. For all $t \in [t_0, c]$,

$$g(t) = |x(t) - y(t)| = \left| \int_{t_0}^{t} f(s, x(s))\, ds - \int_{t_0}^{t} f(s, y(s))\, ds \right|$$

$$= \left| \int_{t_0}^{t} (f(s, x(s)) - f(s, y(s)))\, ds \right| \leq \int_{t_0}^{t} |f(s, x(s)) - f(s, y(s))|\, ds$$

$$\leq \int_{t_0}^{t} L\, |x(s) - y(s)|\, ds = \int_{t_0}^{t} Lg(s)\, ds.$$

Thus by Lemma 14.8, $g(t) = 0$ for all $t \in [t_0, c]$. Since this holds whenever $t_0 < c < S$, $x(t) = y(t)$ for all $t \in [t_0, S)$. ⊣

COROLLARY 14.10. *Suppose $f_y(t, y)$ exists and is continuous on $I \times \mathbb{R}$. Then the initial value problem*

$$\frac{dy}{dt} = f(t, y), \quad y(t_0) = y_0$$

has at most one solution on $[t_0, T)$.

PROOF. By Proposition 14.7 and Theorem 14.9. ⊣

COROLLARY 14.11. *Let J be an open interval which contains the initial value y_0, and suppose that $f_y(t, y)$ exists and is continuous on $I \times J$. Then the initial value problem*

$$\frac{dy}{dt} = f(t, y), \quad y(t_0) = y_0$$

has at most one solution $y(\cdot)$ on $[t_0, T)$ such that $y(t) \in J$ for all $t \in [t_0, T)$.

PROOF. Let $x(t)$ and $y(t)$ be two solutions such that $x(t) \in J$ and $y(t) \in J$ for all $t \in [t_0, T)$, and let $t_0 < c < T$. By the Extreme Value Theorem, $x(t)$ and $y(t)$ have maximum and minimum values for $t \in [t_0, c]$, so there is a closed interval $[a,b] \subseteq J$ such that $x(t) \in [a,b]$ and $y(t) \in [a,b]$ for all $t \in [t_0, c]$. Let $g(t,y)$ be the function defined by

$$g(t,y) = \begin{cases} f(t,b) & \text{if } y > b \\ f(t,y) & \text{if } y \in [a,b] \\ f(t,a) & \text{if } y < a \end{cases}$$

Since g agrees with f on $[t_0, c] \times [a,b]$, $x(\cdot)$ and $y(\cdot)$ are solutions of the initial value problem

$$\frac{dy}{dt} = g(t,y), \quad y(t_0) = y_0 \tag{81}$$

on $[t_0, c)$. Observe that g is continuous on $I \times \mathbb{R}$. By Proposition 14.7, f is locally Lipschitz on $I \times J$, and hence g is locally Lipschitz on $I \times \mathbb{R}$. By Theorem 14.9, there is at most one solution of (81) on $[t_0, c)$, so $x(t) = y(t)$ for all $t \in [t_0, c)$. Since this holds for all $c \in (t_0, T)$, $x(t) = y(t)$ for all $t \in [t_0, T)$. ⊣

Corollary 14.12. *Suppose that for each $c \in (t_0, T)$ the initial value problem (75) has at most one solution on $[t_0, c)$. Then there is an S (a real number or ∞) such that $t_0 < S \leq T$ and:*

(i) There is a unique solution on $[t_0, R)$ when $R \leq S$, and

(ii) There is no solution on $[t_0, R)$ when $S < R \leq T$.

Proof. Let J be the set of all $c \in (t_0, T)$ such that either $c = t_0$ or there exists a solution $y_c(\cdot)$ on $[t_0, c)$. For each $c \in J$ and $b \in (0, c)$, the restriction of $y_c(\cdot)$ to $[t_0, b)$ is a solution on $[0, b)$, and therefore $b \in J$. We conclude that $\{t_0\} \cup J$ is an interval $[t_0, S) \subseteq [t_0, T)$ for some $S \leq T$. The Peano Existence Theorem 14.3 shows that $0 < S$.

Since there is at most one solution on $[t_0, b)$ for each $b \in (t_0, S)$, $y_b(t) = y_c(t)$ whenever $t_0 \leq t < b < c < S$. Thus the union

$$y(\cdot) = \bigcup \{y_c(\cdot) \colon c \in J\}$$

is a function on $[t_0, S)$ and is the unique solution on $[t_0, S)$. This proves (i).

If $S < R \leq T$, then there is a real c with $S < c < R$. Then c does not belong to J, so there is no solution on $[t_0, c)$ and hence no solution on $[t_0, R)$. This proves (ii). ⊣

Corollary 14.13. *Suppose that for each $c \in (t_0, T)$ the initial value problem (75) has exactly one solution on $[t_0, c)$. Then it has exactly one solution on $[t_0, T)$.*

Proof. In this case, we must have $S = T$ in Corollary 14.12 ⊣

We now show that if there is a unique solution, then the solution is also Euler approximable.

Theorem 14.14. *Suppose the initial value problem (75) has the unique solution $y(\cdot)$ on $[t_0, T)$, and has only one solution on $[t_0, c)$ for each $c \in (t_0, T)$. Then $y(t)$ is Euler approximable and is the standard part of every hyperfinite Euler approximation of (75) on $[t_0, T)$.*

Proof. Let $Y(t)$ be the hyperfinite Euler approximation with initial value $z_0 \approx y_0$ and infinitesimal increment dt. It is enough to show that $Y(t)$ is finite whenever $\mathrm{st}(t) \in [t_0, T)$, because then by Theorem 14.1, the standard part of $Y(\cdot)$ is a solution on $[t_0, T)$ and is Euler approximable.

Take a real number $c \in (t_0, T)$ and let $c < d < T$. By assumption, the initial value problem has exactly one solution $x(t)$ on $[t_0, d)$. Then the function $f(t, x(t))$ is continuous for $t \in [t_0, c]$, and by the Extreme Value Theorem there is a positive real N such that $|f(t, x(t))| \le N$ for all $t \in [t_0, c]$. Let $g(t, y)$ be the function defined by

$$g(t,y) = \begin{cases} N+1 & \text{if } f(t,y) > N+1 \\ f(t,y) & \text{if } |f(t,y)| \le N+1 \\ -(N+1) & \text{if } f(t,y) < -(N+1) \end{cases}$$

Then for all $(t, y) \in \times\mathbb{R}$, $g(t, y)$ is continuous, $|g(t, y)| \le N + 1$, and $g(t, y) = f(t, y)$ whenever $|f(t, y)| \le N + 1$. Let $Z(\cdot)$ be the hyperfinite Euler approximation for the initial value problem

$$\frac{dy}{dt} = g(t,y), \quad y(t_0) = y_0 \tag{82}$$

with initial value z_0 and increment dt. By Theorem 14.2, the standard part $z(\cdot)$ of $Z(\cdot)$ exists and is a solution of the new initial value problem (82) on the interval $[t_0, T)$.

We now show that $z(t) = x(t)$ for all $t \in [t_0, c)$. Suppose not, and let v be the greatest lower bound of all $t \in [t_0, c)$ such that $z(t) \ne x(t)$. Then $t_0 \le v < c$, $z(t_0) = x(t_0) = y_0$, and for all $t \in [t_0, v)$ we have $z(t) = x(t)$ and $|f(t, z(t))| \le N$. Since $z(\cdot)$ and $x(\cdot)$ are continuous on $[t_0, c)$, we also have $z(v) = x(v)$ and $|f(v, z(v))| \le N$. $f(t, z(t))$ is also continuous on $[t_0, c)$, so there is an $s \in (v, c]$ such that $|f(t, z(t))| \le N + 1$ for all $t \in [t_0, s)$. Therefore $g(t, z(t)) = f(t, z(t))$ for all $t \in [t_0, s)$, and hence $z(\cdot)$ is a solution of the original initial value problem (75) on $[t_0, s)$. But by hypothesis there is exactly one such solution on $[t_0, s)$, so we must have $z(t) = x(t)$ for all $t \in [t_0, s)$. This contradicts the fact that $s > v$, and proves that $z(t) = x(t)$ for all $t \in [t_0, c)$.

We now have $|f(t, z(t))| \le N$ for all $t \in [t_0, c)$. Now consider hyperreal points $(t_2, y_2) \in [r, c]^* \times \mathbb{R}^*$. Since $z(\cdot)$ is the standard part of $Z(\cdot)$ on $[t_0, c)$, $|f(t_2, Z(t_2))| \le N + 1$ whenever $\mathrm{st}(t_2) \in [t_0, c)$. By Transfer, $g(t_2, y_2) = f(t_2, y_2)$ for all hyperreal t_2, y_2 such that $|f(t_2, y_2)| \le N + 1$. Hence $f(t_2, Z(t_2)) = g(t_2, Z(t_2))$ whenever $\mathrm{st}(t_2) \in [t_0, c)$. It follows that $Z(t_2) = Y(t_2)$ whenever $\mathrm{st}(t_2) \in [t_0, c)$. Therefore $Y(\cdot)$ has the standard part

$z(\cdot)$ on $[t_0, c)$, so $Y(t_2)$ is finite whenever $\mathrm{st}(t_2) \in [t_0, c)$. Since c was an arbitrary real number less than T, $Y(t_2)$ is finite whenever $\mathrm{st}(t_2) \in [t_0, T)$. This completes the proof. ⊣

We conclude this section by showing that the unique solution $y(\cdot)$ in the preceding theorem has a ε, δ property involving standard Euler approximations.

Proposition 14.15. *Assume the hypotheses of Theorem 14.14 and let $y(\cdot)$ be the unique solution of the initial value problem on $[t_0, T)$. Then $y(\cdot)$ has the following standard property: for each real $\varepsilon > 0$ and $c < T$ there is a real $\delta > 0$ such that for every Euler approximation $Y_{u,\triangle t}(\cdot)$ with $|u - y_0| < \delta$ and $0 < \triangle t < \delta$, $Y_{u,\triangle t}(t)$ is within ε of $y(t)$ for all $t \in [t_0, c]$.*

Proof. Suppose the property fails for some $\varepsilon > 0$ and $c < T$. For each real u and $\triangle t > 0$, let $M(u, \triangle t)$ be the maximum value of $|Y_{u,\triangle t}(t) - y(t)|$ for $t \in [t_0, c]$. This maximum exists by the Extreme Value Theorem, because $|Y_{u,\triangle t}(t) - y(t)|$ is continuous on $[t_0, c]$. Since the property fails for ε and c, for every real $\delta > 0$ there exist u and $\triangle t$ such that

$$|u - y_0| < \delta, \quad 0 < \triangle t < \delta, \quad M(u, \triangle t) \geq \varepsilon.$$

By the Partial Solution Theorem, for each positive infinitesimal $\sigma > 0$ there exist hyperreal z and dt such that

$$|z - y_0| < \sigma, \quad 0 < dt < \sigma, \quad M(z, dt) \geq \varepsilon.$$

Using the Partial Solution Theorem again, $M(z, dt)$ is the maximum value of $|Y_{z,dt}(t) - y^*(t)|$ for $t \in [t_0, c]^*$. Then $z \approx y_0$ and dt is positive infinitesimal, so $Y_{z,dt}(\cdot)$ is a hyperfinite Euler approximation but $y(\cdot)$ is not the standard part of $Y_{z,dt}(\cdot)$ on $[t_0, T)$. This contradicts Theorem 14.14, so the property must hold after all. ⊣

14C. An Example where Uniqueness Fails (§14.3)

In *Elementary Calculus* the initial value problem

$$y' = 3y^{2/3}, \quad y(0) = 0$$

was presented as an example with infinitely many solutions. Here we will examine this example more closely, and show that every solution is Euler approximable.

For every real number a, let $y_a(t)$ be the function

$$y_a(t) = \begin{cases} 0 & \text{for } t < a \\ (t-a)^3 & \text{for } t \geq a \end{cases}$$

with domain $[0, \infty)$.

PROPOSITION 14.16. *The family of all solutions of the initial value problem*

$$y' = 3y^{2/3}, \quad y(0) = 0$$

on $[0, \infty)$ *consists of the constant zero function*

$$y(t) = 0 \text{ for all } t \in [0, \infty)$$

and the functions $y_a(\cdot)$ *where* $a \geq 0$.

