ISBN: 978-1-60805-178-6

Scientific Natural Philosophy

Author:
E.E. Escultura
JNT University
India

Bentham e Books

SCIENTIFIC NATURAL PHILOSOPHY

By

E. E. Escultura

Research Professor

GVP – Professor V. Lakshmikantham Institute for Advanced Studies and Department of Physics, GVP College of Engineering, JNT University, Madurawada Visakhapatnam, AP, India

http://users.tpg.com.au/pidro/ http://edgareescultura.wordpress.com/

International Federation of Nonlinear Analysts

CONTENTS

Foreword *i*

Preface *ii*

General Introduction *iii*

CHAPTERS

1 Philosophical Foundations of Mathematics and Physics 3

2 The Mathematics of the Grand Unified Theory 10

3 The Grand Unified Theory 60

4 Applications 108

5 Scientific Natural Philosophy 149

Appendix 162

Glossaries 165

References 167

Index 171

FOREWORD

It is astonishing how wide the enumeration of scientific fields is in this book. One can find here foundations of new mathematics with the new real number system; philosophical foundations of the new physics; theory of evolution; qualitative modeling and quantum gravity; cosmology and macro gravity; theory of intelligence and the nature of thought; geological and atmospheric turbulence and others. In each area of science the author has proposed new nonstandard approaches to its development and solved long-standing problems. The author has formulated about 50 laws of nature that cover vast areas of natural science.

His Grand Unified Theory (GUT) must find a deserving place among other theories claiming to be named Unified Theories. GUT as well as its theoretical applications is based on about 50 natural laws. Foremost among them are: the existence of two fundamental states of matter – visible and dark – and the basic constituent of matter, the superstring, that comprises both fundamental states. Suitable agitation converts dark or non-agitated and semi-agitated superstrings to agitated superstrings or visible matter. The author's attempt to lift the veil over dark matter deserves attention.

The author's important contributions in the new mathematics are the resolution of Fermat's Last Theorem (FLT); the solution of the gravitational n-body problem using the integrated Pontrjagin maximum principle; and the construction of the new real number system (new reals) including the dark number that qualitatively models the superstring.

There are many other new scientific approaches in the book. However, no one on Earth possesses Truth at the highest level and this book while proposing new approaches calls the reader to further investigation.

V. Gudkov

Professor

Institute of Mathematics and Computer Science

University of Latvia, Rainis Boulevard, 29, Riga, LV-1459, Latvia

PREFACE

Scientific natural philosophy is distinguished from previous philosophies for being based on the grand unified theory (GUT). Its methodology of qualitative modeling *explains* how nature works, while quantitative modeling *describes* nature's appearances mathematically. Therefore, GUT can be understood only as self-contained unified physical theory based on the laws of nature. However, the two methodologies are complementary, each indispensable to the other.

Chapter 1 deals with the foundations of mathematics and physics. Chapter 2 surveys the classical and new mathematics involved in the development of GUT but provides full details on the new real number system. Chapter 3 is the formulation of GUT through its three pillars – quantum and macro gravity and thermodynamics. Chapter 4 offers a wide range of applications of GUT from the theory of intelligence and the Earth sciences through astronomy, cosmology, biology, physical psychology, genetic engineering for the treatment of genetic diseases and optimal control theory through the integrated Pontrjagin maximum principle in the derivation of the quantitative component of the solution of the n-body problem. Chapter 5, the final and main chapter, provides the philosophical integration and summation of this work. Except for Chapter 2 the book is accessible to the general readership.

The author acknowledges with deep appreciation the contributions of colleagues, Professors C. G. Jesudason of the University of Malaya, for sustained debate on GUT, E. de la Cruz of California State University, Northridge, for constructive criticisms, and V. Gudkov of the University of Latvia for continued support and collaboration, all of whom contributed immensely to this book. Most of all, greatly indebted to Professor V. Lakshmikantham for paving the way to the frontiers of mathematics and science.

E. E. Escultura

Research Professor

GVP – V. Lakshmikantham Institute for Advanced Studies

GVP College of Engineering, JNT University, Madurawada Visakhapatnam 530041 AP

India

GENERAL INTRODUCTION

The novelty of our methodology calls for this general introduction to initiate the reader to the grand unified theory (GUT), the core of scientific natural philosophy. The conventional methodology of physics, quantitative modeling, has left long-standing problems unsolved, e.g., the turbulence and gravitational n-body problems, and fundamental questions unresolved, e.g., what the basic constituent of matter and the structure of the electron are, an inadequacy that prompted the author to introduce an alternative methodology – qualitative modeling – that digs deeper into nature beyond appearances to find out how it works. It explains not only the appearances of nature revealed by natural phenomena but also their dynamics including natural forces, interactions and behavior in terms of the laws of nature. Natural laws are discovered by observing patterns and regularity in nature and articulating them as natural laws upon which a physical theory such as GUT is built to express scientific knowledge as a deductive system subject to the most updated standards of mathematical rigor and precision. Thus, qualitative modeling gave birth to theoretical physics where there was only mathematical physics. The validity of a physical theory rests on its ability to explain natural phenomena and make verifiable predictions including invention of technology that works. Anytime a contradiction arises in a physical theory within its own structure or from experimental results it goes down the drain unless it is fixed by suitable modification or discovery of natural law that resolves it.

Just as all concepts of a mathematical space are defined by its axioms and conclusions derived from them, all physical concepts and structures, properties and interactions are determined and defined by the laws of nature and conclusions and predictions derived from them. For example, the structure and properties of the superstring are derived from natural laws. In other words, this new methodology axiomatizes physics as a deductive system where the axioms are laws of nature. In effect, it alters the primary task of the physicist from computation and measurement to the search for the laws of nature.

The boundaries between dark and visible matter and semi-and non-agitated superstrings are based on the finest arc length of visible light which may be refined in the future but our analysis will prevail as long as it is based on the laws of nature.

CHAPTER 1

Philosophical Foundations of Mathematics and Physics

Abstract: This chapter stands on the major rectification of foundations by David Hilbert in the early 20th Century and more. Recognizing the ambiguity of an individual thought being inaccessible to others Hilbert proposed that mathematics should be concerned not only with individual thought but with its representation by objects that can be studied collectively, including symbols, subject to consistent axioms. We stand on Hilbert's rectification and avoid other sources of ambiguity, where contradiction usually hides, such as infinity, large and small numbers and vacuous concepts and propositions. Turning on the real number system we find the field axioms inconsistent and rectification paves the way for its reconstruction into the new real number system under new set of axioms.

In physics we wonder why there are long-standing unsolved problems such as the gravitational n-body and turbulence problems and unanswered fundamental questions like what are the basic constituents of matter and the structure of the electrons. We conclude that its present methodology of quantitative modeling that describes nature's appearances mathematically is quite inadequate, and proceed to complement it with qualitative modeling that explains not only its appearances but also how it works in terms of its laws. Either methodology is indispensable to the other. Using qualitative modeling we proceed to participate in the 5,000-year search for the basic constituent of matter and succeed in pinning it down to the last detail of its structure called generalized nested fractal sequence. The superstring was the crucial factor in the solution of the 200-year-old gravitational n-body problem.

INTRODUCTION

We pose the problems and fundamental questions of mathematics and physics to guide the development of later chapters which will provide solutions and answers. We also summarize the critique-rectification of the foundations of mathematics that paved the way for the development of the mathematics of GUT.

THE UNSOLVED PROBLEMS OF MATHEMATICS

Mathematics has a number of unsolved problems but we focus on the most famous one, the 360-year-old conjecture called Fermat's last theorem [17, 118] that says,

For $n > 2$, the equation,

$$x^n + y^n = z^n, \tag{1}$$

has no solution in positive integers.

The conjecture appears simple and clearly stated but on closer look it is actually vague and it took the critique of its underlying fields – foundations, number theory and the real number system – to realize it. Then rectification was carried out to remedy the ambiguity and resolve the problem. For example, Peano's postulates that define the natural numbers are inadequate; number theorists do not even bother with them and use instead the real number system.

CRITIQUE-RECTIFICATION OF FOUNDATIONS, THE REAL NUMBER SYSTEM

We do not bother with the natural numbers because once the real numbers are fixed as the new real number system with the decimals as subspace and base we establish their isomorphism with the integers, i.e., the integral parts of decimals. We build on the great contribution of David Hilbert – the recognition of the ambiguity of the concepts of individual thought being inaccessible to others and, therefore, can neither be studied collectively nor axiomatized nor the subject matter of mathematics. For precision and clarity, the subject matter of mathematics can only be objects in the real world that everyone can look at and study such as symbols and material objects provided they are defined by consistent premises or axioms that specify their existence, behavior, properties and relationship among them. Then we call such system of objects and its axioms mathematical space. Thus, the game of chess is a mathematical space where the axioms are its rules. Tank and naval warfare are mathematical spaces where the appropriate principles of $\mathbf{R}^3$ and nature of weapons, tanks and battle ships are the axioms. In fact, tank battles inspired the fields of mathematics called game theory and operations research. In arithmetic, some properties of the decimals are taken as axioms. Without consistency a mathematical system collapses since any conclusion from it can be contradicted by another. The choice of the axioms is arbitrary depending on what the mathematician wants his mathematical space to do but once chosen, they become binding, i.e., every conclusion must follow from and every construction be justified by them to make it a deductive system.

Axiomatization objectifies a mathematical system and rids it of subjectivism by the mathematician as everything rests on the axioms including the rules of inference. The axioms well-define the mathematical system and its concepts completely and leave no room for universal rules of inference like formal logic since the latter has nothing to do with them. In this regard, all proofs of theorems involving mappings between distinct mathematical spaces are flawed. This applies to Gödel's incompleteness theorems [54 – 61].

This clarification is important as it puts an end to the confusion regarding the equation 1 = 0.99… that has generated much debate online during the last 13 years. These objects are distinct like *apple* and *orange* and to say that apple = orange certainly makes no sense.

The problem that confronted Hilbert was how to insure the consistency of a mathematical system. He proposed a consistent physical or mathematical model, i.e., an isomorphism between the model and the mathematical system. A physical model is, of course, consistent since its behavior is subject to the laws of nature which are consistent, otherwise, our universe would have collapsed a long time ago. It did not and has existed and evolved to higher order since its birth 8 billion years ago [42], e.g., in our young universe there were neither biological species nor biological laws that we now enjoy. To establish the isomorphism the binary operations of the mathematical system must have counterpart binary operations in the physical model. However, since our universe (distinct from the timeless and boundless Universe of dark matter [25] where our visible universe is a local bubble along with other universes) is finite and discrete, only a finite model is possible and the mathematical systems we can check for consistency are limited to finite systems which are quite inadequate for the purposes of science. The simplest and most developed mathematical system during Hilbert's time was arithmetic.

However, the incompleteness of arithmetic makes it ambiguous and unsuitable for modeling mathematical systems since ambiguity often hides inconsistency. The incompleteness of arithmetic means the existence of true proposition that has no proof. The absence of proof is due to the ambiguity of concepts, i.e., ill-defined by its axioms, and the present axioms of arithmetic (field axioms) are inconsistent [93]. Therefore, we look elsewhere to insure consistency of mathematical systems by identifying and avoiding the sources of ambiguity and contradiction or minimizing their impact. They are:

1) Large and small numbers due to our limited capability to compute their digits even with the most advanced technology.
2) Vacuous concept. For example, the concept "the root of the equation $x^2 + 1 = 0$ in the set of real numbers" is vacuous because the equation has no root and yet the concept $i = \sqrt{(-1)}$ is presented as its root which does not exist. Consequently, it yields these contradictions:

$$i = \sqrt{(-1)} = \sqrt{(1/-1)} = 1/\sqrt{(-1)} = 1/i = -i, \quad \textbf{(2)}$$

from which follows, 1 = 0, i = 0 and, for any real number r, r = 0, and both the real and complex number systems collapse. The remedy for the complex plane is in the appendix to [19]. Another vacuous concept is "the largest integer N". By the trichotomy axiom one and only one of the following holds: $N < 1$, $N = 1$, $N > 1$. The first inequality is out and if $N > 1$, then $N^2 > N$, contradicting the definition of N as the largest integer. Therefore, N = 1. This is the original version of the Perron paradox [117, 118]. (The trichotomy axiom is false in the real number system but follows from the lexicographic linear ordering of the consistent new real number system [21])

3) Self-referent or circular proposition where the conclusion applies or refers to the hypothesis. All the Russell paradoxes belong to this type; so does the indirect proof. We cite a few examples to illustrate the problems it gives rise to and find a remedy if possible.

a) (Bertrand Russell) Let M be the set of all sets where each element does not belong to itself, i.e. M $=\{m: m \notin m\}$. Then, either $M \in M$ or $M \notin M$. If $M \in M$, the defining conditions for M holds and $M \notin M$. On the other hand, if $M \notin M$, then M satisfies the defining condition; therefore $M \in M$. Self-reference follows from the fact that a set is defined by its elements and each element is defined by its membership in the set. It is a vicious cycle [78].

b) The famous Russell antimony: A Cretan (native to Crete) saying "All Cretans are liars." Is he telling the truth?

c) The barber paradox: The barber of Seville shaves those and only those who do not shave themselves. Who shaves the barber?

Russell's prescription to avoid this kind of difficulty is simply to keep away from it. The contradiction in (b) can be avoided by inserting the disclaimer "except me" after the phrase, "All Cretans". Similarly, the problem in (c) can be resolved by inserting "except himself" after the phrase "The barber of Seville".

4) An infinite set is ambiguous since we do not know all its elements. Consequently, any statement involving the universal or existential quantifier on infinite set is ambiguous. Such statement is unverifiable and any definite or categorical statement about an ambiguous concept is also ambiguous. For example, to verify that every element of an infinite set has property *A*, we check an element to see if it has the property and keep checking "every" element. Obviously, we cannot exhaust all of its elements and thus we cannot verify if the property *A* holds for all elements of the set. The same problem is true of the existential quantifier. We cannot exhaust the elements of an infinite set to check that there is indeed such element with the specified property.