PROOF. The constant zero function is a solution because its derivative is everywhere zero. For each real a, one can check by differentiation that $y_a'(t) = 0$ whenever $t \leq a$, and $y_a'(t) = 3(t-a)^2 = 3(y_a(t))^{2/3}$ for $t > a$. When $a \geq 0$, $y_a(0) = 0$, so $y_a(\cdot)$ is a solution of the given initial value problem. When $a < 0$, we instead have $y_a(0) = -a^3 > 0$, so in this case $y_a(\cdot)$ is not a solution.

We now show that these are the only solutions. Let $x(\cdot)$ be a solution of the initial value problem, and suppose that $x(\cdot)$ is not the constant zero function. Since $f(t, y) = 3y^{2/3} \geq 0$ for all (t, y), $x(t)$ must be non-decreasing, that is, $x(s) \leq x(t)$ whenever $s \leq t$. We also must have $x(0) = 0$, so $x(t) \geq 0$ for all t. By Theorem 14.11, whenever $b > 0$ and $x(b) > 0$, the initial value problem

$$y' = 3y^{2/3}, \quad y(b) = x(b)$$

has at most one solution on $[b, \infty)$. But it does have one solution on $[b, \infty)$, namely the function $y_a(\cdot)$ where $(b-a)^3 = x(b)$, that is, $a = b - ((x(b))^{1/3}$. It follows from uniqueness that we get the same value a for each starting point $(b, x(b))$ on the curve. Thus for some $a \geq 0$, we have $x(t) = y_a(t)$ whenever $x(t) > 0$. Since $x(\cdot)$ and $y_a(\cdot)$ are continuous, non-decreasing, and map 0 to 0, we must have $x(t) = y_a(t)$ for all t. ⊣

The following result is given in [AFHL 1986], page 32, and is attributed to B. Birkeland and D. Normann.

PROPOSITION 14.17. *Every solution of the initial value problem*

$$y' = 3y^{2/3}, \quad y(0) = 0$$

is Euler approximable. In fact, for each positive infinitesimal dt, every solution is the standard part of a hyperfinite Euler approximation with increment dt and initial value z *with* $z \geq 0$.

PROOF. We first consider the standard Euler approximations $Y_{u,\triangle t}(\cdot)$ with initial value u with $u \geq 0$ and increment $\triangle t > 0$. Since $3y^{2/3} \geq 0$, $Y_{u,\triangle t}(t)$ is non-decreasing in t. The function $f(t, y) = 3y^{2/3}$ is continuous and increasing in y for $y \geq 0$. It therefore follows by induction that for each fixed $\triangle t$ and n, $Y_{u,\triangle t}(n\triangle t)$ is a continuous and increasing function of $u \in [0, \infty)$. Since $Y_{u,\triangle t}(\cdot)$ is a polygonal function, for each fixed $\triangle t > 0$ and $t > 0$, the function $h(u) = Y_{u,\triangle t}(t)$ is also continuous and increasing. We also have $h(0) = 0$ and $u \leq h(u)$. Using the Intermediate Value Theorem we see that $h(u) = Y_{u,\triangle t}(t)$ maps $[0, \infty)$ onto $[0, \infty)$. Thus whenever

$$\triangle t > 0, \quad t \geq 0, \quad v \geq 0$$

there exists $u \in [0, v]$ such that

$$Y_{u,\triangle t}(t) = v.$$

Now let dt be positive infinitesimal. The constant zero solution is the standard part of the hyperfinite Euler approximation with increment dt and initial value 0, because $Y_{0,dt}(t) = 0$. Let $y_a(\cdot)$ be a nonzero solution of the initial value problem with initial value 0. Then a is a non-negative real number. Take a time $t > a$, so that $y_a(t) = (t-a)^3 > 0$. By the Partial Solution Theorem, there is a hyperreal $z \in [0, y_a(t)]^*$ such that

$$Y_{z,dt}(t) = y_a(t).$$

By Theorem 14.2, the standard part of $Y_{z,dt}(\cdot)$ exists on $[0, \infty)$ and is a solution of the initial value problem with initial value $\mathrm{st}(z)$. By Proposition 14.16, $y_a(\cdot)$ is the only solution of the initial value problem whose graph contains the point $(t, y_a(t))$, so $y_a(\cdot)$ must be the standard part of $Y_{z,dt}(\cdot)$. $\dashv$

CHAPTER 15

LOGIC AND SUPERSTRUCTURES

This chapter is optional, and provides a link between the simple treatment of infinitesimal calculus in the text *Elementary Calculus* and the more advanced treatment of infinitesimal analysis found in the literature. The material in this chapter is not needed as background for teaching calculus from *Elementary Calculus*. It is aimed at mathematicians who wish to go more deeply into the subject.

15A. The Elementary Extension Principle

We will show here that the Transfer Axiom is equivalent to an apparently stronger statement called the Elementary Extension Principle. It says that the real numbers (with all real functions and relations) satisfy the same sentences of first order logic as the hyperreal numbers (with the natural extensions of all real functions and relations). Before stating the result precisely we start from the beginning and define the notion of a sentence of first order logic.

In first order logic we start with a set of symbols, called a **language**, appropriate for the structure under consideration. In this case we introduce a language L for the real number system with the following uncountable collection of symbols:

A symbol $\overline{c}$ for each real constant c,
A symbol $\overline{f}$ for each real function f of n variables,
A symbol $\overline{P}$ for each real relation P of n variables.

In addition, L has the following logical symbols common to all first order languages:

Variables $v_1, v_2, v_3, \ldots$
Connectives $\neg$ (not), $\wedge$ (and), $\vee$ (or), $\Rightarrow$ (implies). $\Leftrightarrow$ (if and only if)
Quantifiers $\forall$ (for all), $\exists$ (there exists)
Parentheses and commas.

For simplicity we identify each real constant c, function f, and relation P, with its symbol $\overline{c}, \overline{f}, \overline{P}$. In Section 1C we defined the notion of a term.

We repeat the definition here. A **term** is a finite sequence of symbols built according to the following rules:

- Every variable is a term.
- Every constant is a term.
- If $\tau_1, \dots, \tau_n$ are terms and f is a real function of n variables, then $f(\tau_1, \dots, \tau_n)$ is a term.

When we replace each variable v_i in a term $\tau(v_1, \dots, v_n)$ by a constant c_i, we obtain a **constant term**, which is either equal to some real number or is undefined.

In Section 1C we called an equation or inequality between two terms a formula. These are expressions of the forms

$$\tau = \sigma, \quad \tau \neq \sigma, \quad \tau \leq \sigma, \quad \tau < \sigma, \quad \tau \geq \sigma, \quad \tau > \sigma.$$

The equations and inequalities are special cases of a broader class which we call the atomic formulas of L. If P is a real relation of n variables and $\tau_1, \dots, \tau_n$ are terms, then

$$P(\tau_1, \dots, \tau_n)$$

is called an **atomic formula** of L.

The set of all (first order) **formulas of** L is defined as follows.

- Every atomic formula of L is a formula of L.
- If φ, ψ are formulas of L, so are
$$\neg\varphi, \quad (\varphi \wedge \psi), \quad (\varphi \vee \psi), \quad (\varphi \Rightarrow \psi), \quad (\varphi \Leftrightarrow \psi).$$
- If φ is a formula of L and v_n is a variable, then
$$(\forall v_n \varphi), \quad (\exists v_n \varphi)$$
are formulas of L.

For example, omitting unnecessary parentheses, the following formula of L is the standard ε, δ condition for f to be continuous at x. For readability we use $\varepsilon, \delta, x, y$ for variables instead of v_1, v_2, v_3, v_4.

$$\forall\varepsilon(\varepsilon > 0 \Rightarrow \exists\delta(\delta > 0 \wedge \forall y(|x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon))).$$

In this example, ε, δ, and y are **bound variables**, while x is a **free variable** because it is not in a quantifier. A first order formula with no free variables is called a **sentence**. Thus whenever all the free variables in a first order formula are replaced by constants, the result is a sentence.

The statement that every real solution of a system of formulas S is a real solution of a system of formulas T is expressed by a sentence of L of the form

$$\forall v_1 \cdots \forall v_\ell((\varphi_1 \wedge \cdots \wedge \varphi_m) \Rightarrow (\psi_1 \wedge \cdots \wedge \psi_n)) \tag{83}$$

where the φ_i and ψ_j are atomic formulas.

The statement that every real solution of a system of formulas S is a partial real solution of a system of formulas T is expressed by a sentence of L of the form

$$\forall v_1 \cdots \forall v_k((\varphi_1 \wedge \cdots \wedge \varphi_m) \Rightarrow \exists v_{k+1} \cdots \exists v_\ell(\psi_1 \wedge \cdots \wedge \psi_n)). \tag{84}$$

Each sentence φ of L is either true or false in the real number system. The notion of a true sentence is a precise mathematical concept which is defined by induction on the complexity of a sentence, and corresponds to the intuitive notion.

Definition 15.1. *(i) An atomic sentence $P(\tau_1, \dots, \tau_n)$ is true if each of the constant terms $\tau_1, \dots, \tau_n$ is defined and the n-tuple of values belongs to the relation P.*

(ii) If φ and ψ are sentences, then the truth values of the combinations of φ and ψ by connectives are obtained from the truth values of φ and ψ using the following table.

φ	ψ	$\neg\varphi$	$\varphi \wedge \psi$	$\varphi \vee \psi$	$\varphi \Rightarrow \psi$	$\varphi \Leftrightarrow \psi$
T	T	F	T	T	T	T
T	F	F	F	T	F	F
F	T	T	F	T	T	F
F	F	T	F	F	T	T

(iii) The sentence $\forall x\varphi(x)$ is true if and only if $\varphi(c)$ is true for all constants $c \in \mathbb{R}$.

(iv) The sentence $\exists x\varphi(x)$ is true if and only if $\varphi(c)$ is true for some constant $c \in \mathbb{R}$.

The notation $\mathbb{R} \models \varphi$ means that φ is true in $\mathbb{R}$.

Note that we treat $=$ and $\neq$ as binary relations, and that the sentence $\tau \neq \sigma$ is different than $\neg\tau = \sigma$. In fact, $\mathbb{R} \models \tau \neq \sigma$ if and only if τ and σ are both defined but have different values, while $\mathbb{R} \models \neg\tau = \sigma$ if and only if it is not the case that τ and σ are both defined and equal.

We now introduce a second language L^* for the hyperreal number system which has a symbol for each hyperreal constant, function, and relation. The truth value of a sentence of L^* in $\mathbb{R}^*$ is defined as before but with quantifiers ranging over $\mathbb{R}^*$ instead of $\mathbb{R}$.

Definition 15.2. *Given a formula φ of L, the* **$*$-transform** *of φ is the formula φ^* of L^* obtained by replacing each real function f and relation P occurring in φ by its natural extension f^* and P^*.*

Given a term τ of L, the $*$-transform τ^* is defined analogously.

Thus the $*$-transform of the formula expressing continuity of f at x is the formula

$$\forall\varepsilon(\varepsilon >^* 0 \Rightarrow \exists\delta(\delta >^* 0 \wedge \forall y(|x -^* y|^* <^* \delta \Rightarrow |f^*(x) -^* f^*(y)|^* <^* \varepsilon))).$$

The relation $x \approx y$ is expressed by the following formula of L^*:

$$\forall z(\mathbb{R}(z) \wedge 0 <^* z \Rightarrow |x -^* y|^* <^* z).$$

This formula is not the $*$-transform of any formula of L because it involves the set $\mathbb{R}$, which is a subset of $\mathbb{R}^*$ but is not the natural extension of any set of reals.

Similarly, the property "x is finite" is expressed by

$$\exists z(\mathbb{R}(z) \wedge |x|^* <^* z),$$

which again is not the $*$-transform of any formula of L.

The $*$-transform of a *sentence* of L is always a sentence of L^*. We can now state the Elementary Extension Principle and show that it follows from our Axioms A–E for hyperreal numbers.

THEOREM 15.3. *(Elementary Extension Principle) For every sentence φ of L, φ is true in $\mathbb{R}$ if and only if its $*$-transform φ^* is true in $\mathbb{R}^*$.*

Notice that the Transfer Axiom is just the special case of the Elementary Extension Principle where φ is a sentence of the form (83) above. The Partial Solution Theorem 1.20 is the special case where φ is a sentence of the form (84) above.

PROOF. For each formula $\varphi(v_1, \ldots, v_n)$ of L with at most the free variables $v_1, \ldots, v_n$, let C_φ be the corresponding characteristic function

$$C_\varphi(a_1, \ldots a_n) = \begin{cases} 1 & \text{if } \varphi(a_1, \ldots, a_n) \text{ holds in } \mathbb{R} \\ 0 & \text{otherwise.} \end{cases}$$

Thus

$$\mathbb{R} \models \forall v_1 \cdots \forall v_n(\varphi(v_1, \ldots, v_n) \Leftrightarrow C_\varphi(v_1, \ldots, v_n) = 1).$$

We will prove by induction on the complexity of formulas that for each formula $\varphi(v_1, \ldots, v_n)$ of L,

$$\mathbf{R}^* \models \forall v_1 \cdots \forall v_n(\varphi^*(v_1, \ldots, v_n) \Leftrightarrow (C_\varphi)^*(v_1, \ldots, v_n) = 1). \tag{85}$$

Replacing variables by real constants, it will then follow that the statements below are equivalent:

$$\begin{aligned} \mathbb{R} &\models \varphi(a_1, \ldots, a_n) \\ \mathbb{R} &\models C_\varphi(a_1, \ldots, a_n) = 1 \\ \mathbb{R}^* &\models (C_\varphi)^*(a_1, \ldots, a_n) = 1 \\ \mathbb{R}^* &\models \varphi^*(a_1, \ldots, a_n). \end{aligned}$$

To start the induction, we let $\varphi(v)$ be an atomic formula $P(\tau_1(v), \ldots, \tau_k(v))$. (For simplicity we do this for the case that $\varphi(v)$ has only one variable v; the case of n variables is similar).

Let C_P be the characteristic function of $P(y_1, \ldots, y_k)$. By the definition of P^*,

$$\mathbb{R}^* \models \forall y_1 \cdots \forall y_k(P^*(y_1, \ldots, y_k) \Leftrightarrow (C_P)^*(y_1, \ldots, y_k) = 1),$$

and since $\varphi(v)$ is $P(\tau_1(v), \dots, \tau_k(v))$,

$$\mathbb{R}^* \models \forall v(\varphi^*(v) \Leftrightarrow (C_P)^*(\tau_1^*(v), \dots, \tau_k^*(v)) = 1).$$

Moreover, in $\mathbb{R}$ we have

$$\mathbb{R} \models \forall v(C_P(\tau_1(v), \dots, \tau_k(v)) = 1 \Leftrightarrow C_\varphi(v) = 1).$$

Then by Transfer,

$$\mathbb{R}^* \models \forall v((C_P)^*(\tau_1^*(v), \dots, \tau_k^*(v)) = 1 \Leftrightarrow (C_\varphi)^*(v) = 1).$$

Combining the above formulas, we see that (85) holds for $\varphi(v)$.

To shorten our induction we use the fact that every formula of L is equivalent to a formula which is built up using only the connectives $\wedge$ and $\neg$ and the quantifier $\exists$. The other connectives and quantifiers may be treated as abbreviations for longer expressions. In the following we assume that (85) holds for $\varphi(v_1, \dots, v_n)$ and $\psi(v_1, \dots, v_n)$, and prove that (85) also holds for $\neg\varphi$, $\varphi \wedge \psi$, and $\exists v_n \varphi$.

We first consider $\neg\varphi$. We have

$$\mathbb{R} \models \forall v_1 \cdots \forall v_n(\varphi(v_1, \dots, v_n) \Leftrightarrow C_\varphi(v_1, \dots, v_n) = 1),$$

$$\mathbb{R} \models \forall v_1 \cdots \forall v_n(\neg\varphi(v_1, \dots, v_n) \Leftrightarrow C_{\neg\varphi}(v_1, \dots, v_n) = 1),$$

so

$$\mathbb{R} \models \forall v_1 \cdots \forall v_n(\neg C_\varphi(v_1, \dots, v_n) = 1 \Leftrightarrow C_{\neg\varphi}(v_1, \dots, v_n) = 1).$$

By Transfer the $*$-transforms hold in $\mathbb{R}^*$. Therefore

$$\mathbb{R}^* \models \forall v_1 \cdots \forall v_n(\neg(C_\varphi)^*(v_1, \dots, v_n) = 1 \Leftrightarrow (C_{\neg\varphi})^*(v_1, \dots, v_n) = 1).$$

Using the hypothesis (85) for φ, we have

$$\mathbb{R}^* \models \forall v_1 \cdots \forall v_n(\varphi^*(v_1, \dots, v_n) \Leftrightarrow (C_\varphi)^*(v_1, \dots, v_n) = 1).$$

It follows that

$$\mathbb{R}^* \models \forall v_1 \cdots \forall v_n(\neg\varphi^*(v_1, \dots, v_n) \Leftrightarrow (C_{\neg\varphi})^*(v_1, \dots, v_n) = 1),$$

so (85) holds for $\neg\varphi$.

To verify (85) for $\varphi \wedge \psi$ we observe that

$$\mathbb{R} \models \forall v_1 \cdots \forall v_n(C_{\varphi\wedge\psi}(v_1, \dots, v_n) = C_\varphi(v_1, \dots, v_n) \cdot C_\psi(v_1, \dots, v_n)),$$

and by Transfer the $*$-transform holds in $\mathbb{R}^*$.

We now show that $\exists v_n \varphi$ satisfies (85). For any real $(n-1)$-tuple $a_1, \dots, a_{n-1}$ such that

$$\mathbb{R} \models \exists v_n \varphi(a_1, \dots, a_{n-1}, v_n),$$

choose a real number $f(a_1, \dots, a_{n-1})$ such that

$$\mathbb{R} \models \varphi(a_1, \dots, a_{n-1}, f(a_1, \dots, a_{n-1})).$$

Otherwise put $f(a_1, \dots, a_{n-1}) = 0$. The real function f of $n-1$ variables is called a **Skolem function** for $\exists v_n \varphi$. Working in $\mathbb{R}$ we see that

$$\mathbb{R} \models \forall v_1 \cdots \forall v_{n-1}\, C_{\exists v_n \varphi}(v_1, \dots, v_{n-1}) = C_\varphi(v_1, \dots, v_{n-1}, f(v_1, \dots, v_{n-1}))$$

and also

$$\mathbb{R} \models \forall v_1 \cdots \forall v_n (C_\varphi(v_1, \dots, v_n) = 1 \Rightarrow C_{\exists v_n \varphi}(v_1, \dots, v_{n-1}) = 1).$$

It follows from Transfer that the $*$-transforms of these formulas hold in $\mathbb{R}^*$. Thus for any $b_1, \dots, b_{n-1} \in \mathbb{R}^*$, each statement below implies the next.

$$\begin{aligned}
&\mathbb{R}^* \models \exists v_n \varphi^*(b_1, \dots, b_{n-1}, v_n)\\
&\mathbb{R}^* \models \exists v_n ((C_\varphi)^*(b_1, \dots, b_{n-1}, v_n) = 1)\\
&\mathbb{R}^* \models (C_{\exists v_n \varphi})^*(b_1, \dots, b_{n-1}) = 1\\
&\mathbb{R}^* \models (C_\varphi)^*(b_1, \dots, b_{n-1}, f^*(b_1, \dots, b_{n-1})) = 1\\
&\mathbb{R}^* \models \varphi^*(b_1, \dots, b_{n-1}, f^*(b_1, \dots, b_{n-1}))\\
&\mathbb{R}^* \models \exists v_n \varphi^*(b_1, \dots, b_{n-1}, v_n).
\end{aligned}$$

We conclude that (85) holds for $\exists v_n \varphi$. ⊣

The sentences which arise in beginning calculus are of a very simple form, and the Transfer Axiom and Partial Solution Theorem are broad enough to cover the cases of the Elementary Extension Principle which are needed in proofs. It is important pedagogically that we are able to base the course on the familiar concept of a system of equations and inequalities rather than on the general notion of a formula in first order logic. Beginning calculus students do not have the mathematical experience necessary to work with formulas with even three quantifiers, such as the ε, δ definition of continuity. One advantage of introducing the hyperreal numbers at the beginning of the calculus course is that complicated first order formulas of L can often be replaced by simpler formulas of L^*. This is especially true when the hyperreal relations $x \approx y$ and $x = \mathrm{st}(y)$ are used. For example, the ε, δ condition for f to be continuous at x,

$$\forall \varepsilon (\varepsilon > 0 \Rightarrow \exists \delta (\delta > 0 \wedge \forall y (|x - y| < \delta \Rightarrow |f(x) - f(y)| < \varepsilon))),$$

is equivalent to the simpler L^* formula

$$\forall y (y \approx x \Rightarrow f^*(y) \approx f^*(x)).$$

15B. Superstructures

In classical analysis one goes beyond the real numbers and real functions. A more appropriate object of study is the superstructure over the real numbers, defined as follows.

Definition 15.4. *The* **power set** *of a set X is the set $\mathcal{P}(X)$ of all subsets of X,*

$$\mathcal{P}(X) = \{Y : Y \subseteq X\}.$$

The n-th **cumulative power set** *of X is defined recursively by*

$$V_0(X) = X, \quad V_{n+1}(X) = V_n(X) \cup \mathcal{P}(V_n(X)).$$

The **superstructure** *over X is the union of the cumulative power sets and is denoted by $V(X)$,*

$$V(X) = \bigcup_{n=0}^{\infty} V_n(X).$$

The superstructure $V(X)$ has a **membership relation** *between elements of $V_n(X)$ and $V_{n+1}(X), n = 0, 1, 2, \ldots$. We treat elements of X itself as atoms, and assume that $\emptyset \notin X$ and that no $x \in X$ contains any elements of $V(X)$.*

We observe that for any superstructure $V(X)$, we have

$$X \subseteq V_1(X) \subseteq V_2(X) \subseteq \cdots$$

and

$$X \in V_1(X) \in V_2(X) \in \cdots .$$

Moreover, $\mathcal{P}(X) \in V(X)$, $V_n(X) \in V(X)$ for each n, and if $A \in V(X) \setminus X$ then $A \subseteq V(X)$. One usually takes the set X of atoms to be the set $\mathbb{R}$ of reals, the set $\mathbb{C}$ of complex numbers, or some other structure under investigation. Here we concentrate on the case where the set of atoms is $X = \mathbb{R}$.

The sets $\mathbb{Z}$ of integers and $\mathbb{N}$ of natural numbers are important subsets of $\mathbb{R}$ and elements of $V(\mathbb{R})$. We may define an ordered pair $\langle x, y\rangle$ by

$$\langle x, y\rangle = \{\{x\}, \{x, y\}\}$$

and an ordered n-tuple, $n > 2$, as the function

$$\langle x_1, \ldots, x_n\rangle = \{\langle 1, x_1\rangle, \ldots, \langle n, x_n\rangle\}$$

from $\{1, \ldots, n\}$ into $\mathbb{R}$. Thus all ordered n-tuples belong to $V_3(\mathbb{R})$, and all real relations and functions in n variables belong to $V_4(\mathbb{R})$. Function spaces, measures, and all other structures from classical analysis belong to $V(\mathbb{R})$, and even to, say, $V_{100}(\mathbb{R})$.

The following lemma is easily proved by induction on n.

Lemma 15.5. *If $n > 0$ and $x \in y \in V_n(\mathbb{R})$, then $x \in V_{n-1}(\mathbb{R})$.*

In order to be in a position to use infinitesimals in more advanced areas of mathematics, we must extend the whole superstructure instead of just the real numbers.

Definition 15.6. *A* **superstructure embedding** *is a one to one mapping $*$ of $V(\mathbb{R})$ into another superstructure $V(\mathbb{S})$ such that*

(i) $\mathbb{R}$ is a proper subset of $\mathbb{S}$, $r^* = r$ for all $r \in \mathbb{R}$, and $\mathbb{R}^* = \mathbb{S}$.

(ii) For $x, y \in V(\mathbb{R})$, $x \in y$ if and only if $x^* \in y^*$.