In fact, Lakatos' counterexample to every phase in Cauchy's proof of Euler's formula relating the edges, vertices and faces of a polyhedron in $\mathbf{R}^3$ [77] can be attributed to the fact that the set of such polyhedra is infinite and some exception hides behind each claim in the proof.

In mathematics, particularly theory of numbers, there are many statements involving infinite set of natural numbers that raise questions still unanswered, i.e., the statements are neither proved nor disproved. They usually stem from the inherent ambiguity of infinite set. Consider the following examples taken from [78]:

a) A perfect number has the sum of its proper factors equal to the number itself. The first few known perfect numbers are 6, 28, 496, 8128, and 33, 550, 336 [78]. Question: are all perfect numbers even?

b) Twin primes are prime numbers that differ by 2, like 3 and 5 or 11 and 13. The question is: are there arbitrarily large twin primes? Does there exist an infinite number of twin primes?

c) Goldbach's conjecture [9] that says every even number except 2, is the sum of two primes. For example 4=2+2, 20 = 13 + 7, 30 = 19 + 11, etc. Question: is the conjecture true? This conjecture has been proved recently [23]. The uncertainty in the proof arising from the ambiguity of infinite set can be avoided in this theorem by re-stating it as follows: given any even number N except 2 there exist two primes whose sum is N. This theorem holds for any *given* even number N which is finite.

Consider the statement with the existential quantifier: The decimal expansion of a number p has no row of one hundred threes. True or not, it is unknown although extensive calculation on its decimal expansion has not yielded such row of threes. The probability that this statement is true is $1 - (9/10)^{100}$ which is almost 1. Thus, even a statement with probability that it is true is near 1 is not certain.

Among the field axioms that define the real number system **R** are (a) the trichotomy and (b) completeness axioms. The trichotomy axiom says that for any a, b $\in$ **R**, exactly one of the following is true, $a = b$, $a < b$ or $a > b$; the completeness axiom says: every non empty subset S of R that is bounded above (has an upper bound) has the least upper bound.

The completeness axiom is a variant of the axiom of choice one version of which says, essentially, that if a soft ball is suitably sliced into infinitely small little pieces, then the pieces can be suitably rearranged, without distortion, and reconstructed into a ball, the size of the Earth (Banach-Tarski paradox) [75]. This is a contradiction in R^3 inherited from the reals and attributed to the axiom of choice. Actually the axiom of choice is incidental here. The specific source of the problem aside from the ambiguity of infinite set is the Archimedean property of the reals that says: given any real number $\varepsilon > 0$ no matter how small and any number M, no matter how large, there exists some number N such that $N\varepsilon > M$. This allows one to form an arbitrarily large object from arbitrarily large number of arbitrarily small pieces [15, 75, 78].

A paradox is really a contradiction but others look at it as something unexpected or contrary to intuition.

A counterexample to any of the axioms of a mathematical system reveals inconsistency. Lately, L. E. J. Brouwer constructed a counterexample to the trichotomy axiom [5]. We present our version of the counterexample taken from [21] that shows at the same time that the real number system is not linearly ordered by "<" and that an irrational is not the limit of a sequence of rationals in the standard norm. (In fact, the notion *irrational* is ambiguous [21])

Let C be irrational. We want to isolate C in an interval such that all the decimals to the left of C are less than C and all decimals to the right of C are greater than C. We do this by constructing a sequence of smaller and smaller rational intervals (rational endpoints) such that each interval in the sequence is inside the preceding (this is called nested sequence of intervals). In the construction we skip the rationals that do not satisfy the above condition. Although given two distinct rationals x, y we can tell if $x < y$ or $x > y$, we cannot line them up on the real line under the relation "<"since if x, y are two rationals, $x < y$, there is an infinity of fractions or rationals between them and we cannot verify their arrangement. (Not all reduced fractions or rationals are well defined, e.g., when the denominator of a reduced fraction is a prime other than 2 or 5, since we do not know all the digits of the quotient; in other words, there are missing elements among the fractions, the reason for the difficulty in arranging the decimals linearly with respect to "<")

Therefore, we settle for this scenario: starting with the rational interval [A, B] we find a nested sequence of rational intervals that "insures" C lies between the two endpoints at each stage. We go for an arrangement that will allow us to distinguish the left from the right endpoints of the sequence. We construct rows of rationals starting with numerator 1 in the first row, 2 in the second row, etc., and the denominators in each case consisting of consecutive integers starting from 1 in increasing order going right so that in each case we start with a denominator of a potential right endpoint.

Actually, we can squeeze the rows into a single row since no particular order with respect to "<" is involved. Even this arrangement is a problem. For example, suppose at a certain stage in the construction we have a left endpoint 1/5 then the number 20/100 appears on the right and, in trying to pair the left endpoint 1/5 with a right endpoint we skip 20/100 and all other rationals to the right of 1/5 in the ordering "<" that appear on the right and move further left than all of them. We choose the left endpoint in the succeeding step similarly to the right of 1/5, etc. Without loss of generality, we take this rational 1/5 to be the first left endpoint in the construction. Then once we have found the right pair for 1/5 we either use it as the right endpoint of the next rational interval and pair it with some rational on the right of 1/5 or find a new left endpoint to the right of the first left endpoint to pair with a right endpoint left of 1/5, etc. We make sure that we do not get closer to C than 10^{-n} at the nth step in the choice of the first n endpoints so that C remains inside each interval. While we are sure for all left and right endpoints A, B we have already identified in our construction, that $A < C < B$ and all rationals right of A and left of B in the ordering "<" satisfy this inequality, there remains an interval of rational endpoints containing C and rationals that do not satisfy this inequality no matter how large we choose n. Therefore, the location of C remains unknown.

(In Brouwer's version of the counterexample, there is no limitation on how close the rational end points are to the irrational C. Therefore, by skipping the rationals that do not belong to the left or right endpoints, the right endpoints of the sequence in the construction eventually appear on the left and the left endpoints appear on the right, the irrational C nowhere to be found [5]. This was the first counterexample that showed the field axioms inconsistent and invalidated Wiles' proof of FLT [110, 111])

This construction attests to the ambiguity of the concept *irrational* and the problem of representing it as limit of a sequence of rationals; for every such sequence there is always a gap. Even the rationals in the real line are ambiguous mainly because there is an infinity of rationals between any given two rationals so that we cannot order them under the relation "<", i.e., we cannot line them up in the line interval between 0 and 1, denoted by [0, 1], under this relation. This is due to the ambiguity of infinity. Consequently, the real number system is not linearly ordered by this relation and the trichotomy axiom that says, given two real numbers x, y, one and only one of the following holds: $x < y$, $x = y$, $x > y$, is false. It shows further that the fractions are just as ambiguous as the nonterminating decimals the latter having a bit of advantage for being linearly ordered by the lexicographic ordering. We shall see that the new real numbers are linearly ordered by the lexicographic ordering "<" from which follows the trichotomy axiom.

We state the following two theorems; the proofs are standard and presented in Chapter 2.

Theorem

The rationals and irrationals are separated, i.e., they are not dense in their union (this is the first indication of discreteness of the decimals).

Theorem

The largest and smallest elements of the open interval (0, 1) are 0.99… and 1 – 0.99…, respectively [34].

QUALITATIVE MATHEMATICS AND MODELING

Some mathematical or scientific problems including proving or disproving Fermat's last theorem require more than just computation and measurement to resolve. For example, in an inverse problem in differential equations the boundary conditions are generally unknown or incomplete and what is known is their outcome. An example of an inverse problem is the gravitational n-body problem [41] posed by Simon Marquis de Laplace at the turn of the 17th Century that says:

Given n bodies in the cosmos at some initial time with their respective masses, positions and velocities and subject to their mutual gravitational attractions find their positions, velocities and paths at later time.

The bodies have history and boundary conditions way back in the past that gave rise to their present phase as a physical system. To solve this problem we must know what they were and the nature of the bodies and their interactions. Incidentally, it was this problem and the search for solution that gave birth to the grand unified theory [42].

Mathematical analysis such as the critique-rectification of foundations and the real number system cannot be done by computation or measurement alone. Therefore, we expand the tools of mathematics and science. Consider the following mental activity:

Making conclusions, visualizing, abstracting, thought experimenting, learning, doing creative activity, intuition, imagination and trial and error to sift out what is appropriate, negating what is known to gain insights into the unknown, altering premises to draw out new conclusions, thinking backwards, finding premises for a mathematical space and devising techniques that yield results.

This activity falls under rational thought or intelligence; we call its representation in the real world qualitative or non-quantitative mathematics. A lot of imagination as component of qualitative mathematics is needed to understand this book.

Since individual thought is not accessible to others and not all mental activity can be represented in the real world, there is inherent ambiguity in individual thought and, therefore, only its representation in the real world that can be studied and analyzed collectively is the proper subject matter of mathematics.

Qualitative mathematics introduced for the first time in and the main contribution of [46] includes abstract mathematical spaces, foundations, mathematical reasoning and, most important of all for our purposes, the search for the laws of nature. It is the main component of the new methodology of qualitative modeling that explains nature in terms of its laws applied to physics for the first time in [41] to solve the n-body problem. It provides the remedy for the inadequacy of computation and measurements that has left longstanding problems unsolved and fundamental questions unanswered. It alters the task of the scientist from computation and measurement that describe natural phenomena to the search for the laws of nature upon which physical theory is built to explain natural phenomena and solve scientific problems. Moreover, it raises the quality of scientific knowledge from description of the appearances of nature to a deeper understanding of how nature works by explaining its internal dynamics and the forces and interactions of physical systems in terms of natural laws.

THE PLACE OF MATHEMATICS IN SCIENCE

Like any language, the subject matter of mathematics is not nature but itself; therefore, it is not a science whose subject matter is nature. It is a specialized language that suits the needs of science. It articulates a physical theory so that the latter can explain natural phenomena, provide solution to scientific problem, predict the future course of a dynamical system and serve as guide in designing technology and scientific experiment. Mathematics is the medium of thought for studying our universe including our physical self that is external to our thought. We can also use it to study the representation of thought just as we study the structure or usage of a language. Physical theory is the form by which we express or articulate knowledge of nature.

The traditional role of mathematics has been computation for purposes of describing natural phenomena. Even Einstein's vision of uniting gravity and electromagnetism was aimed at describing both natural phenomena by similar equations, i.e., the same forms except for some constants of nature. He succeeded in doing so for gravity and electromagnetism using the field equations of relativity and Maxwell's equations of electromagnetism, respectively, but only in a very weak sense being sought at the time: Maxwell's and the field equations have similar forms. Then quantum physicists joined in to try to extend the unification in the same sense to the weak and strong forces of physics. It did not prosper, however, as they got stuck in the search for the basic constituent of matter which is central to the description of quantum physics.

Qualitative mathematics came to physics in a dramatic way: it was instrumental in the discovery of the 11 natural laws for the solution of the gravitational n-body problem in 1997 and the discovery of the basic constituent of matter required for it [41]. They anchored the initial formulation of GUT called flux theory of gravitation [33].

As the main component of qualitative modeling qualitative mathematics extended the theoretical applications of GUT to the broad fields of natural science and their applications as far as engineering, medicine [30, 32], physical psychology [28, 29], mathematics-science education [29], theory of evolution [30, 31], geological and atmospheric sciences and oceanography [35, 38] and design of technology along with the discovery of relevant laws of nature. For example, at least two laws of nature were discovered to explain why the final flight of the Columbia Space Shuttle ended in disaster killing all seven flight crew members aboard [40]. That catastrophic failure in technology has not been explained by conventional science and the same problem of breach of the insulation panel recurred during the resumption of the program, the reason for its termination.

Qualitative modeling found applications in another new field – complex systems, i.e., physical or social systems or problems that cannot be analyzed or solved by computation and measurement alone [6, 47]. They include generalized fractal such as the configuration of the superstring and the problem of economic-industrial development of underdeveloped and developing countries [6, 47].

PHILOSOPHICAL FOUNDATIONS OF THE NEW PHYSICS

The new physics elaborated in [78] is hybrid between the grand unified theory (GUT) and its mathematics [25, 42, 78]. Its main content, however, is GUT the foundations of which are the solution of unsolved problems and answers to the fundamental questions of physics [78]. We introduce them here and discuss them later. The unsolved problems include the long-standing gravitational n-body and turbulence problems and the more recent one posed by Einstein in the 1920s, the unification of gravity and electromagnetism and the weak and strong forces of physics in terms of similar mathematical description of their appearances. Our solution of the last problem is much broader: the unification of the forces and interactions of nature by a single physical theory – GUT [42]. The unanswered questions are listed in [78] but we give priority here to what the basic constituent of matter is.

THE UNSOLVED PROBLEMS AND UNANSWERED QUESTIONS OF PHYSICS

A problem is famous because it is simply and clearly stated but defies solution for a long time. Consider the famous gravitational n-body problem posed by Simon Marquis de Laplace at the turn of the 18th Century in his book *Celestial Mechanics* that defied solution for two centuries. It says: given n bodies in the Cosmos at a specific initial time with known masses, positions and velocities subject to their mutual gravitational attraction find their positions, velocities and paths at later time. (It was assumed that masses of cosmological bodies do not change but they do [25, 83]) The problem seems adequately and clearly stated for we think we know what bodies and gravity are which is not true. We do not know how bodies behave; nor do we know the full impact of gravity on them. In other words, the problem is unsolvable as it is. What would it take to solve it? We have to know what a body consists of which leads to the 5, 000-year search for the basic constituent of matter. Then we need to know what gravity is. Newton's so-called law that says the gravitational force of attraction between two bodies of masses m_1 and m_2 s cm apart is given by $F = Gm_1m_2/s^2$, where G is a known cosmological constant [12], is a description of motion of two masses subject to their mutual gravitational attraction. It appears that the solution is a matter of computation but it is not. The questions of what matter consists of and what gravity is cannot be answered by describing the appearances of bodies under the influence of gravity.

The other long-standing problem of physics, so long no one knows who posed it, was the turbulence problem [35]. The problem is to find out what gives rise to it, e.g., typhoon and tornado. Of course, it is not as clearly stated as the n-body problem. One does not even know where to begin the reason the problem has defied solution for so long. Qualitative mathematics provided the key to its solution [35, 78].

In either case, the solution required qualitative modeling that led to the discovery of the basic constituent of matter and the initial 11 laws of nature of GUT [41]

There are unanswered fundamental questions in physics the most ancient one being what the basic constituent of matter is the answer to which is the key to understanding of basic physical concepts such as gravity, black hole and the structure of the electron and atom.

THE SEARCH FOR THE SUPERSTRING

It was the development of quantum physics in the1920s and the splitting of the atom in the 1940s that rekindled the search for the basic constituent of matter in the 1950s that was in limbo for over 5,000 years. During that decade particle physicists embarked on the search for the basic irreducible elementary particles (prima) by smashing the nucleus of the atom to determine its constituents (dark matter was unheard of then). They started with the cyclotron, then the bevatron, linear accelerator, hadron collider and now the large hadron collider or CERN, a circular accelerator 7 km across that straddles between the boundaries of France and Switzerland [12] – all aimed at smashing the nucleus of the atom by an energized proton to find out what is there, specifically, to find the *true* or irreducible elementary particles, i.e., the basic irreducible elementary particles that comprise every atom. By the 1990s the search for the basic irreducible elementary particles was a complete success with the discovery of the +quark (up quark) and −quark (down quark) and the electron by J. J. Thompson in 1897 [12, 43]. They are basic because they comprise the light isotope of every atom; a heavy isotope has, in addition, the neutrino in the neutron [43]. We shall discuss them in detail later. Whatever particle physicists have achieved beyond this point is a bonus for natural science, a bonus for mankind.

The basic prima are produced at staggering rate in the inner core of cosmological vortices, in the vacuum of space and in the cellular membrane of living things.

CHAPTER 2

The Mathematics of the Grand Unified Theory

Abstract: This chapter surveys conventional and new mathematics involved in the development of GUT. The most important conventional mathematics includes the theories of generalized curves and surfaces and the integrated version of Pontrjagin maximum principle developed by L. C. Young. We summarize their original development here because they are quantitative models of many important physical concepts. The boundary year for the new mathematics is 1998, the year of publication of the counterexamples to Fermat's last theorem that proves the false conjecture and catalyzed the development of new mathematics such as (a) the new real number system, (b) generalized integral, derivative and fractal and (c) the complex vector plane (fully developed here except (c)). The introduction to the complex vector plane that rectifies the complex number system is presented but its full development requires re-writing of the rectified complex number system as the basis for rectification of complex analysis. The introduction of qualitative mathematics paved the way for qualitative modeling, the crucial factor for the discovery of the superstring and the 11 initial laws of nature required for the solution of the gravitational n-body problem by serving as foundations of GUT. The d-sequence of a dark number qualitatively models the superstring, the generalized curve quantitatively models its path and that of an elementary particle and the new real number system quantitatively models time and distance of ordinary space. The generalized surface quantitatively models the expanding Cosmic Sphere before its burst at t = 1.5 billion years from the start of the Big Bang.

INTRODUCTION

We survey the mathematics directly involved in GUT's development but provide details on the major ones: the new real number system and the generalized curves, integral, derivatives and fractal. We consider conventional mathematics of GUT developed before 1997 and call the rest new mathematics.

CONVENTIONAL MATHEMATICS

The most important conventional mathematics involved in the development of GUT are the theories of generalized curves and surfaces and the integrated Pontrjagin maximum principle the last one developed by L. C. Young in 1969 which he adapted to generalized curves of optimal control theory called relaxed trajectories [117]. It is an improvement over the original Pontrjagin maximum principle developed by Pontrjagin and his associates in 1962 [91].

Special Functions and Generalized Curves

We consider mischievous functions that render certain established methods ineffective. However, once tamed they are useful. Examples of mischievous functions are the infinitesimal zigzag and the wild oscillation,

$$\mathrm{Sin}^m 1/x,\ (\sin^n 1/x)(\cos x^m 1/x), \qquad \textbf{(1)}$$

where m and n are integers.

The Infinitesimal Zigzag

We generate a sequence of functions (polygonal lines),

$$C_n\!: y_n = y_n(x),\ 0 \le x \le 1;\ n = 1, 2, \ldots, \qquad \textbf{(2)}$$

that converges to curve C_0 point-wise (or in the sup norm) as follows:

Without loss of generality, take C_1: $y_1 = y_1(x)$ $0 \le x \le 1$, the polygonal line joining A and B formed by sides AD and DB of triangle ABD with vertices at A(0, 0), B(1, 0) and D(1/2, 1/2) (Fig. **1**). Note that their slopes are +1 and –1, respectively. For the second term in the sequence, join the midpoint P of AD to the midpoint Q of AB and point Q to the midpoint R of DB to form the polygonal line APQRB from A to B. We denote this function by C_2: $y_2 = y_2(x)$, $0 \le x \le 1$. We continue similar construction on the polygonal APQRB. From the geometry of the figure, the slope of

E. E. Escultura

C_n, n = 1, 2, …, at any point in [0, 1] is +1 or –1 except at corner points where it is simultaneously +1 and –1 (derivative does not exist in this set of measure 0 being countable). Also, the length $|C_2|$ of $C_2 = |C_1|$ = length of $C_1 = \sqrt{2}$. Continuing similar construction on the finer polygonal lines we obtain a sequence of polygonal lines C_n: $y_n = y_n(x)$, n = 1, 2,..., $0 \le x \le 1$, having the following properties:

(a) $$|C_n| = \int_\Delta ((1 + (y')^2)^{1/2}dx + \int_\Sigma ((1 + (y')^2))^{1/2}dx = \sqrt{2} = |C_1| \quad \textbf{(3)}$$

where y_n' is the derivative of y_n; $\Delta = \{x \mid y_n'(x) = +1\}$ and $\sum = \{x \mid y_n'(x) = -1\}$ (the set at which $y_n' = \pm 1$ has measure 0).

(b) The sequence C_n: n = 1, 2, …, is uniformly convergent point-wise and since each C_n is continuous the limit, C_0: $y_0 = 0$, $0 \le x \le 1$, is continuous. In fact, y_0 coincides with y(x) = 0, $x \in [0, 1]$, which is absolutely continuous. Hence, y_0 is also absolutely continuous.

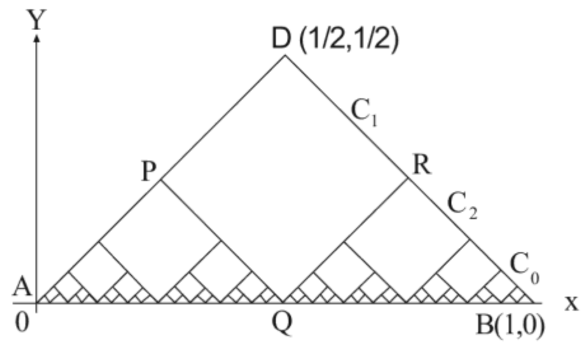


Figure 1: The first two terms of the sequence of polygonal lines that tends to the infinitesimal zigzag in the interval [0, 1].

(c) What about the derivative of y_0? Does it exist? If it does, what is it? We cannot have $y_0' = 0$. For, if that were so, it would violate the dominated convergence theorem applied to the integrals in (a) since this would imply,

$$\int_{[0,1]}((1 + (y_0')^2))^{1/2}dx = 1 \neq \sqrt{2} = \lim |C_n| = \lim \int_{[0,1]} ((1 + (y_n')^2))^{1/2}dx, \quad \textbf{(4)}$$

as $n \to \infty$. Derivative of y_0 does not exist in the ordinary sense because the sequence y_n', n + 1, 2, …, or –1, +1, –1 +1, …, does not converge to a single point since its set limit is {–1, +1}, *i.e.*, y_0' is set-valued.

The function C_0:(y_0, y_0'): $y_0 = 0$, $y_0' = \pm 1$, $0 \le x \le 1$ is a counterexample to a theorem in [93] that says, an absolutely continuous function is differentiable, almost everywhere. C_0 is absolutely continuous but nowhere differentiable. It raises some important points which are the source of this particular contradiction:

(a) Inadequacy of the present notion of function; this was pointed out in [112, 113] 70 years ago and, again, more recently in [117]. A function defined by its values alone cannot distinguish the function C: y = 0 from C_0:$y_0 = 0$ which are distinct in at least two ways: one is differentiable and the other is not and they also have different lengths.

(b) Inadequacy of the notion *derivative*; that the derivative of a function cannot be adequately expressed by its values because derivative is a property belonging to an extension of its underlying space (extension of n-space in the general case) whose restriction to the space of real-valued functions contradicts some of its properties (e.g., properties of absolutely continuous function). Therefore, there is a need to extend the notion of function to include those with set-valued derivative. Also, the present defect in the notion of limit is passed on to other notions defined by limits including the derivative [117].

The function C_0:$y_0 = 0$, $y_0' = \pm 1$, $0 \le x \le 1$, belongs to a wider class of curves called generalized curves different from the ordinary curve C: y = 0, $0 \le x \le 1$. Yet their values coincide point-wise. Furthermore, their arc lengths differ; in fact, there are countably infinite functions of this kind. One can see that although the sequence of functions C_n, n = 1, 2, …, converges to the segment AB point-wise, its standard limit, say, in the sup norm, is something else: the infinitesimal zigzag, C_0:$y_0 = 0$, $y_0' = \pm 1$. This example raises two very important points:

(1) Fallacious proof of existence of a mathematical object by approximation or convergence as well as the erroneous use of numerical and algorithmic methods without existence theory (in fact, this flaw is a variant of vacuous statement).

(2) The inadequacy of the values of a function in characterizing its derivative; thus, the present notion of derivative is inadequate to capture the complexity of the property of a function.

This inadequacy of the notion of function and derivative as well as numerical method without rigorous justification has far-reaching significance for all of analysis and beyond. In particular, any theorem on derivative inherits this problem. That is drawn out by a property of a mischievous function that we shall deal with later. The infinitesimal zigzag is our first example of a mischievous function. It serves as counterexample to a number of well-established theorems.

Significance of the Infinitesimal Zigzag

This example says that the sup norm and the metric induced by point-wise convergence are not the natural metric for purposes of optimization, especially, in the calculus of variations; moreover, the properties of a curve are not fully accounted for by its values or parametric representation. We must put into account the behavior of its derivative and, as remedy, if f(t), $t \in [0, 1]$, is the parametric representation of a curve C we represent it by the pair C: (f, g) where g is the derivative of f. Then the natural metric for purposes of optimization is the Young measure which we shall present later or curvilinear integral of some objective function along it (which can be the cost function) [117]. If we represent that measure by the integral,

$$I(C) = \int_{[0, 1]}(f(t), g(t))dt, \quad \textbf{(5)}$$

then I(C) is the Young measure of the curve C. When the integrand is 1, I(C) is called the length of the curve. Thus, a curve is a linear functional and curves of the same Young measure belong to the same equivalence class representing that linear functional. This makes functional analysis available to optimal control theory. In an optimal control problem the derivative g is the control parameter so that it is independent of f; in other words, the system is controlled by finite set of values of the derivative.

Another case of optimization where the "obvious" curve is not the optimal solution is this example: find the minimum of the integral,

$$\int_{[0, 1]}((1 + x^2)(1 + ((x'^2 - 1)^2)^{100})dt \quad \textbf{(6)}$$

(where x' is derivative) from 0 to 1 among admissible functions x(t) subject to x(0) = x(1) = 0. The "obvious' optimal curve among conventional curves is x = 0, subject to x(0) = x(1) = 0 and the minimum is 2^{100}. However, by admitting infinitesimal zigzag, which is like the ordinary curve x = 0 but whose derivative is set-valued and concurrently takes the values +1 and −1, and attaching a probability weight 1/2 to each of these values, we obtain a minimum of 1. Thus, the conventional theory of curves yields incorrect solution of this variational problem.

Here, the infinitesimal zigzag is a generalized curve or, to be precise, this generalized curve is the equivalence class of curves of the same Young measure (with set-valued derivative) [117]. Incidentally, all four types of cosmic waves and the superstring are generalized curves because they have one thing in common: set-valued derivatives; so is the path of a primum (elementary particle) [119].

The Wild Oscillation Sin1/x

Our next mischievous function is the wild oscillation, F(x) = sin1/x. This is a special case of the more general mischievous function $\sin^m 1/x^k$, where k, m, are positive integers. It reveals a flaw in the Lebesgue theorem on the Riemann integral that says:

A bounded function is Riemann integrable if and only if its set of discontinuity has measure zero [93]. The bounded function F(x) = sin1/x whose only discontinuity is at x = 0 is not Riemann integrable in any neighborhood of the origin. Known proof of integrability of sin1/x involves construction of a Riemann integral outside an ε-neighborhood of x = 0, where ε > 0, which exists, and taking a sequence of such integral as ε → 0, which converges.