In view of (i), we will always write $\mathbb{R}^*$ instead of $\mathbb{S}$, and denote the superstructure embedding by $*\colon V(\mathbb{R}) \to V(\mathbb{R}^*)$. Notice that by (i), $V(\mathbb{R}) \subseteq V(\mathbb{R}^*)$ and $*$ contains the identity map on $\mathbb{R}$, but $*$ does not contain the identity map on $V(\mathbb{R})$.

To go further we need an analogue of the Elementary Extension Principle for superstructure embeddings. An arbitrary embedding will not do; we want the embedding to preserve all properties of a certain kind. Assume hereafter that $*\colon V(\mathbb{R}) \to V(\mathbb{R}^*)$ is a superstructure embedding.

We first form the first order predicate logic with the equality symbol, a binary relation symbol $\in$, and a constant symbol $\overline{a}$ for each element $a \in V(\mathbb{R}^*)$. For simplicity we identify a with its constant symbol $\overline{a}$. We call the language $\mathcal{L}$.

DEFINITION 15.7. *A* **bounded formula** *of $\mathcal{L}$ is an expression built according to the following rules.*

- *If x and y are variables or constants, $x = y$ and $x \in y$ are bounded formulas, called* **atomic formulas***.*
- *If φ, ψ are bounded formulas, so are*
$$(\neg\varphi), \quad (\varphi \wedge \psi), \quad (\varphi \vee \psi), \quad (\varphi \Rightarrow \psi), \quad (\varphi \Leftrightarrow \psi).$$
- *If u is a variable, c is a constant, and φ is a bounded formula, then $(\forall u \in c)\varphi$ and $(\exists u \in c)\varphi$ are bounded formulas.*
- *If u, v are variables, φ is a bounded formula, and v does not appear in φ in the form $(\forall v \in w)$ or $(\exists v \in w)$, then $(\forall u \in v)\varphi$ and $(\exists u \in v)\varphi$ are bounded formulas.*

The adjective *bounded* refers to the fact that the quantifiers in bounded formulas of $\mathcal{L}$ are of the form $(\forall u \in x)$ or $(\exists u \in x)$.

A **bounded sentence** is a bounded formula in which each occurrence of a variable u is within the scope of a quantifier of the form $(\forall u \in x)$ or $(\exists u \in x)$ where x is another variable or constant. Note that if a bounded sentence begins with a quantifier $(\forall u \in x)$ or $(\exists u \in x)$, x must be a constant.

For example, the property $y = \bigcup x$ is expressed by the bounded formula
$$(\forall u \in y)(\exists v \in x)u \in v \wedge (\forall v \in x)(\forall u \in v)u \in y.$$
If x and y are constants, this is a bounded sentence.

Each bounded sentence is either true or false in the superstructure $V(\mathbb{R}^*)$. The definition of the relation "φ is true in $V(\mathbb{R}^*)$" is by induction on the complexity of the bounded sentence φ. The quantifier clauses are:

$(\exists v \in c)\varphi(v)$ is true in $V(\mathbb{R}^*)$ iff $\varphi(b)$ is true in $V(\mathbb{R}^*)$ for some $b \in c$,
$(\forall v \in c)\varphi(v)$ is true in $V(\mathbb{R}^*)$ iff $\varphi(b)$ is true in $V(\mathbb{R}^*)$ for all $b \in c$.

The notation $V(\mathbb{R}^*) \models \varphi$ means "φ is true in $V(\mathbb{R}^*)$". Since $V(\mathbb{R}^*)$ is the only structure under discussion, we sometimes suppress mention of $V(\mathbb{R}^*)$ and say "φ is true" instead of "φ is true in $V(\mathbb{R}^*)$".

Remember that $\mathbb{R} \subseteq \mathbb{R}^*$ and $V(\mathbb{R}) \subseteq V(\mathbb{R}^*)$. We call the elements of $\mathbb{R}$ **real numbers**, the elements of $V(\mathbb{R}) \setminus \mathbb{R}$ **real sets**, and arbitrary elements of $V(\mathbb{R})$ **real entities**. Similarly, elements of $\mathbb{R}^*$, $V(\mathbb{R}^*) \setminus \mathbb{R}^*$, and $V(\mathbb{R}^*)$ are called **hyperreal numbers**, **hyperreal sets**, and **hyperreal entities** respectively. A **real bounded formula** is a bounded formula of $\mathcal{L}$ all of whose constants are real entities. Since each element of a real set is a real entity, a real bounded sentence has the same meaning in $V(\mathbb{R})$ as in $V(\mathbb{R}^*)$.

The superstructure embedding $*\colon V(\mathbb{R}) \to V(\mathbb{R}^*)$ induces a mapping, called the **$*$-transform**, from real bounded formulas to bounded formulas. The $*$-transform φ^* of a real bounded formula φ is defined as the bounded formula obtained by replacing each constant c occurring in φ by its image c^*.

For example, the $*$-transform of the real bounded sentence

$$(\forall x \in \mathbb{R})(x < 0 \vee (\exists y \in \mathbb{R}) y \cdot y = x)$$

is the bounded sentence

$$(\forall x \in \mathbb{R}^*)(x <^* 0^* \vee (\exists y \in \mathbb{R}^*) y \cdot^* y = x).$$

Definition 15.8. *A* **nonstandard universe** *is a superstructure embedding* $*\colon V(\mathbb{R}) \to V(\mathbb{R}^*)$ *which satisfies* **Leibniz' Principle**, *which is the property that for each real bounded sentence $\varphi \in \mathcal{L}$, φ is true if and only if φ^* is true.*

Of course, Leibniz did not formulate his principle in anything like the present form. In fact, first order predicate logic was not available until the work of Frege and Peano in the late nineteenth century. The name "Leibniz' Principle" is used in the literature because Leibniz suggested that the real numbers should be extended to a larger system which has the same elementary properties but contains infinitesimals. The formal notion given here captures this intuitive idea.

In Section 15D we will build a nonstandard universe. We now show that Leibniz' Principle implies the Elementary Extension Principle of Section 15A.

By the **elementary part** of the superstructure $V(\mathbb{R})$ we will mean the set

$$\mathbb{R} \cup \bigcup_{n=1}^{\infty} \mathcal{P}(\mathbb{R}^n)$$

of all elements of $\mathbb{R}$ and all relations (and hence functions) on $\mathbb{R}$ with finitely many variables. By the **elementary part** of the nonstandard universe $*\colon V(\mathbb{R} \to \mathbb{R}^*$ we mean the structure $(\bullet, \mathbb{R}, \mathbb{R}^*)$ where $\bullet$ is the restriction of the mapping $*$ to the elementary part of $V(\mathbb{R})$. This mapping $\bullet$ associates with each real relation P or function f a hyperreal relation P^* or function f^*.

Theorem 15.9. *The elementary part of a nonstandard universe $*\colon V(\mathbb{R}) \to V(\mathbb{R}^*)$ satisfies the Elementary Extension Principle.*

Proof. Given a first order sentence φ of L, let $\varphi^{\mathbb{R}}$ be the real bounded sentence of $\mathcal{L}$ obtained by replacing each quantifier $\exists v_n$ occurring in φ by the bounded quantifier $(\exists v_n \in \mathbb{R})$, and replacing each quantifier $\forall v_n$ occurring in φ by the bounded quantifier $(\forall v_n \in \mathbb{R})$. Define $\varphi^{\mathbb{R}^*}$ similarly. Then for each sentence φ of L, we have

$$(\varphi^*)^{\mathbb{R}^*} = \left(\varphi^{\mathbb{R}}\right)^*.$$

From the definition of truth value we see that the following are equivalent:

$$\mathbb{R} \models \varphi, \quad \varphi^{\mathbb{R}} \text{ is true}, \quad \left(\varphi^{\mathbb{R}}\right)^* \text{ is true}, \quad \mathbb{R}^* \models \varphi^*.$$

This completes the proof. $\dashv$

THEOREM 15.10. *The elementary part of a nonstandard universe* $*: V(\mathbb{R}) \to V(\mathbb{R}^*)$ *is a hyperreal number system which satisfies Axioms A–E for the hyperreal numbers.*

PROOF. Axiom A, that $\mathbb{R}$ is a complete ordered field, is satisfied by definition. Axiom B says that $\mathbb{R}^*$ is an ordered field extension of $\mathbb{R}$. By the definition of a superstructure embedding, $r^* = r$ for each $r \in \mathbb{R}$, so $\mathbb{R}^*$ is an extension of $\mathbb{R}$. Each of the ordered field axioms is a first order sentence which holds in $\mathbb{R}$, so its $*$-transform holds in $\mathbb{R}^*$ by the Elementary Extension Principle. Therefore Axiom B is satisfied.

Axiom D, the Function Axiom, holds because the superstructure embedding $*$ gives us the natural extension f^* of each real function f and the natural extension $<^*$ of the real order relation $<$. Axiom E, the Transfer Axiom, is a special case of the Elementary Extension Principle, which holds by Theorem 15.9.

It remains to prove Axiom C, that $\mathbb{R}^*$ has a positive infinitesimal. By the definition of a superstructure embedding, $\mathbb{R}^*$ is a proper field extension of $\mathbb{R}$, so there is an element $x \in \mathbb{R}^* \setminus \mathbb{R}$. Suppose first that x is infinite. Using only Axioms A and B, it now follows that $|x|$ is positive infinite, $|x|^{-1}$ is positive, and $|x|^{-1}$ is infinitesimal (here we use Theorem 1.7, whose proof uses only Axioms A and B). So in this case $\mathbb{R}^*$ has a positive infinitesimal. Now suppose that x is finite. The Standard Part Principle, Theorem 1.9, is also proved using only Axioms A and B, and shows that there is a real number r such that $r - x$ is infinitesimal. Since $x \notin \mathbb{R}$, $r - x \neq 0$, and therefore $\mathbb{R}^*$ has the positive infinitesimal element $|r - x|$. This proves Axiom C. $\dashv$

Theorem 15.10 has the following converse, which we will state without proof. It shows that any hyperreal system can be extended to a nonstandard universe. A proof can be found in the book [CK 1990], Section 4.4.

THEOREM 15.11. *Suppose that the triple* $(\bullet, \mathbb{R}, \mathbb{R}^*)$ *satisfies Axioms A–E for the hyperreal numbers. Then there is a nonstandard universe* $*: V(\mathbb{R}) \to V(\mathbb{R}^*)$ *with elementary part* $(\bullet, \mathbb{R}, \mathbb{R}^*)$.

15C. Standard, Internal, and External Sets

In this section we assume that $*: V(\mathbb{R}) \to V(\mathbb{R}^*)$ is a nonstandard universe. By Theorem 15.10, it follows that Axioms A–E and all their consequences hold for the real and hyperreal numbers. We will adopt the convention of dropping asterisks on terms $f^*(x_1, \ldots, x_n)$ where f is a real function, and on the hyperreal order relations $<^*$, $\leq^*$, $>^*$, $\geq^*$.

A number of new distinctions which did not arise in the hyperreal numbers become important in a nonstandard universe. The image of the power set of $\mathbb{R}$ will contain some but not all subsets of $\mathbb{R}^*$, that is, $[\mathcal{P}(\mathbb{R})]^*$ will be a proper subset of $\mathcal{P}(\mathbb{R}^*)$. For example, the set $\mathbb{R}$ of real numbers and the set $\mathbb{N}$ of natural numbers are subsets of $\mathbb{R}^*$ but do not belong to $[\mathcal{P}(\mathbb{R})]^*$. To see this we note that the $*$-transform of the sentence

"Every nonempty bounded set in $\mathcal{P}(\mathbb{R})$ has a least upper bound"

holds in $V(\mathbb{R}^*)$. However, $\mathbb{R}$ and $\mathbb{N}$ are bounded but have no least upper bound in $\mathbb{R}^*$, so $\mathbb{R}$ and $\mathbb{N}$ cannot belong to $[\mathcal{P}(\mathbb{R})]^*$.

It is useful to distinguish between four kinds of entities in $V(\mathbb{R}^*)$. An entity $b \in V(\mathbb{R}^*)$ is said to be

real	if $b \in V(\mathbb{R})$,
standard	if $b = a^*$ for some $a \in V(\mathbb{R})$,
internal	if $b \in a^*$ for some $a \in V(\mathbb{R}) \setminus \mathbb{R}$,
external	if b is not internal.