The limit of such a sequence, however, is not necessarily Riemann integrable, certainly, not sin1/x because no Riemann sum of this function can be formed in any neighborhood of 0. The best we can say here is that one can construct a convergent sequence of Riemann integrals with some relation to the function sin1/x in the same ε-neighborhood of x = 0 but its limit is something else. This is, in fact, a form of the Perron paradox on the use of necessary conditions without an existence theory [117, 118]. It also illustrates the same fallacy mentioned earlier in the proof of existence of some mathematical object by approximation or convergence as well as the use of an algorithmic solution of a problem the existence of solution of which is not establish.

In the development of the Henstock integral in [80] the function $\sin 1/x^2$ plays a central role. However, the theory is flawed by the inadequate notion *derivative*. While this function is shrunk to zero by the factor x^2 its derivative is not for it belongs to a higher space independent of the function. The function considered in [80] is $x^2\sin 1/x^2$, $0 \leq x \leq 1$. It is asserted that its derivative F'(x) exists at x = 0 and F'(x) = 0 because at that point its one-sided derivative can be trivially computed since, using the ordinary definition of derivative, we have,

$$|\Delta F|/|\Delta x| \leq |x^2|/|x| = |x|, \tag{7}$$

so that lim $|\Delta F|/|\Delta x|$, as $x \to 0^+$, exists.

The inequality follows from the fact that F(x) is bounded by its envelope $y = \pm x^2$. F(x) is continuously differentiable outside x = 0. In fact, we have, at $x \neq 0$,

$$F'(x) = 2x\sin 1/x^2 - (2/x)\cos 1/x^2, \tag{8}$$

and its graph is shown in Fig. **2**. where the term $2x\sin 1/x^2$ has been discarded since it vanishes as $x \to 0^+$. However, the term $(2/x)\cos 1/x^2$ oscillates rapidly along all values in the interval $(-\infty, \infty)$ as $x \to 0^+$ and does not converge. This is a particular kind of discontinuity, an example of what we shall call chaos. Moreover, this is another example of a derivative of a function that is independent of it.

Our final mischievous function is a function of the type,

$$(e^{1/z}/x^k)(\sin^m 1/x^2 + \cos^n 1/x^2) \text{ or } (e^{1/z}\, x^k)(\sin^n 1/x^2), \tag{9}$$

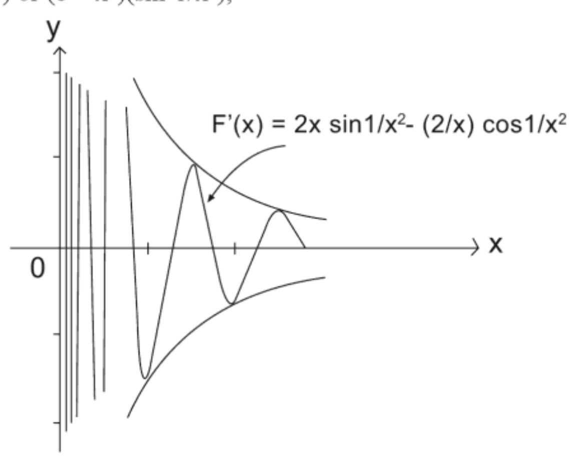


Figure 2: Graph of $F'(x) = -(2/x)\cos 1/x^2$ where the term $2x\sin 1/x^2$ is discarded since it tends to 0 with x; it takes all values in $(-\infty, +\infty)$ as $x \to 0$.

where $z = x^2$, k, m, n are positive integers. Finding the limits of these functions, as $x \to 0$, quickly reveals that L'Hospital's rule breaks down on (9). The reason: these functions do not satisfy its hypothesis at the origin, that the function should not have a zero in any neighborhood; each of the functions in (9) has countably infinite zeros in any neighborhood of the origin. Also by rearranging the factors one gets different standard limits. The generalized derivatives of $(e^{1/z}/x^k)\sin^m 1/x^2$ and $(e^{1/z}/x^k)(\sin^m 1/x^2)$ or their expectations are evaluated in [2] and used to generalize L'Hospital's rule; the latter is applied to functions (9) to evaluate their limits as $x \to 0$ [2].

Rapid Spiral and Oscillation

A primum (unit of visible matter) is mathematically modelled by the rapid spiral $x = t$, $r(t) = \beta(\sin n\pi t)(\cos^m k\pi t)$, $t \in [-1/k, 1/k]$, $\theta = nt$, n, m, k, integers, $n >> k$, m even, whose profile is a sinusoidal curve of even power [25]. Its cycle energy is Planck's constant $h = 6.64\times10^{-34}$joules [12], the irreducible unit of energy that we shall discuss later. Energy conservation and flux compatibility pull the primal cycles together to form a set-valued function that requires the generalized integral [19] to do calculation on it because the ambiguity or uncertainty of large number induces uncertainty on such large number of primal cycles.

(Note: a function multiplied by an oscillation is an oscillation)

Rectification of Inadequacy *Function* and *Derivative*

Partial rectification of the inadequacy of the concepts *function* and *derivative* is done in [112, 113] by representing a vector function f with values in **R** by a pair (f, g), where g is the derivative of f that takes values in a field of vectors belonging to a separate space isometric to $\mathbf{R}^n$. This representation formalizes the independence of the derivative from function. (For full development of this idea and the requirements on f and g see [112, 113]; we provide a summary later). Further rectification is required by the reconstruction of the real numbers into the consistent new real number system.

Regarding the derivative component of the pair (f, g) Young went so far as admitting set-valued derivative with the introduction of chattering controls [117]; in the study of convex vector functions the notion of a set of landing hyperplanes at a point is admitted. This corresponds to a set of tangent lines for a real-valued function [117]).

The thrust of rectification focuses on the derivative component of the pair (f, g) in the representation of a function. It is this approach that led to the introduction of generalized curves which, in turn, established an existence theory and resolved the Perron paradox [117] in the calculus of variations, paving the way for the latter's modern formulation in optimal control theory. However, with the appearance of set-valued functions in the study of fractal and chaos this rectification effort still falls short; we need to allow set-valued function in the pair (f, g) as a way to capture the complexity of certain notions, particularly, *function* and *derivative*.

Note that the approach in [113] reflects the methodology of enrichment: enriching the space with new elements to achieve an existence theory or convergence.

Let $\{f_n(t), g_n(t)\}$, n = 1, 2. …, $t \in [a, b]$ be a sequence of functions in the new sense, where each $f_n(t)$ is continuous and $g_n(t)$ measurable and well-defined almost everywhere. We suppose further that the end points $[a_n, b_n]$ of the domains of definition of $f_n(t)$ and $g_n(t)$ tend, respectively, to 0 and 1 as $n \to \infty$. For our purposes here we require that $a_n \geq 0$, and each of $f_n(t)$ and $g_n(t)$ has common extension to some interval T containing both [0, 1] and the sequence of intervals $[a_n, b_n]$, n = 1, 2, …. We define the limit set of $\{f_n(t), g_n(t)\}$as the pair $(\{f_{0,0}(t)\}, \{g_{0,0}(t)\})$, where $\{f_{0,0}(t)\}$ = Slim$\{f_{n,k}(t)\}$ = the set of limit points of a diagonal element $\{f_n(t_n)\}$, as $n \to \infty$ and $t_k \to t$ and $\{g_{0,0}(t)\}$ = Slim$\{g_{0,0}(t)\}$= the set of limit points of a diagonal element $\{g_n(t_k)\}$, as $n \to \infty$ and $t_k \to t$. Since f_n is continuous its limit is independent of the sequence $\{t_k\}$; not so with g_n since we only require measurability. Thus, we distinguish the limit set of a sequence of pairs $\{f_n(t), g_n(t)\}$ by the particular sequence $\{t_k\}$, where $t_k \to t$. This is consistent with our observation above that there is infinity of coincident but distinct curves on the segment AB, each element being determined by the particular curve that tends towards it. (As we shall see later those curves are countably infinite) As special case, let $t_k \to t$ for all t in T. Then the closure of such sequence $\{f_n(t), g_n(t)\}$ under this convergence is called the space of generalized curves. The complete formulation of the theory of generalized curves as linear functionals is given in [113]; we summarize it below. Essentially, a generalized curve is one with set-valued [113].

Applications of the Infinitesimal Zigzag

We make references to the superstring although we shall take it up later since we want to introduce its mathematical models. A superstring is a nested fractal sequence of superstrings where the first term is a close helix; it has a flux called toroidal flux in its helical cycle which is a superstring traveling at 7×10^{22} cm/sec or 10^{12} times the speed of light [4]; the toroidal flux has a toroidal flux in its helical cycle which is a superstring traveling at the same speed, etc. The projection of a helix on the plane through its axis is sinusoidal or oscillatory curve, by the energy conservation equivalence [25].

Given any curve in the plane we can deform it into an oscillatory curve y = $\sin^m bx$ which is rectifiable; we can further deform it into some isosceles triangle ADB so that its length is preserved and equal to the sum of the lengths of AD and DB. In turn, we can deform this triangle into a finer oscillatory curve K_1, with length preserved (Fig. **3**). We iterate this deformation forming an alternate sequence of polygonal lines and oscillatory curves K_n from A to B. Again, the sequence K_n tends towards a generalized curve called infinitesimal oscillation whose function component coincides with the zero function C: $y= 0$, $0 \leq x \leq 1$. Its length is equal to the original length $|K|$ of K and its derivative at any point $x \in [0, 1]$ is set-valued and equals the set of limit points of the derivatives of the sequence of oscillations at x.

Since the segment AB is arbitrary we can prescribe its length to be an arbitrary number $\varepsilon > 0$. Then we have the following:

Theorem 1

Given an oscillatory curve K, any number $\varepsilon > 0$ and a line segment AB, there exists a continuous deformation of K into a fine oscillatory curve inside an ε-neighborhood of AB that preserves the length of K [24, 34].

Theorem 2

Given an oscillatory curve K, there exists a continuous deformation of K, with length preserved, into an arbitrarily small neighborhood of a point [24, 34].

Proof

We prove both theorems. Let A be a given point and B a point in the ε-neighborhood of A and suppose $|AB| = \varepsilon/2 > 0$. There exists a deformation of K, with length preserved, into two sides of an isosceles triangle ADB where $|AD| + |DB| = |K|$. Following the construction above there exists a sequence of polygonal curves C_n and corresponding oscillatory curves K_n such that for each n, $K_n = C_n = AD + DB = K$ and K_n tends to the segment AB (Fig. **3**). Hence there exists a positive integer N such that whenever $n \geq N$, the curve K_n lies inside the ε-neighborhood of A. (This establishes the first theorem) since the length of AB is arbitrary, $\varepsilon > 0$ and $\neg AB\neg = \varepsilon/2$. Then the second theorem follows from the first).

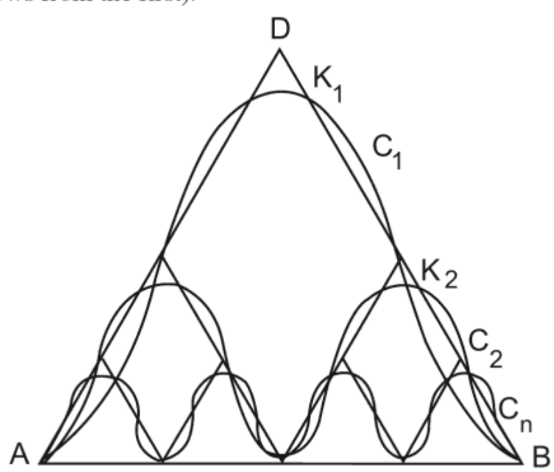


Figure 3: Sequence of sinusoidal curves of length $|K|$ that tends towards infinitesimal oscillation of the same length.

Note that in each case the oscillatory structure is preserved as well as its length. Thus, it is possible to shrink an oscillatory curve of any length into an infinitesimal oscillation at a point. Now, let $\beta > 0$, where β is small, and let K be an oscillatory curve of large length K. Let $\varepsilon = \beta/2 < |K|/2$. As before, we deform K into the two sides of an isosceles triangle ADB with base AB, where ┐AB┐ = ε. Let h be the altitude of this triangle, then for suitably small ε, h ≈ |K|/2. By the

Archimedean property of the decimals there exists some positive integer n such that

$$|K|/2^{n+2} < |K|/2^{n+1} \leq |K|/2^n. \quad \textbf{(10)}$$

Therefore, in the sequence of oscillatory curves K_i with $|K_i| = |K|$, for each i = 1, 2,..., which tends towards the line segment AB, there is one whose amplitude satisfies the inequality (10). We state this as a theorem.

Theorem 3

Let k be an oscillatory curve with large length |K| and let ε > 0, ε = β/2 < |K|/2. Then one can continuously deform the oscillatory curve K into an arbitrarily small neighborhood of a point with its length and amplitude prescribed to satisfy,

$$|K|/2^{n+2} < |K|/2^{n+1} \leq \varepsilon \leq |K|/2^n, \quad \textbf{(11)}$$

for some integer n [24, 34].

The following theorem [24] is now obvious and follows from the above theorems:

Theorem 4

The real line is chaos.

Theorems 1 – 4 model different aspects of the shrinking of a superstring. They have other implications for physics that can explain certain phenomena such as the tremendous but undetected (latent) energy in the nucleus of an atom. Tremendous because we can pack infinitesimal helical loops into an arbitrarily small neighborhood of a point at very high energy level hζ where h is Planck's constant and ζ is number of helical cycles. They cannot be detected by our means of observation such as light since if the wavelength of the latter is sufficiently fine there would be no interference or discordant resonance with this infinitesimal helix due to difference in orders of magnitude of its cycles (the helix can be semi- or non-agitated superstring [25]). Helix and oscillation (sinusoidal) are universal configurations of matter and they are related: the projection of a helix on a plane through its axis is sinusoidal. Infinitesimal helix and oscillation are both generalized curves because their derivatives are set-valued. The superstring, basic constituent of matter, is an infinitesimal helical loop, a generalized curve [25, 113].