Proposition 15.12. *(i) Every standard entity is internal.*
(ii) An entity x is internal if and only if $x \in [V_n(\mathbb{R})]^*$ for some n.
(iii) Every element of an internal set is internal.

Proof. (i) If x is standard, $x = a^*$ for some real entity a. If a is a real number, then $x \in \mathbb{R}^*$ and $\mathbb{R} \in V(\mathbb{R})$. If a is a real set, then $a \in \mathcal{P}(a) \in V(\mathbb{R})$, so $x = a^* \in [\mathcal{P}(a)]^*$. In each case, x is internal.

(ii) Since $V_n(\mathbb{R}) \in V(\mathbb{R})$, every element of $[V_n(\mathbb{R})]^*$ is internal. Suppose x is internal, so that $x \in a^*$ for some $a \in V(\mathbb{R}) \setminus \mathbb{R}$. Then $a \subseteq V_n(\mathbb{R})$ for some n, so

$$(\forall u \in a) u \in V_n(\mathbb{R}).$$

By Leibniz' Principle,

$$(\forall u \in a^*) u \in [V_n(\mathbb{R})]^*,$$

and hence $x \in [V_n(\mathbb{R})]^*$.

(iii) Suppose x is an internal set and $y \in x$. By (ii), $x \in [V_n(\mathbb{R})]^*$ for some n. Since x is a set, $n > 0$. By Lemma 15.5,

$$(\forall u \in V_n(\mathbb{R}))(\forall v \in u) v \in V_{n-1}(\mathbb{R}).$$

By Leibniz' Principle,

$$(\forall u \in [V_n(\mathbb{R})]^*)(\forall v \in u) v \in [V_{n-1}(\mathbb{R})]^*.$$

Hence $y \in [V_{n-1}(\mathbb{R})]^*$, so y is internal. $\dashv$

Here are some examples.
Standard and real: Each $r \in \mathbb{R}$. Each finite subset of $\mathbb{R}$.
External and real: $\mathbb{N}, \mathbb{R}$.
Standard but not real: $\mathbb{N}^*, \mathbb{R}^*$.
Internal but not standard:
Each $c \in \mathbb{R}^* \setminus \mathbb{R}$.
$[a, b]^*$ where $a, b \in \mathbb{R}^* \setminus \mathbb{R}$.
The function $h(x) = \sin(Hx)$ where H is infinite.

External but not real:
The monad of 0.
The galaxy of 0.
The standard part function st$(\cdot)$.

We remark that for every standard function $g(x, y)$ and hyperreal constant c, the function $h(x) = g(x, c)$ is internal. Similarly, for any standard relation $P(x, y)$ are hyperreal constant c, the relation $P(x, c)$ is internal.

A bounded formula of $\mathcal{L}$ is said to be internal if every constant occurring in the formula is internal. For example, the $*$-transform of a real bounded formula is an internal formula. Any formula formed from an internal formula by replacing variables by internal constants is again an internal formula.

THEOREM 15.13. *Let A be a real set, $A \in V(\mathbb{R}) \setminus \mathbb{R}$.*
(i) If $B \subseteq A$ then $B^* \subseteq A^*$.
(ii) $[\mathcal{P}(A)]^*$ is equal to the set of all internal subsets of A^*.

PROOF. (i) Since $B \subseteq A$, we have

$$(\forall x \in B) x \in A.$$

By Leibniz' Principle,

$$(\forall x \in B^*) x \in A^*,$$

so $B^* \subseteq A^*$.

(ii) First suppose $X \in [\mathcal{P}(A)]^*$. Then X is internal, because $\mathcal{P}(A) \in V(\mathbb{R})$. We have

$$(\forall u \in \mathcal{P}(A))(\forall v \in u) v \in A.$$

By Leibniz' Principle,

$$(\forall u \in [\mathcal{P}(A)]^*)(\forall v \in u) v \in A^*.$$

Therefore $(\forall v \in X) v \in A^*$, so X is an internal subset of A^*.

Now suppose that X is an internal subset of A^*. Then $X \in B^*$ for some real set B. We have

$$(\forall u \in B)((\forall v \in u) v \in A \Rightarrow u \in \mathcal{P}(A)).$$

By Leibniz' Principle,

$$(\forall u \in B^*)((\forall v \in u) v \in A^* \Rightarrow u \in [\mathcal{P}(A)]^*).$$

But $X \in B^*$ and $(\forall v \in X) v \in A^*$. Therefore $X \in [\mathcal{P}(A)]^*$. $\dashv$

Several basic notions such as open set, continuity, and differentiability, split into two separate notions when applied to internal functions and relations. Consider an internal function f on $\mathbb{R}^*$ and a point $c \in \mathbb{R}^*$. Let $\mathbb{R}_+$ be the set of positive reals. f is said to be **S-continuous** at c if it satisfies the real ε, δ condition

$$(\forall \varepsilon \in \mathbb{R}_+)(\exists \delta \in \mathbb{R}_+)(\forall x \in \mathbb{R}^*)(|x - c| < \delta \Rightarrow |f(x) - f(c)| < \varepsilon).$$

This is not an internal formula, because it has the constant $\mathbb{R}_+$. f is said to be $*$**-continuous** at c if it satisfies the hyperreal ε, δ condition

$$(\forall \varepsilon \in \mathbb{R}_+^*)(\exists \delta \in \mathbb{R}_+^*)(\forall x \in \mathbb{R}^*)(|x - c| < \delta \Rightarrow |f(x) - f(c)| < \varepsilon),$$

which is an internal formula.

Theorem 5.5 shows that if f is a standard function and c is real, then f is S-continuous at c if and only if f is $*$-continuous at c. This can also be proved by Leibniz' Principle. First we use Leibniz' Principle to show that for each real $\varepsilon, \delta \in \mathbb{R}_+$, we may replace $(\forall x \in \mathbb{R}^*)$ by $(\forall x \in \mathbb{R})$ in the definition of S-continuous, and then we use Leibniz' Principle again to show that f is S-continuous if and only if it is $*$-continuous. The two notions are not equivalent when f is only assumed to be internal. Here are two examples. Let H be positive infinite. The internal function $f(x) = \sin(Hx)$ is everywhere $*$-continuous but nowhere S-continuous. The internal function

$$g(x) = \begin{cases} 1/H & \text{if } x \in \mathbb{Q}^* \\ 0 & \text{if } x \notin \mathbb{Q}^* \end{cases}$$

where $\mathbb{Q}$ is the set of rational numbers, is everywhere S-continuous but nowhere $*$-continuous.

Some properties of the real numbers are never preserved by a superstructure embedding. For example, the field $\mathbb{R}$ of real numbers is Archimedean but the larger field $\mathbb{R}^*$ of hyperreal numbers is not. The Archimedean Property cannot be expressed by an internal formula. A field is Archimedean if every element is less than some natural number. Formally we have

$$(\forall x \in \mathbb{R})(\exists n \in \mathbb{N})x < n \text{ but } \neg(\forall x \in \mathbb{R}^*)(\exists n \in \mathbb{N})x < n.$$

The $*$-transform of the Archimedean Property has a different meaning and is true for the hyperreal numbers,

$$(\forall x \in \mathbb{R}^*)(\exists n \in \mathbb{N}^*)x < n.$$

Thus $\mathbb{R}^*$ is $*$-Archimedean but not Archimedean.

A second example is the completeness property. $\mathbb{R}$ is complete but $\mathbb{R}^*$ is not. Formally,

$$(\forall x \in \mathcal{P}(\mathbb{R}))\, \varphi(x) \text{ but } \neg(\forall x \in \mathcal{P}(\mathbb{R}^*))\, \varphi^*(x),$$

where $\varphi(x)$ is the bounded real formula stating that if x is nonempty and has an upper bound in $\mathbb{R}$ then x has a least upper bound in $\mathbb{R}$. However, $\mathbb{R}^*$ is $*$-complete, that is,

$$(\forall x \in [\mathcal{P}(\mathbb{R})]^*)\, \varphi^*(x).$$

$*$-completeness states that every internal subset of $\mathbb{R}^*$ which is nonempty and has an upper bound in $\mathbb{R}^*$ has a least upper bound in $\mathbb{R}^*$. A third example is induction on the natural numbers. We have

$$(\forall x \in \mathcal{P}(\mathbb{N}))(0 \in x \wedge (\forall y \in x)(y + 1 \in x) \Rightarrow x = \mathbb{N}),$$

but by considering $x = \mathbb{N}$ as an element of $\mathcal{P}(\mathbb{N}^*)$ we see that

$$\neg(\forall x \in \mathcal{P}(\mathbb{N}^*))(0 \in x \wedge (\forall y \in x)(y+1 \in x) \Rightarrow x = \mathbb{N}^*).$$

However, $\mathbb{N}^*$ does satisfy $*$-induction, that is, induction for internal subsets of $\mathbb{N}^*$,

$$(\forall x \in [\mathcal{P}(\mathbb{N})]^*)(0 \in x \wedge (\forall y \in x)(y+1 \in x) \Rightarrow x = \mathbb{N}^*).$$

We conclude this section with some consequences of Leibniz' Principle. The next theorem is very useful in practice when one wants to show that a set is internal.

THEOREM 15.14. *(Internal Definition Principle) If $\varphi(x, y_1, \dots, y_n)$ is a bounded formula with no constants, and $c, b_1, \dots, b_n$ are internal sets, then*

$$\{x \in c \colon \varphi(x, b_1, \dots, b_n)\}$$

is an internal set.

PROOF. By Proposition 15.12, there is a natural number m such that the elements $c, b_1, \dots, b_n$ all belong to $[V_m(\mathbb{R})]^*$. By the axioms of set theory, the following real bounded sentence is true:

$$(\forall y_1, \dots, y_n, z \in V_m(\mathbb{R}))(\exists u \in V_{m+1}(\mathbb{R}))u = \{x \in z \colon \varphi(x, y_1, \dots, y_n)\}.$$

By Leibniz' Principle, the $*$-transform of this sentence is true. It follows that

$$\{x \in c \colon \varphi(x, b_1, \dots, b_n)\} \in [V_{m+1}(\mathbb{R})]^*.$$

Therefore the set on the left is internal. ⊣

THEOREM 15.15. *(Overspill Principle) Let X be an internal subset of $\mathbb{R}^*$.*

(i) If $c \in \mathbb{R}^*$ and X contains the monad of c, then there is a real $\delta > 0$ such that

$$(c - \delta, c + \delta)^* \subseteq X.$$

(ii) If X contains infinitely many natural numbers then X contains a positive infinite hyperinteger.

PROOF. (i) Let

$$Y = \{y \in \mathbb{R}^* \colon 0 < y \leq 1 \wedge (\forall u \in (c - y, c + y)^*)u \in X\}.$$

By the Internal Definition Principle 15.14, the set Y is internal. Y is a subset of $\mathbb{R}^*$ and is bounded above by 1. Since X contains the monad of c, every positive infinitesimal belongs to Y. By the $*$-transform of the completeness property of the reals, Y has a least upper bound $b \in \mathbb{R}^*$. Then b is positive but not infinitesimal, so there is a real number δ such that $0 < \delta < b$. Then $\delta \in Y$ and hence

$$(c - \delta, c + \delta)^* \subseteq X.$$

(ii) Let K be a positive infinite hyperinteger. By the Internal Definition Principle, the set

$$Y = \{x \in X \colon x \in \mathbb{N}^* \wedge x \leq K\}$$

is internal, and has the upper bound K. Y is nonempty because it contains the infinite set $X \cap \mathbb{N}$. By the $*$-transform of the completeness property of the reals, Y has a least upper bound b. Since Y contains an infinite subset of $\mathbb{N}$, b must be positive infinite. Take an infinite $c < b$. Then c is not an upper bound of Y, so there exists $H \in Y$ such that $H > c$. Then H is infinite, $H \in \mathbb{N}^*$, and $H \in X$. ⊣

For example, suppose f is an internal function from $\mathbb{R}^*$ into $\mathbb{R}^*$, which is $*$-continuous at every point $x \approx c$. By the Internal Definition Principle, the set

$$X = \{x \in \mathbb{R}^* \colon f \text{ is } *\text{-continuous at } x\}$$

is internal and contains the monad of c. Then by the Overspill Principle, there is a real $\delta > 0$ such that $(c-\delta, c+\delta)^* \subseteq X$ and thus f is $*$-continuous at every point of $(c-\delta, c+\delta)^*$.