The Generalized Curves

We summarize the development of the generalized curves [112, 113, 117] as an offshoot of the search for existence theory in the Calculus of Variations to resolve Perron paradox [117, 118] coming from the use of necessary condition without existence theory. We reconstruct its basic formulation to get a sense of the method of enrichment.

Consider the parametric equation x(t), $t_1 \geq t \geq t_2$, of the curve C, where x(t) is absolutely continuous and takes values in $\mathbf{R}^n$. At each t we take a vector y(t) taken from a vector field in a separate space isometric to $\mathbf{R}^n$ which we also denote by $\mathbf{R}^n$. We require the vector y(t) to lie in a unit sphere **S**, *i.e.*, |y(t)| = 1. We also require C to lie in a compact cube in $\mathbf{R}^n$. and denote by **A** the compact Cartesian product of these two sets. Our Lagrangian belongs to the space $C_0(A)$ of continuous functions with compact support **A**. We further assume that our Lagrangian is homogeneous in y so that it is determined in $\mathbf{R}^n$ by its restriction to **A** since if L is the Langrangian and (x, y) is any point in $\mathbf{R}^n \times \mathbf{R}^n$ then there exists some scalar α ≥ 0 such that y = αŷ, where ŷ ∈ **S**. From the homogeneity of L in y,

$$L(x,y) = L(x,\alpha\hat{y}) = \alpha f(x,\dot{y}), \quad \textbf{(12)}$$

where f ∈ $C_0(A)$ is the restriction of L to **A**.

We make a minor adjustment for simplicity of notation by representing C(x(t), y(t)), a ≥ t ≥ b, as a curve defined on the compact set **A**. Here we have attached a derivative y(t) to stress the independence of derivative from x(t). We also assume the parameter t to be an arc length from the initial point of C to avoid dependence of the curvilinear integral along C on its parametrization but more on intrinsic properties.

We consider two curves C_1 $(x_1(t), y_1(t))$, $a \geq t \geq b$ and C_2 $(x_2(t), y_2(t))$, $c \geq t \geq d$, equivalent if their curvilinear integrals satisfy,

$$I(C_1) = \int_{[a,b]} f(x_1(t),y_1(t))dt = \int_{[c,d]} f(x_2(t),y_2(t))dt = I(C_2), \tag{13}$$

for all $f \in C_0(\mathbf{A})$. Thus, a curve C_0 is completely determined by the values of its curvilinear integrals all $f \in C_0(\mathbf{A})$, *i.e.*, a linear functional g in the dual space $C^*(\mathbf{A})$ defined on $C_0(\mathbf{A})$.

We define fine convergence of a sequence g_n, n = 1, 2,..., of elements of $C^*(\mathbf{A})$ and say that g_n converges to g_0 or $g_n \to g_0$ if $g_nf \to g_0f$ for all $f \in C_0(\mathbf{A})$. The closure of this space of curves in the new sense is called the space of generalized curves. In this sense a generalized curve is the fine limit of ordinary curve.

A generalized curve is also called a generalized flow, where the latter is an element of the positive cone of $C^*(\mathbf{A})$, *i.e.*, $gf \geq 0$ for all $f \in C(\mathbf{A})$ with $f \geq 0$. It is the space of generalized curves that provides an existence theory in the calculus of variations and optimal control theory and renders Perron paradox inoperative there, showing again how a contradiction is resolved by some sort of enrichment, embedding or completion. This fine convergence induces a metric on $\mathbf{R}^n$ called Young metric. Thus, the distance between two curves C_1 and C_2 is given by,

$$I((C_1) - I(C_2)). \tag{14}$$

This norm is really the sup norm in the space of generalized curves defined as,

$$\sup | g_1f - g_2f | \text{ for all } f \in C_0(\mathbf{A}). \tag{15}$$

We require these linear functionals g_1, g_2 to be well-defined in the intervals of definitions of their respective duals (ordinary curves) in **A** and have common endpoints, the same requirement for fine convergence, that the dual sequence of g_n in **A** must have endpoints tending towards those of its limit g_0. The length of a generalized curve g is the value of the integral (15) when f = 1. This norm is the right one consistent with the counterexample presented earlier, the polygonal line that converges to an infinitesimal zigzag. In that case the function f is the arc length $((1 + (y_0'(t))^2)^{1/2}$ which is constant and equal to $\sqrt{2}$ for each C_n. Thus, $g_nf = \int_{[0,1]}((1 + (y_0'(t))^2)^{1/2}dt = \sqrt{2}$. Of course, the derivative of f, $y_0'(t)$, is set-valued with value $\{1, -1\}$, and its dual in $C^*(\mathbf{A})$, where $\mathbf{A}$ $[0, 1] \times [0, -1]$, is the infinitesimal zigzag.

The development of the generalized curves started 70 years ago but the study of fractal bifurcation and chaos can benefit much from it, not to mention its many applications. In fact, a generalized curve such as the infinitesimal zigzag is both chaos and limit set of fractal but it needs to be rediscovered because contemporary studies on fractal have recent origin. In the case of bifurcation (or more generally, multifurcation), which is the transition from chaos to fractal, it can be explained by the fact that even well-behaved functions y = f(x) passing through some point (x_0, y_0) is only one of the countable infinity of local solutions of some differential equation near (x_0, y_0) satisfying an initial condition on its derivative there. Put another way, near the point (x_0, y_0), there is a countable infinity of functions (f(x), g(x)), which are local solutions of some set-valued differential equation. For each choice of g in a set-valued differential equation of the form $x' \in \{g(t, x(t)\}$ we have, for a given probability distribution, a corresponding branch of the solution. This is how multifurcation occurs at every point on the initial function as well as on each branch. This is how chaos ultimately results. This is similar to the formulation of the development of generalized surfaces that we used for dealing with the undecidable proposition FLT. Later, it was extended to relaxed trajectories of control theory [117].

(For extension of this methodology to the development of generalized surfaces see [50, 114 – 116])

To summarize, the generalized curve quantitatively models the superstring, primum and its path in flight [119] and spiral path of visible matter falling into the core and spinning around the eye of a cosmological vortex which are continuous arcs with set-valued derivatives [20, 78]. What we see in a spiral nebula are its stars and minor vortices falling into its core and spinning around the eye. Rapid spiral quantitatively models a simple primum and rapid oscillation the photon and primum in flight [25]. **R*** quantitatively models physical time and distance, non-standard g-sequence of d* the nested fractal superstring and d* the tail end of its toroidal fluxes, a superstring and continuum. The decimals model the metric system and the integers the countably infinite and discrete dark and visible matter; and, as we shall see later, GUT qualitatively models our universe. Note that a physical system may have more than one mathematical model but only one qualitative model within equivalence.

The Generalized Surfaces

We summarize Young's theory of generalized surfaces and his joint work with W. H. Fleming [50, 114 – 116]. We shall focus on the more developed version in [116] for $\mathbf{R}^m$, $m \geq 3$.

Let **X** be the Euclidean m-space and let the parameters (u, v) range in the unit square $R(0 \leq u \leq 1, 0 \leq v \leq 1)$. We say that x(u, v) is a generalized Dirichlet representation of if it is defined and bounded in R and takes values in **X**, absolutely continuous on the intersection of R with almost every line u = constant or v = constant, its extension to the perimeter of R is continuous and its dirichlet integral,

$$L(x,R) = (1/2) \int_{R} ((x_u)^2 + (x_v)^2)dudv, \quad \textbf{(16)}$$

is finite.

We say that $x \in \mathbf{X}$ if x is an element of m-space, $x \in D$ and x is a generalized Dirichlet representation x(u, v). We also write $x \in D(N)$ to mean $x \in D$ and $L(x, R) < N$ for some positive integer N. The jacobian j(u, v) of x(u, v) is the vector product of the partial derivative x_u and x_v, *i.e.*, their normal, and exists almost everywhere in R. we write $(x, j) \in D$ to mean $x \in D$ and j is the jacobian of x. The values of any j(u, v) are bivectors. They lie in the space **J** whose elements can be identified with the skew-symmetric matrices $j = (j_{rs})$, r, s, 1, …, m, of rank 2 or 0. We define the norm $|j|$ for $j \in J$ as the square root of $\Sigma(j_{rs})^2$ summed up for $1 \leq r < s < m$). We denote by J_1 the subset of J consisting of the j of norm 1 and by J* the hyperspace obtained from J by deleting the condition that j be of rank 2 or 0. We write further (E, J), (E, J_1), (E.J*) for the Cartesian product of a subset E of X with J, J_1, J*.

We write for the space of continuous functions $f = f(x, j) \in (X, J_1)$ with norm $\|f\| = \sup |f(x, j)|$ for $x \leq B$, where B is sufficiently large constant depending on the context. Just as for generalized curves we require homogeneity of f in j, *i.e.*, $f(x, \alpha j) = \alpha f(x, j)$ for $\alpha \geq 0$ so that the extension of f is defined by its restriction to F. the function f is called integrand and this is analogous to the Larangian for generalized curves. The classical generalized integral is defined as,

$$L(f) = \int_R f(x(u,v),j(u,v))dudv, \ f \in F \quad \textbf{(17)}$$

A parametric surface is the equivalence or maximal class consisting of $x \in D$ all of which give the same value to the functional L(f) for each integrand f (or for its symmetric integrand). Each integrand f in the class is a representation of L(f). We define L(f) a generalized surface. A generalized surface L(f) is termed fine limit of the sequence of generalized surfaces $L_n(f)$, n = 1, 2, …, if for each relevant integrand f (each f in or symmetric f) the values of $L_n(f)$ converge weakly to L(f). It can be seen that every generalized sequence is expressible as the limit of a sequence of uniformly bounded parametric surfaces (ordinary surfaces). Such convergence induces a norm on the linear functional representing surfaces defined by the expression sup $|L(f) – L'(f))$, where the supremum is taken over all $f \in F$ whose norm is ≤ 1 (or symmetric f). When $f_0(x, j) = 1$ for $(x, j) \in (X, J_1)$ we term $L(f_0)$ the area of the generalized surface L. Note that in this formulation a generalized surface is a linear functional. The area of a generalized surface L(f) is L(1) (which corresponds to the length of a generalized curve gf when f is the constant 1).

The requirement of continuity for f is not necessary; measurability suffices. This can be further weakened by allowing set-valued integrand with appropriate probability distribution. In fact, in taking the limit of parametric representation of surfaces we may have set-valued integrand. Therefore, we introduce the notion of generalized derivative of an integrand as the expectation or average of its set-values at a point in each argument. This is the same as the equivalence class of probability distributions that yields the same expectation. For surfaces we can also admit set-valued jacobian with a probability distribution. This way we can deal with the notions of bifurcation and multifurcation of surfaces and chaos. Each surface that emerges has a certain probability of being actualized, such probability being determined by the probability distribution of its jacobian.

The space of generalized surfaces is complete in the norm defined by this fine convergence. This provides the needed existence theory which is local, *i.e.*, the existence of a surface subject to suitable initial and boundary conditions is valid only in a suitable neighborhood of a point. This is all we can expect since it is defined by differential equations describing local properties of the underlying space. However, it is shown in [114 – 116] that a global generalized surface is the fine limit of a sequence of suitable conventional surfaces. In Young's theory, the space of surfaces is enlarged to insure existence. This is achieved by allowing set-valued jacobians of a vector partial differential equation satisfying the requirements in [114 – 116] with unit measure or probability distribution. A generalized jacobian is the expectation of a set-valued jacobian with probability distribution. Thus, the local solution of a partial differential equation with set-valued jacobian depends on its probability distribution; consequently, the solution space forms a family of surfaces each of which is a representation of a generalized surface. A conventional surface is a solution of a conventional partial differential equation (with well-defined jacobian) satisfying requirements in [114 – 116]. Probability distribution introduces uncertainty in the space of generalized surfaces. Another level of uncertainty is brought in by the proof of local existence of a surface which uses fine convergence of a sequence of conventional surfaces along with the completeness of the space of generalized surfaces.

The Cosmic Sphere [42] which we shall discuss later is modeled quantitatively by a generalized surface.

The Integrated Pontrjagin Maximum Principle

We summarize the integrated version of the Pontrjagin maximum principle that we shall need in Chapter 4.

The Integrated Version

This version of the Pontrjagin maximum principle is developed in [117] (with slight improvement in [22]) from the original principle by Pontrjagin and his associates [91] to apply to generalized curves, specifically, relaxed trajectories of optimal control theory. Anchored on some existence theorems, it was meant to rid optimal control theory of the Perron paradox, a contradiction arising from the use of necessary condition without existence theory [117]. The principle itself is a necessary condition consisting of three parts but it rests on proof of existence of optimal solution of the optimal control problem among relaxed trajectories.

This principle was used to calculate the trajectories, positions and velocities of the n bodies in the gravitational n-body problem once the qualitative solution was provided by GUT [41]. But why do we really need to go through this level of sophistication when GUT has already determined that the n-bodies fall to or orbit around the cores of their respective cosmological vortices along rotating spiral streamlines? The streamlines are not ordinary curves but generalized curves, *i.e.*, they have set-valued derivatives, and for purposes of applications we need the specific equation of the trajectory of each body in the problem. In fact, we can look at a falling body as being subjected to two-valued control set one value being the pull of gravity and the other the impact of the centrifugal force of spin. The path of a body is what is called infinitesimal simplicial curve, *i.e.*, piece-wise arcs corresponding to alternating constant values of the control set U consisting of two elements. A generalized curve called relaxed trajectory is the limit of a sequence of simplicial curves in the Young measure [117]. Superposed on it is another generalized curve due to the micro component of turbulence but has no visible impact on this macro problem.