Corollary 15.16. *(Robinson's Principle) Let f be an internal function from $\mathbb{N}^*$ into $\mathbb{R}^*$. If $f(n)$ is infinitesimal for all finite n, then there is an infinite $H \in \mathbb{N}^*$ such that $f(K)$ is infinitesimal for all $K \leq H$ in $\mathbb{N}^*$.*

Proof. Let

$$Y = \{n \in \mathbb{N}^* \colon (\forall m \in \mathbb{N}^*)(m \leq n \Rightarrow |f(m)| \leq 1/(n+1)\}.$$

By the Internal Definition Principle, Y is internal. By hypothesis, $\mathbb{N} \subseteq Y \subseteq \mathbb{N}^*$. By Theorem 15.15, Y contains a positive infinite hyperinteger H. Then for all $K \leq H$ in $\mathbb{N}^*$, $|f(K)| \leq 1/(H+1)$ and hence $f(K)$ is infinitesimal. ⊣

15D. Bounded Ultrapowers

In this section we build a nonstandard universe. One way to do this is to apply the logical compactness theorem for the theory of types ([Henkin 1950]), as in the book [Robinson 1970]. Another method, which we will adopt here, is the bounded ultrapower.

Our method here will have two stages. First, we build a bounded ultrapower of the superstructure $V(\mathbb{R})$. This is a generalization of the ultrapower of the real number system $\mathbb{R}$ given in Section 1G. Then we use a method known as the Mostowski collapse to map this ultrapower into a new superstructure $V(\mathbb{R}^*)$.

The notion of a free ultrafilter over a set I was given in Definition 1.41. Theorem 1.42 shows that there exists a free ultrafilter over every infinite set. Hereafter we let U be a free ultrafilter over $\mathbb{N}$.

Definition 15.17. *A countable sequence $a = \langle a_0, a_1, a_2, \dots \rangle$ of elements of $V(\mathbb{R})$ is said to be* **bounded in** $V(\mathbb{R})$ *if there is a fixed $n \in \mathbb{N}$ such that each a_i belongs to $V_n(\mathbb{R})$.*

We will build the bounded ultrapower of $V(\mathbb{R})$ modulo U. The elements of the bounded ultrapower are going to be equivalence classes of bounded sequences in $V(\mathbb{R})$, and the equivalence relation is determined by U.

DEFINITION 15.18. *Two bounded sequences a, b in $V(\mathbb{R})$ are said to be* U**-equivalent***, in symbols $a =_U b$, if*

$$\{n\colon a_n = b_n\} \in U.$$

LEMMA 15.19. *The relation $=_U$ is an equivalence relation on the set of bounded sequences in $V(\mathbb{R})$.*

PROOF. $=_U$ is obviously reflexive and symmetric. We show that $=_U$ is transitive. Assume $a =_U b$ and $b =_U c$. Let

$$X = \{n\colon a_n = b_n\},\ Y = \{n\colon b_n = c_n\},\ Z = \{n\colon a_n = c_n\}.$$

Then $X \in U$ and $Y \in U$, so $X \cap Y \in U$. But $X \cap Y \subseteq Z$, so $Z \in U$, and hence $a =_U c$. ⊣

DEFINITION 15.20. *For each a be a bounded sequence in $V(\mathbb{R})$, we define a_U to be the U-equivalence class of a,*

$$a_U = \{b\colon a =_U b\}.$$

The **bounded ultrapower** $\prod_U V(\mathbb{R})$ *of the set $V(\mathbb{R})$ is the set*

$$\prod_U V(\mathbb{R}) = \{a_U\colon a \text{ is bounded in } V(\mathbb{R})\}.$$

The **natural embedding** *of $V(\mathbb{R})$ into $\prod_U V(\mathbb{R})$ is the mapping*

$$i\colon V(\mathbb{R}) \to \prod_U V(\mathbb{R})$$

defined by

$$i(x) = \langle x, x, x, \dots \rangle_U$$

for each $x \in V(\mathbb{R})$. That is, $i(x) = x$ if $x \in \mathbb{R}$, and $i(x)$ is the U-equivalence class of the constant sequence $\langle x, x, x, \dots \rangle$ otherwise.

The U**-membership relation** $\in_U$ *on $\prod_U V(\mathbb{R})$ is defined by*

$$a_U \in_U b_U \text{ iff } \{n\colon a_n \in b_n\} \in U.$$

LEMMA 15.21. *The relation $a_U \in_U b_U$ depends only on the equivalence classes a_U, b_U. That is, if $a =_U a'$ and $b =_U b'$ then*

$$\{n\colon a_n \in b_n\} \in U \text{ iff } \{n\colon a'_n \in b'_n\} \in U.$$

PROOF. Suppose $\{n\colon a_n \in b_n\} \in U$. Then

$$\{n\colon a'_n \in b'_n\} \supseteq \{n\colon a_n \in b_n\} \cap \{n\colon a_n = a'_n\} \cap \{n\colon b_n = b'_n\}.$$

The right side belongs to U, so the left side belongs to U, and hence $a' \in_U b'$. ⊣

LEMMA 15.22. *The natural embedding $i\colon V(\mathbb{R}) \to \prod_U V(\mathbb{R})$ is one to one.*

Proof. If $x \neq y$ then $\langle x, x, x, \dots \rangle$ and $\langle y, y, y, \dots \rangle$ are never equal, and $\emptyset \notin U$, so $i(x) \neq i(y)$. ⊣

Definition 15.23. *The* ***ultrapower*** $\prod_U \mathbb{R}$ *of* $\mathbb{R}$ *modulo* U *is the set*

$$\prod_U \mathbb{R} = \{b_U : b \text{ maps } \mathbb{N} \text{ into } \mathbb{R}\}.$$

In Section 1G we defined $\prod_U \mathbb{R}$ in a slightly different way, replacing the equivalence class of a constant sequence $\langle r, r, \dots \rangle$ by r itself to make $\prod_U \mathbb{R}$ an extension of $\mathbb{R}$. This time we postpone that replacement to the next step, when we form the set $\mathbb{R}^*$ of hyperreal numbers.

Observe that the following conditions are equivalent:

$$a_U \in \prod_U \mathbb{R},$$
$$\{n : b_n \in \mathbb{R}\} = \mathbb{N} \text{ for some } b =_U a,$$
$$\{n : a_n \in \mathbb{R}\} \in U,$$
$$a_U \in_U i(\mathbb{R}).$$

Lemma 15.24. *The set* $\{i(r) : r \in \mathbb{R}\}$ *is a proper subset of* $\prod_U \mathbb{R}$.

Proof. $\mathbb{R}$ is a subset of $\prod_U \mathbb{R}$ because $i(r) = \langle r, r, \dots \rangle_U \in \prod_U \mathbb{R}$ for each $r \in \mathbb{R}$. Let a be a one to one mapping of $\mathbb{N}$ into $\mathbb{R}$. Then $a_U \in \prod_U \mathbb{R}$. However, for each $r \in \mathbb{R}$, the set $\{n : a_n = r\}$ has at most one element and hence is not in U, so $a_U \neq i(r)$. ⊣

Lemma 15.25. *Suppose* $a_U, b_U \in \prod_U V(\mathbb{R}) \setminus \prod_U \mathbb{R}$ *and* $a_U \neq b_U$. *Then*

$$\{c_U : c_U \in_U a_U\} \neq \{c_U : c_U \in_U b_U\}.$$

Proof. The intersection of the two sets $\{n : a_n \subseteq b_n\}, \{n : b_n \subseteq a_n\}$ does not belong to U, because their intersection $\{n : a_n = b_n\}$ does not belong to U. Therefore one of these sets does not belong to U, say

$$\{n : a_n \subseteq b_n\} \notin U.$$

Then its complement X does belong to U,

$$X = \{n : a_n \setminus b_n \neq \emptyset\} \in U.$$

For each $n \in X$ choose $c_n \in a_n \setminus b_n$, and for $n \in \mathbb{N} \setminus X$ choose $c_n = 0$. Then $c_U \in_U a_U$ but not $c_U \in_U b_U$. ⊣

We have defined the bounded ultrapower and proved some basic lemmas. We now turn to the second stage, the Mostowski collapse.

We begin at the bottom level by forming the set $\mathbb{R}^*$ of hyperreal numbers. As in Section 1G, the idea is to start with $\prod_U \mathbb{R}$ and replace $i(r)$ by r itself for each real r. But this time, we want to use $\mathbb{R}^*$ as the set of atoms of a superstructure $V(\mathbb{R}^*)$, so we must also make sure that no element of $\mathbb{R}^*$ contains any elements of $V(\mathbb{R}^*)$.

LEMMA 15.26. *There is a set* $\mathbb{R}^*$ *and a function* j_0 *such that:*
(i) j_0 *maps* $\prod_U \mathbb{R}$ *one to one onto* $\mathbb{R}^*$*,*
(ii) $j_0(i(r)) = r$ *for each* $r \in \mathbb{R}$*, and hence* $\mathbb{R} \subseteq \mathbb{R}^*$*,*
(iii) $\emptyset \notin \mathbb{R}^*$ *and no element of* $\mathbb{R}^*$ *contains any elements of* $V(\mathbb{R}^*)$*.*

PROOF. There are many ways to do this. One way is to take a set λ of cardinality greater than $V(\mathbb{R})$ and define

$$j_0(a_U) = \begin{cases} r & \text{if } r \in \mathbb{R} \text{ and } a_U = i(r), \\ a_U \times \{\lambda\} & \text{otherwise.} \end{cases}$$

$\dashv$

Hereafter we let $\mathbb{R}^*$ and j_0 be as in Lemma 15.26.

THEOREM 15.27. *There is a unique one to one mapping*

$$j\colon \prod_U V(\mathbb{R}) \to V(\mathbb{R}^*),$$

called the **Mostowski collapse** *of* $\prod_U V(\mathbb{R})$*, such that:*
(i) $j(x) = j_0(x)$ *for* $x \in \prod_U \mathbb{R}$*.*
(ii) $j(a_U) = \{j(b_U)\colon b_U \in_U a_U\}$ *for all* $a_U \in \prod_U V(\mathbb{R}) \setminus \prod_U \mathbb{R}$*.*

PROOF. For each $k \in \mathbb{N}$ we consider the ultrapower of the set $V_k(\mathbb{R})$, which is defined by

$$\prod_U V_k(\mathbb{R}) = \{a_U\colon a \text{ maps } \mathbb{N} \text{ into } V_k(\mathbb{R})\}.$$

Then

$$\prod_U V(\mathbb{R}) = \bigcup_{k=0}^{\infty} \prod_U V_k(\mathbb{R}),$$
$$\prod_U \mathbb{R} = \prod_U V_0(\mathbb{R}) \subseteq \prod_U V_1(\mathbb{R}) \subseteq \prod_U V_2(\mathbb{R}) \subseteq \cdots .$$

We define j restricted to $\prod_U V_k(\mathbb{R})$ by induction on k. For $k = 0$, j restricted to $\prod_U V_0(\mathbb{R})$ is the mapping j_0 from Lemma 15.26. Suppose we already have defined j restricted to $\prod_U V_k(\mathbb{R})$. For $a_U \in \prod_U V_{k+1}(\mathbb{R}) \setminus \prod_U V_k(\mathbb{R})$, define

$$j(a_U) = \{j(b_U)\colon b_U \in_U a_U\}.$$

This definition is unambiguous because by Lemma 15.5, whenever $b_U \in_U a_U$ we have

$$\{n\colon b_n \in V_k(\mathbb{R})\} \supseteq \{n\colon b_n \in a_n\} \in U,$$

so $b_U \in \prod_U V_k(\mathbb{R})$. Conditions (i) and (ii) hold by definition. j is one to one by Lemma 15.25. To show that there is at most one function satisfying (i) and (ii), one proves by induction on k that the restriction of j to $\prod_U V_k(\mathbb{R})$ is the only function with properties (i) and (ii). $\dashv$

Definition 15.28. *The embedding* $*\colon V(\mathbb{R}) \to V(\mathbb{R}^*)$ *is the composition of the natural embedding* i *and the Mostowski collapse* j*, that is, for* $a \in V(\mathbb{R}), a^* = j(i(a))$.

$$\begin{tikzcd}
\prod_U V(\mathbb{R}) \arrow[rr, "j"] & & V(\mathbb{R}^*) \\
& V(\mathbb{R}) \arrow[ul, "i"] \arrow[ur, "*"'] &
\end{tikzcd}$$

Lemma 15.29. *The mapping* $*\colon V(\mathbb{R}) \to V(\mathbb{R}^*)$ *built from a bounded ultrapower* $\prod_U V(\mathbb{R})$ *is a superstructure embedding.*

Proof. Since i and j are one to one, their composition $*$ is one to one. By Lemmas 15.24 and 15.26, $\mathbb{R}$ is a proper subset of $\mathbb{R}^*$ and $r^* = r$ for each $r \in \mathbb{R}$. Taking $\mathbb{R}$ as an element of $V(\mathbb{R})$, we have

$$j(i(\mathbb{R})) = j(\langle \mathbb{R}, \mathbb{R}, \dots \rangle_U) = \left\{ j(a_U)\colon a_U \in \prod_U \mathbb{R} \right\} = \mathbb{R}^*.$$

Finally, for $x, y \in V(\mathbb{R})$,

$$x \in y \text{ iff } i(x) \in_U i(y) \text{ iff } j(i(x)) \in j(i(y)) \text{ iff } x^* \in y^*.$$

$\dashv$

The following theorem of Łoś will show that Leibniz' Principle holds for this superstructure embedding.

Theorem 15.30. *(Łoś' Theorem) Given a bounded ultrapower* $\prod_U V(\mathbb{R})$*, let* $\varphi(v^1, \dots, v^n)$ *be a bounded formula with no constants and let* $a_U^1, \dots, a_U^n \in \prod_U V(\mathbb{R})$*. Then*

$$\varphi(j(a_U^1), \dots, j(a_U^n)) \textit{ is true}$$

if and only if

$$\{k \in \mathbb{N}\colon \varphi(a^1(k), \dots, a^n(k))\} \in U.$$

Proof. We use induction on the complexity of formulas. The theorem holds for atomic formulas by the definition of $a_U = b_U$ and $a_U \in_U b_U$. For the induction steps it suffices to assume the theorem for $\varphi(v^1, \dots, v^n)$ and $\psi(v^1, \dots, v^n)$ and prove the theorem for $\neg\varphi$, $\varphi \wedge \psi$, and $(\exists v^1 \in v^2)\varphi$.

For $\neg\varphi$ we observe that the following are equivalent.

$$\neg\varphi^*(j(a_U^1), \dots, j(a_U^n))$$
$$\{k \in \mathbb{N}\colon \varphi(a^1(k), \dots, a^n(k))\} \notin U$$
$$\{k \in \mathbb{N}\colon \neg\varphi(a^1(k), \dots, a^n(k))\} \in U$$

For $\varphi \wedge \psi$ we use the following equivalences.

$(\varphi \wedge \psi)^*(j(a_U^1), \dots, j(a_U^n))$

$\varphi^*(j(a_U^1), \dots, j(a_U^n)) \wedge \psi^*(j(a_U^1), \dots, j(a_U^n))$

$\{k \in \mathbb{N} \colon \varphi(a^1(k), \dots, a^n(k))\} \in U$ and $\{k \in \mathbb{N} \colon \psi^1(a(k), \dots, a^n(k))\} \in U$

$\{k \in \mathbb{N} \colon \varphi(a^1(k), \dots, a^n(k))\} \cap \{k \in \mathbb{N} \colon \varphi(a^1(k), \dots, a^n(k))\} \in U$

$\{k \in \mathbb{N} \colon (\varphi \wedge \psi)(a^1(k), \dots, a^n(k))\} \in U.$

The theorem for $(\exists v^1 \in v^2)\varphi$ again follows from a sequence of equivalent statements.

$(\exists v^1 \in j(a_U^2))\varphi^*(v^1, j(a_U^2), \dots, j(a_U^n))$

For some $a_U^1 \in_U a_U^2, \varphi^*(j(a_U^1), \dots, j(a_U^n))$

$\{k \in \mathbb{N} \colon$ For some $a^1(k) \in a^2(k), \varphi(a^1(k), \dots, a^n(k))\} \in U$

$\{k \in \mathbb{N} \colon (\exists v^1 \in a^2(k))\varphi(v^1, a^2(k), \dots, a^n(k))\} \in U.$

$\dashv$

Theorem 15.31. *There exists a nonstandard universe. In fact, for each free ultrafilter U over $\mathbb{N}$, the superstructure embedding $* \colon V(\mathbb{R}) \to V(\mathbb{R}^*)$ built by the bounded ultrapower modulo U is a nonstandard universe.*

Proof. By Theorem 1.42 there exists a free ultrafilter U over $\mathbb{N}$. For a real bounded sentence ψ, Łoś' Theorem 15.30 shows that ψ^* is true if and only if $\{k \in \mathbb{N} \colon \psi \text{ is true}\} \in U$. Since $\mathbb{N} \in U$, this is equivalent to the truth of ψ. Hence Leibniz' Principle holds for $* \colon V(\mathbb{R}) \to V(\mathbb{R}^*)$. $\dashv$

15E. Saturation and Uniqueness

By Theorem 1.38, the field of real numbers is, up to isomorphism, the unique complete ordered field. Furthermore, there is a definable complete ordered field.

In this section we present the analogous results for hyperreal number systems and for nonstandard universes. The theorems in this section use some fairly advanced methods from model theory and are stated without proof.

The hyperreal number system $(*, \mathbb{R}, \mathbb{R}^*)$ is not uniquely characterized by Axioms A–E, even up to isomorphism. This lack of uniqueness does not matter much in practice, but can be unsettling. In this section we remedy the situation by introducing one more axiom, the Saturation Axiom, which is important in applications beyond calculus and which does give uniqueness up to isomorphism. Saturation is different from completeness but has a similar appeal.

The most natural formulation of the Saturation Axiom requires the existence of an uncountable inaccessible cardinal. The default foundation for mathematics is usually taken to be the system ZFC, Zermelo-Fraenkel set theory plus the Axiom of Choice. The Axiom of Inaccessibility in set theory, which says that there are uncountable inaccessible cardinals, cannot be proved in ZFC (if ZFC is consistent). But the justification of the axioms of ZFC, based on the idea of a cumulative hierarchy of sets, also justifies the Axiom of Inaccessibility. For this reason, it is considered acceptable to add the Axiom of Inaccessibility to ZFC when convenient.

We now introduce the notion of an inaccessible cardinal and state the Axiom of Inaccessibility in set theory. We will then introduce the Saturation Axiom for hyperreal numbers, and an analogous property for nonstandard universes.

Definition 15.32. *A cardinal number κ is said to be* **inaccessible** *if:*

(i) For any set x of cardinality less than κ, the power set $\mathcal{P}(x)$ has cardinality less than κ.

(ii) For any set of sets $X = \{x_i : i \in I\}$ such that I and each x_i has cardinality less than κ, the union $\bigcup_{i \in I} x_i$ has cardinality less than κ.

The **Axiom of Inaccessibility** *in set theory is the axiom that there exists an uncountable inaccessible cardinal.*

The first infinite cardinal $\aleph_0$ is inaccessible. In ZFC, the axiom of infinity gives the existence of $\aleph_0$. The Axiom of Inaccessibility is the simplest example of a strong axiom of infinity, and is a common addition to the axioms of ZFC.

We now state the Saturation Axiom for a hyperreal number system $(*, \mathbb{R}, \mathbb{R}^*)$.

Saturation Axiom

Let S be a set of equations and inequalities involving real functions, hyperreal constants, and variables, such that S has smaller cardinality than $\mathbb{R}^*$. If every finite subset of S has a hyperreal solution, then S has a hyperreal solution.

To state the uniqueness theorem, we need the notion of an isomorphism between two hyperreal number systems.

Definition 15.33. *Let $(*, \mathbb{R}, \mathbb{R}^*)$ and $(\bullet, \mathbb{R}, \mathbb{R}^\bullet)$ be two hyperreal number systems with the same real part $\mathbb{R}$. An* **isomorphism** *between them is a mapping $h \colon \mathbb{R}^* \to \mathbb{R}^\bullet$ such that:*

(i) $h(r) = r$ for each $r \in \mathbb{R}$,

(ii) h is an ordered field isomorphism from $\mathbb{R}^*$ onto $\mathbb{R}^\bullet$,

(iii) For each real function f of n variables and $x_1, \dots, x_n \in \mathbb{R}^*$,

$$f^\bullet(hx_1, \dots, hx_n) = h(f^*(x_1, \dots, x_n)).$$

Two hyperreal number systems with the same real part $\mathbb{R}$ are said to be **isomorphic** *if there is an isomorphism between them.*

Theorem 15.34. *Assume the Axiom of Inaccessibility. There for each complete ordered field* $\mathbb{R}$ *there is up to isomorphism a unique structure* $(*, \mathbb{R}, \mathbb{R}^*)$ *which satisfies Axioms A–E and the Saturation Axiom, such that the cardinality of* $\mathbb{R}^*$ *is the first uncountable inaccessible cardinal.*

One might ask whether there exist structures in cardinalities other than the first uncountable inaccessible cardinal which satisfy Axioms A–E and the Saturation Axiom. The set theory ZFC plus Axiom of Inaccessibility is not strong enough to decide this question, which depends on the behavior of cardinal exponents. However, the analogue of Theorem 15.34 can be shown to hold for every uncountable inaccessible cardinal. If the generalized continuum hypothesis never holds, i.e. $\kappa^+ < 2^\kappa$ for all infinite cardinals κ, then such structures exist only in uncountable inaccessible cardinals. So it is natural to require the size of the hyperreal structure to be the first uncountable inaccessible cardinal.

There is a similar uniqueness theorem for nonstandard universes.

Definition 15.35. *A nonstandard universe* $*\colon V(\mathbb{R}) \to V(\mathbb{R}^*)$ *is said to be* ***saturated*** *if for every set* X *of internal sets such that the cardinality of* X *is less than the cardinality of* $\mathbb{R}^*$*, if every finite subset of* X *has nonempty intersection then* X *has nonempty intersection.*

We say that $*\colon V(\mathbb{R}) \to V(\mathbb{R}^*)$ *has the* **inaccessibility property** *if both* $\mathbb{R}^*$ *and the set of all internal sets have cardinality equal to the first uncountable inaccessible cardinal.*

Definition 15.36. *Let* $*\colon V(\mathbb{R}) \to V(\mathbb{R}^*)$ *and* $\bullet\colon V(\mathbb{R}) \to V(\mathbb{R}^\bullet)$ *be two nonstandard universes with the same real part* $\mathbb{R}$*. An* **isomorphism** *between them is a mapping* $h\colon V(\mathbb{R}^*) \to V(\mathbb{R}^\bullet)$ *such that:*

(i) $h(r) = r$ *for each* $r \in \mathbb{R}$*,*

(ii) h *maps* $\mathbb{R}^*$ *one to one onto* $\mathbb{R}^\bullet$*,*

(iii) For each $X \in V(\mathbb{R}^*) \setminus \mathbb{R}^*$*,*

$$h(X) = \{h(u) : u \in X\}.$$

(iv) For each $A \in V(\mathbb{R})$*,* $h(A^*) = A^\bullet$*.*

Two nonstandard universes with the same real part $\mathbb{R}$ *are said to be* **isomorphic** *if there is an isomorphism between them.*

Here is an easy lemma.

Lemma 15.37. *If* h *is an isomorphism between two nonstandard universes* $*\colon V(\mathbb{R}) \to V(\mathbb{R}^*)$ *and* $\bullet\colon V(\mathbb{R}) \to V(\mathbb{R}^\bullet)$*, then* $h(A) = A$ *for each* $A \in V(\mathbb{R})$*,* h *maps* $V(\mathbb{R}^*)$ *one to one onto* $V(\mathbb{R}^\bullet)$*, and* h *is an extension of an isomorphism between the elementary parts of* $*\colon V(\mathbb{R}) \to V(\mathbb{R}^*)$ *and* $\bullet\colon V(\mathbb{R}) \to V(\mathbb{R}^\bullet)$*.*

We can now state a uniqueness theorem for nonstandard universes.