Formulation of the Problem

We first summarize the formulation of the time-optimal control pre-problem for each body in the original naïve optimal control theory [117], *i.e.*, without the benefit of existence theory. Then we update the formulation to be able to utilize the Integrated Pontrjagin Maximum Principle. We ask for the minimum of the integral,

$$\int f(t,x,u,w)dt, \qquad \textbf{(18)}$$

for trajectories x(t), controls u(t) and constants w and subject to,

$\dot{x} = g(t,x,u,w)dt,$ **(19)**

where u(t) ranges in the set U, w in W and subject to suitable specified conditions (dimensionality is at our disposal). We can eliminate the constants w by regarding the pair (x, w) as a point in higher space and adjoining the differential equation,

$dw/dt = 0,$ **(20)**

which insures that w is constant along any trajectory. Then we add to our end condition that the initial or final values of the projection w of (x, w) lie in W. Thus, the only effect of the constants is to alter dimensionality and end conditions.

To further simplify the problem we introduce another coordinate x_0 subject to,

$\dot{x}_0 = f(t,x,u),$ **(21)**

and write x and g for the pairs (x_0, x) and (f, g). We also add the end condition x_0 at the final end of a trajectory which correspond to the time t = 0; in other words, we reverse time and find the minimum of $-x_0$ for trajectories x(t) and u(t) subject to

$\dot{x} = g(t,x,u),$ **(22)**

where u(t) ranges in U and x(t) satisfies appropriate end conditions. Thus, in this problem what is being minimized is the function x_0 which in applications can be the cost function. Without loss of generality we assume that the pair of endpoints of x(t) belongs to some pre-assigned closed set B of the Cartesian product of x-space with itself (*i.e.*, the initial values of t is not restricted directly, e.g., we can set $t + t_0$ in place of t). We denote by G(t, x) the set of values of the vector g(t, x, u)

for fixed (t, x) as u ranges in U. Then we ask for the minimum of $-x_0$ subject to the condition,

$\dot{x} \in G(t).$ **(23)**

The problem with constraint (22) is called the controlled pre-problem; the one with constraint (23) the uncontrolled pre-problem. The latter has larger space of trajectories from which to find the minimum. As the space of trajectories becomes larger the better is the chance of existence of minimum. The space of trajectories of either problem is still conventional; we can still improve the solution by enlarging the space beyond conventional trajectories into the space of generalized curves. This is exactly what we need for the gravitational in-body problem because the trajectories we are looking for are the spiral streamlines each of which is the local resultant of the effect of gravity and centrifugal force mathematically modeled by relaxed trajectory. L. C. Young set the machinery for doing so [117] as follows.

Instead of the constraint equation,

$\dot{x} = u,$ (24)

where the velocity vector is controlled directly and coincides with the control vector u we attach a probability or unit measure to u (normalized probability distribution) so that the actual velocity dx/dt becomes the integral of u with respect to this unit measure. This is called the weighted average or expectation value for this probability measure. Then the control function u(t) that yields a specific trajectory be replaced by the probability measure v(t). Such measure is called chattering control value v, and we say that it reduces to a conventional control u if the measure is totally concentrated at u. We write V for the space of chattering control values v, *i.e.*, the set of unit measures on U.

Existence Theorems

We quote some existence theorems on conventional and relaxed trajectories from [117].

(1) Existence of Solutions to the General Initial Value Problem for Ordinary Differential Equations

Let f(t, x) be a vector-valued function with values in n-space and suppose in some neighborhood of (t_0, x_0), f(t, x) is continuous in x for each t, measurable in t for each x, and uniformly bounded in (t, x). Then there exists an absolutely continuous function x(t) defined in some neighborhood of t_0 such that $x(t_0) = x_0$ and, almost everywhere in that neighborhood,

$$\dot{x} = f(t,x(t)) \tag{25}$$

This theorem is of fundamental importance in both optimization and approximation theories, calculus of variations and optimal control.

(2) Halfway Principle of McShane and Warfield

Suppose given a continuous map p from Q to P, and a measurable map p** from R to P such that*

$$p^{**}(R) \subset p^{*}(Q) \subset P \tag{26}$$

Then there exists a measurable (lifting) map q* from R to Q such that

$$p^{**} = p^{*}q^{*}. \tag{27}$$

We denote by $\hat{G}(t, x)$ the set of the values of G(t, x, v) when (t, x) is kept fixed and v allowed to vary in V.

(3) First Corollary

Let x(t) be continuous in the finite time interval T and let z(t) be measurable vector-valued function in T such that $z(t) \in G(t, x)$ or $z(t) \in \hat{G}(t, x)$. Then there exists a measurable conventional or chattering control u(t) or v(t) such that z(t) = g(t, x(t), u(t)) or z(t) = g(t, x(t), v(t)).

(4) Second Corollary (the Filippov Lemma)

If, in particular, x(t) is an (uncontrolled) conventional trajectory subject to $\dot{x} \in G(t, x)$ or relaxed trajectory subject to $\dot{x} \in \hat{G}(t, x)$ almost everywhere. Then there exists a measurable conventional or chattering control u(t) or v(t).so that x(t) coincides with the corresponding controlled trajectory satisfying the differential equation

$$\dot{x} = g(t,x(t),u(t)) \text{ or } dx(t)/dt = g(t,x(t),v(t)), \tag{28}$$

almost everywhere.

(5) Uniqueness Theorem for the Initial Value Problem of an Ordinary Differential Equation dx/dt = f(t, x)

Suppose, in addition to the hypothesis of (1), that for some constant K, the function f(t, x) satisfies, whenever (t, x_1) and (t, x_2) lie in some neighborhood N of (t_0, x_0), the Lipschitz condition

$$|f(t,x_2) - f(t,x_1)| \leq K \, |x_2 - x_1|. \tag{29}$$

Then in some neighborhood of t_0 there exists one and only one absolutely continuous function x(t) such that

$$x(t) = x_0 + \int_{\Delta} f(\tau,x(\tau))d\tau, \tag{30}$$

where $\Delta = [t, t_0]$.

Let $f \in C_o(T \times U)$, the space of continuous functions on f(t, u) on $T \times U$, where U is the set of control values u, and T is some fixed time interval. We write Δ for some variable time interval of U. For such pair (f, Δ) consider the function w of (f, Δ) defined by the integral

$$w(f,\Delta) = \int_\Delta f(t,v(t))dt, \quad \textbf{(31)}$$

where v(t), $t \in T$ is a measurable chattering control (or v is conventional control, *i.e.*, unit measure concentrated at some point $u \in U$). We understand the integrand f(t, v(t)) as shorthand notation for an integral of f, for constant t, with respect to probability measure v(t) on U. then we regard w as a measure and identify w(f, Δ) with the integral

$$\int_{\Delta\times U} f dw. \quad \textbf{(32)}$$

Then we write,

$$w = v(t)dt, \quad \textbf{(33)}$$

Bearing in mind that (31) is a double integral; thus, every control measure is determined by a chattering control v(t).

The control measure w will be termed simplicial if it is defined by (31) where v(t) reduces to a conventional piece-wise constant control u(t). Then we use formula (32) and reinterpret u(t) as a unit measure on U concentrated at one point u(t). We denote by W the space of all control measures w. We introduce fine convergence in this measure. A sequence of control measures w_ν, ν = 1, 2, … is termed convergent if, for each f, the values $w_\nu(f, \Delta)$ tend to a limit w(f, Δ) uniformly in Δ; then we say that w_ν tends to w.

(6) Theorem

(i) The space W is sequentially complete. (ii) In order that a real function w of f and Δ be of the form w(f, Δ) where $w \in W$, the following system of conditions is both necessary and consistent.

- *(a) w(f, Δ) is linear in f and additive in Δ;*
- *(b) $f(t, u) \geq 0$ in $\Delta \times U$ implies $w(f, \Delta) \geq 0$;*
- *(c) $f(t, u) = 1$ in $\Delta \times U$ implies $w(f, \Delta) = |\Delta|$.*
- *(d) Theorem. (i) The space W is sequentially compact. (ii) Simplicial control measures are dense in W.*

We term bundle of relaxed, conventional or simplicial trajectories the family of trajectories which meet a given bounded closed subset of (t, x)-space corresponding to closed time intervals, possibly degenerate ones all contained in a fix time interval.

A sequence of functions

$$x_\nu, t \in T_\nu, t \in T_0, \nu = 1, 2, \dots, \quad \textbf{(34)}$$

where T_ν are closed time intervals all contained in some fixed time interval will be said to converge uniformly to

$$x_0, t \in T, \quad \textbf{(35)}$$

if, first, T_0 x_ν, $t \in T_0$, a closed time interval whose extremities are the limits of those of T_ν, and, second, for some choice of a closed time interval T containing T_0 and all but a finite number of the T_ν, these exist, for large ν, extensions of our functions of the form,

$$x_\nu, t \in T, \quad \textbf{(36)}$$

which tend uniformly to a corresponding extension to T of $x_0(t)$ (T_0 may be a point).

(7) Theorem

A bundle of relaxed trajectories is sequentially compact and complete and the corresponding bundle of simplicial trajectories is dense in it.

(8) Corollary

Suppose the set G(t, x) of the values of g(t, x, u) for fixed (t, x) is convex. Then any bundle of conventional trajectories is sequentially complete and compact, and the corresponding bundle of simplicial trajectories is dense in it.

(9) Existence Theorem for Relaxed Solutions

Let Q be a bounded closed set of (t, x)-space, P a closed set in the Cartesian product of (t, x)-space with itself and T a closed finite interval of t. We denote by Σ the set of relaxed trajectories x(t) defined on closed subintervals of T which meet Q and p a pair of extremities situated in P. The function g(t, x, u) which appears in the differential equation (5) is supposed continuous and subject to the Lipschitz condition in x. Then either Σ is empty or there exists in Σ a relaxed trajectory for which the difference at the endpoints of the coordinates x_0 of x assumes its minimum.

The Pontrjagin Maximum Principle: Integrated Version

For greater generality we suppress dependence on the chattering control v(t) by writing g(t, x) for g(t, x, v(t)). Consider a convex family G of such functions, a family such that every convex combination

$$\Sigma\alpha_i g_i, \tag{37}$$

of a finite number of members g_i, of G with constant coefficients $\alpha_i \geq 0$, where $\Sigma\alpha_i = 1$, is itself a member of G (in the chattering case the family G are functions g of the form g(t, x, v(t)) and convexity holds in a stronger sense in which the coefficients α_I are allowed to be measurable functions of t instead of constants). In addition we require that every function g(t, x) in G is continuously differentiable in x for fixed t and measurable in t for fixed x, and also that each g and its partial derivative g_x are bounded functions of (t, x) or, more generally, bounded in absolute value by some integrable function of t only. These various requirements are to hold only in some bounded open set O of (t, x)-space. In the chattering case all these requirements are satisfied if we make the stipulation that g(t, x, u) is continuously differentiable.

We consider the family

$$H = yG \tag{38}$$

of Hamiltonian functions h(t, x, y) = yg(t, x), where y is a variable vector and each $g \in G$ gives rise to a corresponding $\lambda \in H$. we shall be concerned with points (t, x) that lie in a sufficiently fine neighborhood of the set described by a given fixed trajectory C of the form x(t), $t_1 \leq t \leq t_2$. In terming C a trajectory we imply that the function x(t) is, almost everywhere in its interval, a solution of the differential equation,

$$\dot{x} = g(t, x(t)), \tag{39}$$

for some fixed corresponding member $g \in G$; moreover, x(t) is to be absolutely continuous.

We term ordinary point of C a point at which (39) holds; in particular, we say that C has ordinary endpoints if the derivatives $\dot{x}(t_i)$ exist and have the values $g(t_i, x(t_i))$, i = 1, 2. We assume function x(t) continued outside its interval domain, when convenient, subject to the same differential equation and absolute continuity. In view of the uniqueness theorem any such extension is uniquely determined once we fix the member $g \in G$ and an initial condition of the type $x(t_0) = x_0$.

We write q for the ordered pair of endpoints of C and P for a small neighborhood of q. Thus, p lies in the space of such ordered pairs q, *i.e.*, in the Cartesian product of (t, x)-space with itself or, equivalently, O with itself. We denote by Q the subset of P consisting of ordered pairs (p, q) of endpoints of trajectories in O, sufficiently close to C. Any such trajectory has the form χ(t), $\tau_1 \leq \tau \leq \tau_2$, where χ(t) is absolutely continuous and satisfies, for almost all t in its interval of definition, a differential equation similar to (34) with g replaced by some member $\hat{g}$ of form g(t, x, v(t)).

In P suppose given a smooth manifold M with q as boundary point where M is represented by local coordinates as a smooth one-to-one map of a smooth Euclidean domain with its boundary. We take the interior of M and boundary of

M as corresponding images of the interior and boundary of this domain. We suppose the dimension of M to be ≥ 1 so that it does not reduce to a point. We suppose, further, that the boundary of M at q has a tangent subspace, which is itself the boundary of a tangent half-space (a tangent half-subspace by taking half-lines tangent to M at q). We suppose that a neighborhood of q in M has a continuous one-one map onto a neighborhood of q in this tangent half-subspace such that q corresponds to itself and that, if $q + \delta p$ denotes the image of p in M, we have,

$$P = q + \delta p + o(p - q) \quad \textbf{(40)}$$

where $o(p - q)$ is small compared with $p - q$ as $p \to q$. In particular, local coordinates can be thought of as coordinates on the tangent half-subspace. Moreover, a vector $\phi \neq 0$ in the underlying $(2n + 2)$-dimensional Euclidean space will be termed an inward normal of M at the point q if it is, first, orthogonal at q to the boundary of M, *i.e.*, to a hyperplane through q that contains the tangent subspace of this boundary and, second, directed towards the side of this hyperplane that contains the tangent half-subspace of M.