Theorem 15.38. *Assume the Axiom of Inaccessibility. Then for each complete ordered field* $\mathbb{R}$ *there is up to isomorphism a unique nonstandard universe* $*\colon V(\mathbb{R}) \to V(\mathbb{R}^*)$ *which is saturated and has the inacessibility property.*

Theorems 15.34 and 15.38 have analogues in ordinary ZFC set theory, without the Axiom of Inaccessibility. However, these results replace the notion of a saturated structure with a more complicated notion, called a special structure. Here we have chosen to assume the Axiom of Inaccessibility in order to get a more natural uniqueness result.

One can also prove a result like Theorem 15.34 but where $\mathbb{R}^*$ is a proper class. This approach would require an extension of ZFC with proper classes, and a corresponding result for nonstandard universes would need a notion of superstructure over a proper class.

In Section 1G we presented the result of Kanovei and Shelah [KS 2004], which gives a definable hyperreal number system $(*, \mathbb{R}, \mathbb{R}^*)$ that satisfies Axioms A–E. This was done with an iterated ultrapower. One can also build definable structures in Theorems 15.34 and 15.38.

In the set theory ZFC plus the Axiom of Inaccessibility, we say that a set X is **definable** by a first order formula $\theta(v)$ if one can prove that X is the unique set such that $\theta(X)$ holds. The following two theorems are implicit in the paper of Kanovei and Shelah, and are proved with an iterated ultrapower which is similar to but more elaborate than the one in Section 1G .

Theorem 15.39. *Assume the Axiom of Inaccessibility. There is a definable hyperreal number system* $(*, \mathbb{R}, \mathbb{R}^*)$ *which satisfies Axioms A–E and the Saturation Axiom, such that the cardinality of* $\mathbb{R}^*$ *is the first uncountable inaccessible cardinal.*

Theorem 15.40. *Assume the Axiom of Inaccessibility. There is a definable nonstandard universe* $*\colon V(\mathbb{R}) \to V(\mathbb{R}^*)$ *which is saturated and has the inaccessibility property.*

The superstructure method given here is not the only approach to infinitesimal analysis. Other approaches extend the language of ZFC by adding new primitive symbols to the language in some way. One of these, Nelson's Internal Set Theory, has been used extensively, and several extensions of this theory have been studied (see the article of Hrbacek in [CNR 2006] for a survey).

The superstructure method has been the most common approach in the literature, because it stays close to the traditional classical foundations of mathematics, working within ZFC, or ZFC with the Axiom of Inaccessibility. Theorems 15.38 and 15.40 strengthen this point by showing that one can work with a nonstandard universe which is definable and is uniquely characterized up to isomorphism by Leibniz' Principle, saturation, and the inaccessibility property.

REFERENCES

[AFHL 1986] S. Albeverio, J.-E. Fenstad, R. Hoegh-Krohn, and T. Lindstrom, *Nonstandard Methods in Stochastic Analysis and Mathematical Physics*, Academic Press (1986)

[ACH 1997] L.O. Arkeryd, N.J. Cutland, and C.W. Henson, *Nonstandard Analysis, Theory and Applications*, Kluwer (1997).

[Buck 1965] R.C. Buck, *Advanced Calculus*, McGraw-Hill (Second Edition, 1965).

[CK 1990] C.C. Chang and H.J. Keisler, *Model Theory*, North-Holland (1990)

[CNR 2006] N.J. Cutland, M. Di Nasso, and D.A. Ross, *Nonstandard Methods and Applications in Mathematics*, *Lecture Notes in Logic*, Association for Symoblic Logic (2006).

[Goldblatt 1991] R. Goldblatt, *Lectures on the Hyperreals*, Springer (1991).

[Henkin 1950] L.A. Henkin, Completeness in the Theory of Types, *Journal of Symbolic Logic* vol. 15 (1950), pages 159-166.

[Hewitt 1948] E. Hewitt, Rings of real-valued continuous functions, *Transactions of the American Mathematical Society* vol. 64 (1948), pages 45-99.

[HL 1985] A.E. Hurd and P.A. Loeb, *An Introduction to Nonstandard Real Analysis*, Academic Press (1985).

[KS 2004] V. Kanovei and S. Shelah, A Definable Nonstandard Model of the Reals, *Journal of Symbolic Logic* vol. 69 (2004), pages 159-164.

[Keisler 2000] H.J. Keisler, *Elementary Calculus*, Online Edition, www.math.wisc.edu/~keisler.

[Kei 1986] H.J. Keisler, *Elementary Calculus*, Prindle, Weber and Schmidt (First Edition 1976, Second Edition 1986).

[Keisler 1976] H.J. Keisler, *Foundations of Infinitesimal Calculus*, Prindle, Weber and Schmidt (1976).

[Łoś 1955] J. Łoś, Quelques remarques, théorèmes et problèmes sur les classes définissables d'algèbres, *Mathematical Interpretation of Formal Systems*, North-Holland (1955).

[Robinson 1970] A. Robinson, *Nonstandard Analysis*, North-Holland (1970).

[Skolem 1934] T. Skolem, Über di Nicht-Charakterisierbarkeit der Zahlenreighe mittels endlich oder abzählbar unendlicht vieler Aussagen mit ausschliesslich Zahlenvariablen, *Fundamentae Mathematica* vol. 23 (1934), pages 150-161.

[Stroyan 1997] *Mathematical Background: Foundations of Infinitesimal Calculus*, Second Edition, Academic Press (1997).

[SL 1976] K.D. Stroyan and W.A.J. Luxemburg, *Introduction to the Theory of Infinitesimals*, Academic Press (1976).

Index

absolute value 20
absolutely convergent 108
Addition Property 59,75,135,146
almost linearly dependent 115
almost parallel 115
antiderivative 64
Archimedean Property 12,22,187
area function 57
Associative Law 19
atomic formula 174,180
Axiom of Choice 23
Axiom of Inaccessibility 193
Barwise, J. ix
basic closed region 128,129
basic open region 129
basic polar region 94
basis vectors 113
Behrens, M.F. 93
Bolzano-Weierstrass Theorem 107
bound variables 174
boundary 119
boundary curve 154
bounded 17,22
bounded formula 180
bounded in $V(\mathbb{R})$ 187
bounded sentence 105,180
bounded ultrapower 188
Cauchy Condition 21,96
chain 23
Chain Rule 41
circumscribed rectangle 136,143
closed set 16,119
closure 16
Commutative Law 19
commutative ring 19
compact 43
complete ordered field 21,187
components 113
constant 7
Constant Rule 38
constant term 8,174
contains a solution 9
continuous 42,43,117,120
$*$-continuous 185
converge 83,103,107
coset 19
countable 25
critical point 51,128
Critical Point Theorem 53,130
cumulative power set 178
Cylinder Property 136,146
decreasing 48
definable complete ordered field 21
definable 28,195
definite integral 59
dependent variable 32
depends on x 32
derivative 31
differentiable 31,121
differential equation 159
directed line segment 113
distance 17
Distributive Law 19
diverge 86,103,107
double integral 137
double Riemann sum 136
element of area 138
Elementary Extension Principle 178
elementary part 181
equation 8
equivalence relation 19
Extreme Value Theorem 51,54,128,129
Euler approximation 160,164
exponential function 99
external 183
field 20
field extension 20
field of view 34
filter 24
finite 2,18,114
formula 8,174
Frèchet filter 24
free ultrafilter 24
free variable 174
Frege, F.L.G 183
Function Axiom 9
function of x 32
Fundamental Theorem of Calculus 66
galaxy 2,114
Generalized Mean Value Theorem 74
geometric series formula 100,109
Goldblatt, R. ix
Green's Theorem 156
Heine-Borel Theorem 107
Henkin, L.A. 189
Hewitt, E 23

homomorphism 19
Hrbacek, K 197
hyperfinite Euler approximation 160
hyperintegers 45
hyperrationals 98
hyperreal closed interval 46
hyperreal cover 130
hyperreal entity 180
hyperreal number 2,9,180
hyperreal plane 17
hyperreal sets 180
hyperreal solution 9
hyperreal subinterval 46
hyperreal vector 114
ideal 19
Identity Law 19,20
image 144
implicit function 124
Implicit Function Theorem 126,128
improper integral 85
inaccessibility property 194
inaccessible 193
increasing 48
increment 32,121
Increment Theorem 35
indefinite integral 65
independent variable 32
index set 23
induction 22
inequality 8
infinite 2,114
infinite lower Riemann sum 62
infinite partial sum 106
infinite partition 47
infinite Riemann sum 58,137,150
infinite sequence 103
infinite series 107
Infinite Sum Theorem 77,140
infinite telescope 36
infinitely close 2,18,114
infinitely close compared to $\triangle x$ 32
infinitesimal 2,114
infinitesimal interval 46
infinitesimal microscope 33
initial value problem 159
inner product 113
integral equation 159
interior 16,119
interior critical point 51
Intermediate Value Theorem 49,54
internal 183
Internal Definition Principle 188
interval 22
inverse function 89
Inverse Function Theorem 92
Inverse Law 19,20
inverse relation 89
isomorphism 19,193
Iterated Integral Theorem 139
Jacobian 144
Jordan content 143
Kanovei, V. vii,1,23,197
Keisler, H.J. vii
language 173
Leibniz, G.W. vii,183
Leibniz' Principle 181
length 80,82,113
l'Hospital's Rule 75
limit 41,42
Limit Comparison Test 109
line integral 149
linearly dependent 115
Lipschitz bound 165
Local Inverse Function Theorem 93
local maximum 49
locally Lipschitz 165
logarithmic function 100
Łoś, J. 23,193
lower Riemann sum 62
Luxemburg, W.A.J. 93
maximum 49,126
maximum critical point 130
maximum value 126
Mean Value Theorem 53,54,134
membership relation 179
minimum 49,126
minimum value 126
monad 2,18,114
Mostowski, A 189
Mostowski collapse 190
n-th differential 40
n-th increment 110
natural embedding 188
natural extension 7,9,14,17
natural number 22
Nelson, E. 197
nonstandard universe 181
open set 16,119
order isomorphic 21
ordered field 20
Overspill Principle 188
parallel 115
parametric curve 82
partial derivatives 120

Partial Solution Theorem 12
partial sum sequence 106
path 82
Peano, G. 183
Peano Existence Theorem 165
piecewise smooth curve 152
plane 17
polar area function 94
polar rectangle 94
polygonal function 160
Power Rule 40
power set 178
Product Law 20
Product Rule 38
Quotient Rule 38
real bounded formula 180
real direction 115
real entity 180
real function 7
real length 115
real neighborhood 16,17,119
real number 2,9,21,180
real plane 17
real relation 17
real set 180
real solution 8
real vector function 116
Rectangle Property 59
Reflexive Law 19
region under the curve 57
reparametrization 82
Reparametrization Theorem 84
Riemann sum 58
ring 19
Robinson,A vii,viii,189
Robinson's Principle 189
S-continuous 184
Saturation Axiom 23,194,195
scalar multiple 113
Second Derivative Test 54
Second Fundamental Theorem 68
sentence 174
Shelah, S vii,1,23,197
simple curve 82
Skolem, T 23
Skolem function 177
slope 31
smooth curve 149
smooth transformation 144
smooth 80,82,121
standard part 5,114,161
Standard Part Principle 5
standard set 183
Stroyan, K ix,33,35,93
Symmetric Law 19
subfield 20
subring 19
subsequence 105
Sum Law 20
Sum Rule 38,146
superstructure 179
superstructure embedding 179
surface area 84
system of formulas 8
tangent line 32
tangent plane 121
Taylor's Formula 110
term 8,174
total differential 121
Transfer Axiom 9
$*$-transform 175, 181
Transitive Law 11,19
Trichotomy Law 11,20
U-equivalent 188
U-equivalent 24
U-membership relation 188
ultrafilter 24
ultrapower 25
unbounded 22
uniformly continuous 43
uniformly differentiable 53,121
Uniqueness Theorem 168
unit vector 115
upper Riemann sum 62
variable 7
vector 113
vector derivative 117
vector difference 113
vector differential 117
vector increment 117
vector sum 113
volume 137
volume function 135
Zermelo-Fraenkel set theory 19
zero vector 113
Zorn's Lemma 23