A trajectory C is an M-extremal if it contains no interior point of M; we term conjugate vector along C an absolutely continuous and nowhere vanishing vector-valued function y(t) with values in n-space defined on the same interval as x(t). If h is the Hamiltonian corresponding to the element $g \in G$ that enters into the differential equation (38) satisfied by x(t), we term corresponding momentum and denote by $\eta(t)$ the (n+1)-dimensional vector derived from y(t) by the adjunction of the initial component

$$-h(t, x(t), y(t)) = \eta_0(t). \quad \textbf{(41)}$$

We term corresponding transversality vector for C the (2n+2)-dimensional vector,

$$(-\eta(t_1), \eta(t_2)). \quad \textbf{(42)}$$

<u>The Integrated form of the Pontrjagin Maximum Principle</u>

Let $g \in G$, h be the corresponding Hamiltonian function yg(t, x) and let C be an M-extremal of the form x(t), $t_1 \leq t \leq t_2$, satisfying, almost everywhere, the corresponding differential equation (21) which we now write $\dot{x} = \partial h/\partial y$. If C has ordinary endpoints or M consists of pairs with the same coordinates as q, then there exists a conjugate vector y(t) along C such that the pair (x(t), y(t)) satisfies the following three condition:

(a) The canonical Euler equations:

$$\dot{x} = \partial h/\partial y,\ \dot{y} = -\partial h/\partial y. \quad \textbf{(43)}$$

(b) The Weierstrass condition: As function of $\lambda \in H$, the quantity,

$$\int_\Delta h(t, x(t), y(t))dt, \quad \textbf{(44)}$$

assumes its maximum when $\lambda = h$.

(c) The transversality condition:

Then the transversality vector (42) is an inward normal of M.

NEW MATHEMATICS

We shall consider here the new mathematics involved the development of GUT, namely, the new real number system, chaos and turbulence and generalized fractal, integral and derivatives but we focus on the new real number system and the generalized integral; the rest we shall summarize or refer to original sources. We have already introduced qualitative mathematics and qualitative modeling, the most important new mathematics involved in GUT. The new real number system provides qualitative models for important physical systems like the superstring and our universe. Therefore, it is a very important component of the new methodology.

Fractal, Generalized Fractal and Chaos

The generalized fractal of particular importance to GUT that we consider here is nested generalized fractal. (See [11, 48, 53, 87, 109] on geometrical fractal)

Definitions and Examples

Classical fractal is iterated affine transformation of a given generator, some geometrical figure in the case of a geometrical fractal; it is mainly quantitative [13, 24]. Affine transformation is a combination of contraction and translation the effect being to generate a sequence of self-similar figures, *i.e.*, each term except for the first is similar to the preceding term, at decreasing scale. We generalize this fractal construction to include as well, rotation, taking mirror image, sliding along a curve and replication all of which preserve similarity. The last characteristic is the most important for applications, especially, in biology. We have seen this replication in the form of splitting or branching of roots and branches of a tree and the veins of its leaf. This also happens in mitosis or cell-division where self-similarity is in terms of the replication of the genes in every offspring cell. The last two examples are physical fractals where the sequences involved are finite since visible physical systems are finite our visible universe being finite [25]. Moreover what is replicated need not be geometrical figure but general properties of the terms of the fractal sequence. We call this formation generalized fractal; only qualitative mathematics is capable of modeling generalized fractal, especially, when there is multiple replication or multifurcation. In the root of a tree every branch continues as nested fractal, *i.e.*, each term in the sequence is contained in or a part of the preceding term (in this case at decreasing scale). The characteristic here is replication more than self-similarity and decreasing scale because the terms may not be geometrical but some characteristics or processes as in the above examples of replication. Mitosis in a living cell is a special fractal that biologists use to describe its replication in the offspring cell, especially, its genetic content; here decreasing scale does not apply but self-similarity and replication do.

Nested fractal is nature's way of packing huge energy in a physical system or attaining maximum efficiency in a physical process, an expression of energy conservation and a universal configuration of nature that applies even to man-made structures such as machines, buildings and bridges due to its optimal properties. In biology, encoding of information in the brain is a fractal process [29].

The fractal structure of the roots of a tree allows it to absorb maximum nutrients from the ground and that of its branches and veins of its leaf allows optimal efficiency in the distribution of nutrients to the stomata for food production (photosynthesis), e.g., fruit, and distribution to where they are needed including fruits that humans can harvest.

Chaos is mixture of order none of which is identifiable. For example, in regions where there is much under-ocean volcanic activity the ocean surface heats up, throws the gas molecules above it into motion (kinetic energy) that pushes them apart and creates low pressure. Low pressure sucks gas molecules around it and the initial rush throws trillions of gas molecules into collision that makes it impossible to monitor or even predict the path of any single one. At the same time, every molecule is subject to the laws of nature. This is then a classic case of chaos but it is transitional since collision is energy dissipating. Energy conservation induces global order and turns it into coherence of order called turbulence in this case a hurricane.

The formation of tropical cyclone is an example of standard dynamic system. It starts on a calm summer day which is order; then chaos ensues as a transition to coherence of order called turbulence, in this case, a cyclone which is a vortex of gas molecules in the atmosphere. This transition from chaos to turbulence is due to energy conservation. Chaos is energy dissipating; in this example it is due to collision of gas atoms and molecules which is distortion of order; therefore, energy conservation induces its evolution into global order, the cyclone. Then the cyclone vanishes when nothing infuses energy on it, e.g., warm corridors on the ocean, and its energy dissipates into the atmosphere; when a cyclone passes through a warm corridor its power rises because heat or kinetic energy is fed to it. It was once suggested that an atom bomb be dropped into the eye of a typhoon to end it. That would be like pouring gasoline into brush fire. At the same time, when the eye hits and gets plugged in by a mountain its power declines because friction dissipates its energy.

It is impossible to model chaos computationally, only qualitative modeling can. Another example is fundamental chaos or dark matter one of the two fundamental states of matter that we shall consider later; fundamental in the sense that it is part of the cycle of matter: all matter comes from it and will ultimately return to it. However, fundamental chaos is not energy dissipating because the semi-and non-agitated superstrings do not interact among themselves; therefore, it is stable and has zero entropy, the only stable physical chaos. (The real line is mathematical chaos [24, 78]; so are infinitesimal zigzag and oscillation [24, 78]).

The Peano Space-Filling Curve

The problem here is to map the unit interval [0, 1] *onto a* unit square by continuous function. The usual tool is a theorem that says: the limit of a uniformly convergent sequence of continuous functions is continuous. But all constructions so far are difficult to visualize. We make the geometrical construction of this curve quite simple [27, 34] using the concept of nested fractal sequence.

Divide the unit square into 9 little squares by the lines,

$$x = 1/3, x = 2/3, y = 1/3, y =/2/3. \quad \textbf{(45)}$$

We label the blocks, B_1, B_2, B_3, B_4, B_5, B_6, B_7, B_8, B_9, starting at the bottom row from left to the right, then up to the middle, then left to the first block at the middle row, then up on this block to the top row, then right to the last block of the top row to the right hand corner of the square (see Fig. **4**); details are discussed below. We take as the first initial generator the segment $g_{1,1}(x)$: $y = x, 0 \leq x \leq 1/3$; take the mirror image of (flip over) $g_{1,1}(x)$ with respect to the y-axis to obtain the second initial generator $g_{1,-1}(x)$: $y = -x, 0 \leq x \leq 1/3$; take the mirror image of the second initial generator to obtain to obtain the third initial generator, $g_{-1,-1}(x)$: $y = x, 0 \geq x \geq -1/3$ and take the mirror image of the third generator to obtain the fourth generator, $g_{-1,-1}(x)$: $y = x, 0 \geq x \geq -1/3$. The four initial generators are:

$$g_{11}(x): y = x, 0 \leq x \leq 1/3; g_{-11}(x): y = -x, 0 \leq x \leq 1/3;$$

$$g_{-1-1}(x): y = x, 0 \geq x \geq -1/3; g_{1-1}(x): y = -x, 0 \geq x \geq -1/3. \quad \textbf{(46)}$$

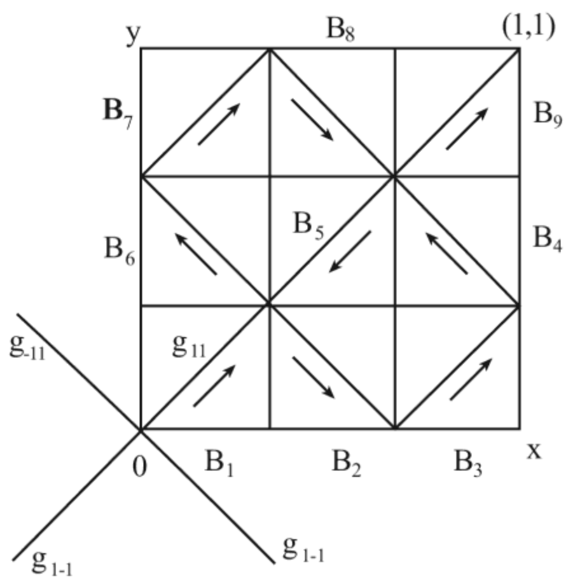


Figure 4: The first term $f_1(t)$ of the fractal sequence and its generators g_{11}, g_{-11}, g_{1-1} and g_{1-1} in the construction of the Peano space-filling curve obtained by translating them suitably to form the ordered polygonal line from the origin to the point (1, 1). Arrows indicate direction of the polygonal line.

Then the initial function $f_1(t)$ in the iteration process follows:

$$f_1(x): g_{11}(x), g_{1-1}(x), + (1/3.1/3), g_{11}(x) + (2/3),$$

$g_{-11}(x) + (1, 1/3)$, $g_{-1-1}(x) + (2/3, 2/3)$,

$g_{-11}(x) + (1/3, 1/3)$, $g_{11}(x) + (0, 2/3)$,

$g_{1-1}(x) + (1/3, 1)$, $g_{11}(x) + (2/3, 2/3)$, $0 \leq x \leq 1/3$. **(47)**

The functions that comprise $f_1(x)$ are suitable translations of the generators in (43) to form a polygonal line through the diagonals of

blocks $B_1, B_2, \ldots, B_9$, in that order with suitable orientation. Construction of $f_1(x)$:

(1) The first segment of $f_1(x)$ is $g_{11}(x)$.

(2) For the second segment in block $\mathbf{B_2}$, use $g_{1-1}(t)$ by translating it to the right by 1/3 units and up by 1/3 units so that its initial point coincides with the terminal point of $g_{11}(x)$ and terminal point coincides with the lower right hand vertex of $\mathbf{B_2}$.

(3) For the second segment, simply translate $g_{11}(x)$ to the right by 2/3 units so that the initial point coincides with the terminal point of the previous segment and the terminal point with the upper right hand vertex of $\mathbf{B_3}$.

(4) Then we move up to $\mathbf{B_4}$. The segment that will go here is suitable translation of $g_{-11}(x)$, etc. Then follow the same construction with suitable translations of generators through $\mathbf{B_5}$, $\mathbf{B_6}$, …, $\mathbf{B_9}$ so that the end point of $f_1(x)$ coincides with the point (1, 1). The initial construction of $f_2(x)$ is shown in Fig. **5**.

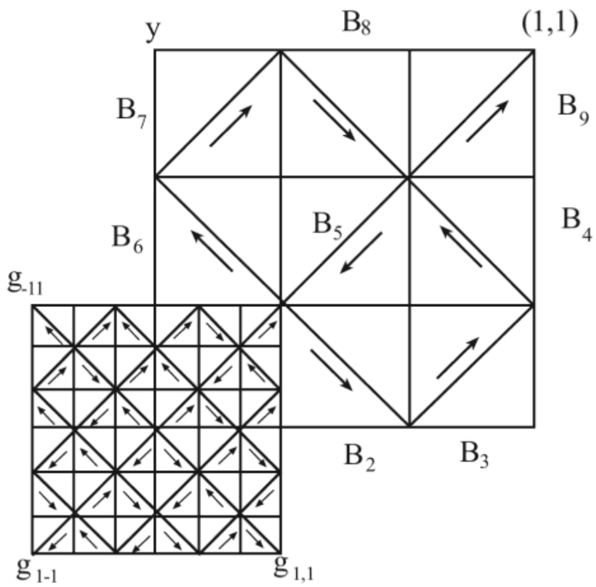


Figure 5: Contracting $f_1(t)$ by 1/3 at the origin yields one of the generators $g_{22}(t)$; the other generators are as follows: g_{-22}, mirror image of g_{22} about the y-axis, g_{-2-2}, mirror image of g_{-22} about the −x-axis and g_{2-2} that of g_{-2-2} about the −y-axis. Suitable translation of these generators forms a finer polygonal line from the origin to the point (1, 1), the second term $f_2(t)$ of the fractal sequence [71].

The second set of generators for the second term $f_2(x)$ of the fractal sequence is constructed as follows:

(1) Contract f_1 through its projection cone with vertex at the origin and call its image g_{22}.

(2) Flip g_{22} over the y-axis, *i.e.*, take its mirror image with respect to it, and call it g_{-22};

(3) Flip g_{-22} over the −x-axis and call the image g_{-2-2};

(4) Flip g_{-2-2} over the −y-axis and call the image g_{2-2}.

(Fig. **5** shows the construction of the second set of generators)

We construct $f_2(x)$ using the generators $g_{22}(x)$, $g_{-22}(x)$, $g_{-2-2}(x)$ and $g_{2-2}(x)$ as follows:

$f_2(x)$: $g_{22}(x)$, $g_{2-2}(x) + (1/3, 1/3)$, $g_{22}(x) + (2/3, 0)$,

$g_{-22}(x) +(1.1/3)$, $g_{22}(x) + (2/3, 0)$, $g_{-2-2}(x) + (2/3, 2/3)$,

$g_{-22}(x) + (1/3, 1/3)$, $g_{22}(x) + (0, 2/3)$,

$g_{2-2}(x) + (1/3, 1)$, $g_{22}(x) + (2/3, 2/3)$, $0 \leq t \leq 1/3$. **(48)**

The next phase is, again, contraction of $f_2(x)$ by 1/3 to form the first of the four generators of f_3 denoted by $g_{33}(x)$ and taking suitable mirror images to complete its four generators. Then we iterate this phase of the construction to generate the sequence of functions, $f_1(x)$, $f_2(x)$,..., $f_n(x)$, with the following properties:

(1) Each $f_n(x)$, n = 1, 2,..., is made up of polygonal lines as suitable translations of the generators of $f_n(x)$;

(2) Self-similarity is obvious from the construction since the orientation of $f_n(x)$, for each n, is preserved;

(3) Each $f_n(x)$ in the sequence is continuous;

(4) The sequence $f_n(x)$, n = 1, 2,..., is uniformly convergent;

(5) Therefore, the sequence $f_n(x)$, n = 1, 2,..., converges to a continuous function f(x) whose graph clearly fills up the unit square.

Filling up the Unit Cube

The above construction can be extended to fill up the unit cube by the continuous mapping of the unit interval as follows:

(1) Consider the unit cube with vertices A(0, 0, 0), B(1, 0, 0), C(0, 1, 0), D(0, 0, 1).

(2) Subdivide the cube by the planes,

$x = 1/3, x = 2/3, y = 1/3, y = 2/3, z = 1/3, z = 2/3,$ **(49)**

into 27 little cubes.

(3) As in the construction of the Peano space-filling curve, connect the diagonals of the little cubes from A(0, 0, 0) suitably in the right order to form a polygonal line through the little cubes once and ending up at so that the entire cubes is covered stretch and deform the unit interval AB to a polygonal line and map each segment in suitable order into the diagonals of at the origin joining the vertices B(0, 0, 0) and B(1/3, 1/3, 1/3),..., B(2/3, 2/3, 2/3) and B(1, 1, 1), where the last segment is mapped into the diagonal of the little cube opposite the first little cube at the origin and the terminal point at the opposite corner of the cube at its diagonal. Call this polygonal line f_1.

(4) Contract f_1 to 1/3 into the first cube (by pushing it through its projection cone with vertex at the origin. This contracted polygonal line becomes the first generator g_{111} of f_2; take the mirror image of g_{111} about the y-axis; the image is the second generator g_{1-11} of f_2; take the mirror image of g_{1-11} about −x-axis to obtain the third generator g_{1-1-1}; take the mirror image of g_{1-1-1} about the –y-axis to obtain the fourth generator of f_2.

(5) Suitably translate these generators as in the construction of f_1 to form the appropriate finer polygonal line from the origin to the point (1, 1) of the unit square made up of the segments of f_1 through the contracted cubes with the terminal point at the vertex of the original cube opposite the origin.

(6) Continue this iteration procedure to obtain a uniformly convergent sequence of continuous functions f_1, f_2, … whose point-wise limit is a continuous function. We shall call this the Peano cube space-filling curve.

This construction is constructivist, intuitive and the simplest so far. Earlier constructions use this theorem: the limit of a uniformly convergent sequence of continuous functions is continuous. This construction can be extended to the n-hypercube where n is odd.

The Infinitesimal Zigzag as Limit Set of Fractal

Consider without loss of generality the equilateral triangle ADB of Fig. **1** with one vertex at the origin A and with base AB. We denote by $g_{11}(t)$ side AD which we take as the initial generator. We take the mirror of $g_{11}(t)$ with respect to the y-axis and denote it by $g_{-1,\,1}(t)$ as the second generator. The third generator is the mirror image of the second with respect to the negative x-axis denoted by $g_{-1-1}(t)$ and the fourth generator is the mirror image of the third with respective to the negative y-axis denoted by $g_{1-1}(t)$. Their parametric equations are given by:

$$g_{11}(t)\text{: } (x, y) - (t, t),\ g_{1-1}(t) : (x, y) = (y, -t),\ 0 \leq t \leq 1/2. \qquad \textbf{(50)}$$

To construct function $f_1(t)$, we take $g_{11}(t)$, $0 \leq t \leq ½$, as part of $f_1(t)$ and combine it with the translation of $g_{1-1}(t)$ given by $(x, y) = g_{1-1}(t) + (1/2, 1/2)$, $0 \leq t \leq 1/2$. Its graph is the polygonal line formed by segments AD and DB.

To construct $f_2(t)$ in the iteration process, we contract $g_{11}(t)$ by ½ and denote this by $g_{22}(t)$. This is one of the generators. Rotate $g_{22}(t)$ by $-\pi/2$ to form another generator $g_{2-2}(t)$. The generators are given by $g_{22}(t)$: $(x, y) = (t, t)$, $g_{2-2}(t) = (t, -t)$, $0 \leq t \leq 1/4$.

The function $f_2(t)$ is given by the following system of equations,

$$f_2(t)\text{: } (x, y) = g_{22}(t),\ (x, y) = g_{2-2}(t) + (1/3, 1/4),$$

$$(x, y) = g_{22}(t) + (1/2, 0),\ (x, y) = g_{2-2}(t) + (3/4, 1/4),$$

$$0 \leq t \leq 1/4, \qquad \textbf{(51)}$$

and is represented by the polygonal line APQRB with x replaced by t and y replaced by x.

We iterate this construction to generate a sequence of functions $f_n(t)$ represented by the curve C_n, n = 1, 2,..., with the following properties:

(1) The sequence $f_n(t)$, n = 1, 2,..., is uniformly convergent and each f_n is continuous,

(2) For each n, the length $|C_n|$ of C_n satisfies

$$|C_n| = |C_1| = \sqrt{2}, \qquad \textbf{(52)}$$

(3) The sequence f_n converges to a continuous function $f_n(t)$ represented by the curve C_0: $y_0(t) = 0$ which coincides point-wise with the ordinary constant curve C: $x(t) = 0$, $0 \leq t \leq 1$, distinct from it since

$$|C_0| = \lim |C_n| = \lim |\sqrt{2}| = \sqrt{2} \neq |C|. \qquad \textbf{(53)}$$

This curve C_0 is the infinitesimal zigzag.

One property of the superstring is that, left alone, it shrinks steadily. We model this behavior mathematically using the superposition of a sinusoidal curve over a polygonal line as shown in [24, 34] (the sinusoidal curve is the projection of the helix on the plane through its axis). Given any real number $r > 1$, there exists an isosceles triangle ADB with sides AD and DB having total length $|AD| + |DB| = r$. Following the above construction, we form a sequence of polygonal lines whose limit in the sup norm is $|AB|$ but whose length is r. Thus, AB is a coincidence of countably many curves distinct from AB and from each other. The ordinary segment AB is the visible element of the countably infinite space of generalized fractals. The rest is dark matter.

The preceding construction can be generalized to apply to any triangle, ADB, where the slopes of AD and DB are m_1 and m_2, respectively, and $-\infty < m_1 < \infty$, $-\infty < m_2 < \infty$. Then we can pack countably infinite infinitesimal zigzags into AB whose lengths vary along all values in the interval $|AB|, \infty)$ each of which has set-valued derivative (m_1, m_2) at each point on AB.

A More General Geometrical Fractal

We first illustrate a convenient way of contracting a compacts set.

Given compact set B in $\mathbf{R}^n$ (subspace of the n-Cartesian product of the real line) and any real number s, $0 \le s \le 1$, then the set,

$$sB = \{sb \mid b \in B\}, \quad \textbf{(54)}$$

is similar to B in the sense that for any point $b \in B$ with components bk, k = 1, 2,..., n, sbk/bk = s. *i.e.*, ratios of components are preserved. For a plane set B (Fig. **6**) the set sB is a contraction of B along the projection cone B* towards its vertex at the origin where remains in the line segment joining it to the origin during contraction. In fact, B* can be expressed as

$$B^* = \cup\{tB\}, 0 \le t \le 1, \quad \textbf{(55)}$$

obtained by taking the union of B with its projection cone whose vertex is the origin. Note that since for any real number s, $0 < s < 1$, $\lim s^m = 0$, the compact set B in $\mathbf{R}^n$ can be shrunk towards a point at the origin by iterated construction,

$$sB, s^2B, \ldots, s^mB, \ldots. \quad \textbf{(56)}$$

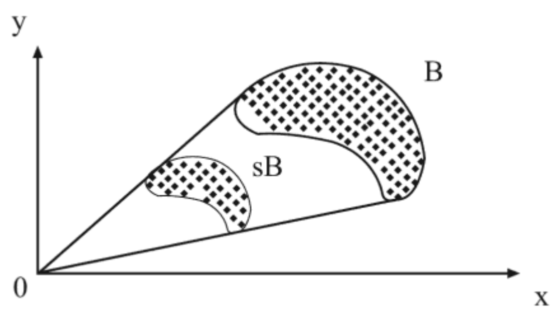


Figure 6: Plane geometrical figure B contracted by s by pushing it suitably through its projection cone towards its vertex at the origin.

Suppose B has diameter of length δ, where

$$\delta = \sup\{d(p,q) \mid p,q \in B\} \quad \textbf{(57)}$$

and d(p, q) is the Euclidean distance between p and q. Let s = 1/3. We rotate and translate B/3 suitably so that its diameter coincides with the x-axis and its left extreme point is at the origin. We translate B/3 along the x-axis twice one at a time so that their images join end to end at extreme points shown in Fig. **7a**. Consider B/3 at the origin and its two images and denote them by B_{11}, B_{12}, B_{13}, respectively; denote their union by

$$B_1 = B_{11} \cup B_{12} \cup B_{12}. \quad \textbf{(58)}$$

B_1 is the generator and first term of the nested fractal sequence we are constructing. We contract B_1, again, through its projection cone by 1/3. The image consists of contraction of B/3 to $B/3^2$ at the origin and similar translation of the latter twice one at a time so that they join end to end as in the previous construction shown in Fig. **7b**. We denote the components of the contraction of B_1 by B_{21}, B_{22}, B_{33}, respectively, where each is the contraction of B/3 to $B/3^2$. Note that the image of B_1 lies inside B_{11} and the extreme right endpoint of B_{23} coincides with the extreme right endpoint of B_{13}. We translate B_{11} (which contains B_1) twice one at a time along the x-axis so that their images are end to end with B_{11} and denote them by B_{21}, B_{22}, B_{23} each component of which is equal to the contraction of B/3 to $B/3^2$. Then the second term of our nested geometrical fractal sequence is given by

$$B_2 = B_{21} \cup B_{22} \cup B_{23}. \quad \textbf{(59)}$$

We iterate the construction and form the nested fractal sequence,

$$B_n,\ n = 2, 3, \ldots . \quad \textbf{(60)}$$

and take k successive translations of $g_{11} = \delta B$:

$$G_{11} = g_{11}$$

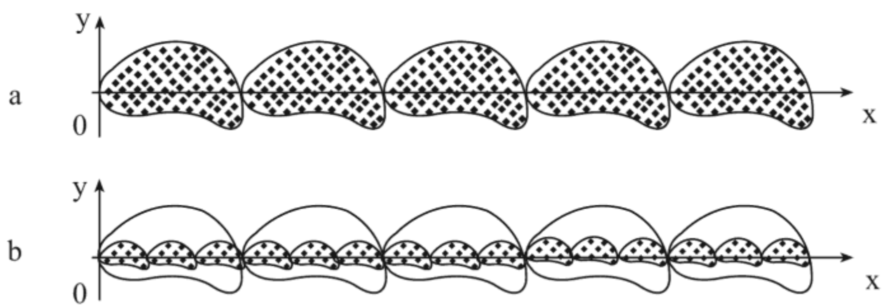
$$G_{12} = g_{11} + (\delta/k, 0, \ldots, 0)$$

$$G_{13} = g_{11} + (2\delta/k, 0, \ldots, 0)$$

$$G_{1k} = g_{11} = ((k-1)\delta/k, 0, \ldots, 0). \quad \textbf{(61)}$$

Take the union of the G_{1m}, m =1, 2, …, k – 1, to obtain,

$$G_2 = \cup G_{1m},\ m = 1, 2, \ldots, k-1. \quad \textbf{(62)}$$



shown in Fig. **7** for a three-component generator.

We iterate this construction to obtain a sequence G_1, G_2, …, G_m, forming a nested geometrical fractal that shrink to the x-axis joining the origin and the point (δ, 0, …, 0). Again, obvious generalization can be done by allowing linear combination of the different contractions of the generators at each stage in the affine transformation.

We note that the set limit, point-wise or in the sup norm, of a nested geometrical fractal set is chaos.

This method can be further generalized by allowing linear combination of different contractions of the generator at each stage in the iteration process.

This fractal mathematics, particularly, infinitesimal oscillation is the key to an understanding of physical singularities such as black hole and the tremendous but dark or latent energy in the nucleus of an atom. As we shrink an oscillation to a point with its length preserved, its energy hν, where h is the Planck's constant and ν its frequency, rises without bounds. However, it is dark (undetectable) with respect to our present means of observation such as light due to difference in orders of magnitude between their frequencies (the same principle that applies to non-resonance of radio or TV reception).

Theorem 3 is really a prescription for pushing matter (of which oscillation is its universal motion and fractal its universal configuration) into the hidden or dark region of matter. It is quite well known in physics today that 95% of matter in our universe is dark [8, 42]. Dark matter consists of non-agitated superstrings, visible matter of agitated superstrings; a primum is agitated superstring [43].

Note that the proofs of the above theorem are geometrical and much simpler and suitable for animation. In fact, fractal construction of the Peano space-filling curve was animated in the presentation of [44]. This construction is an extension of the method used in the Peano space-filling curve as limit of nested fractal sequence.

Fractal is everywhere in nature because it is one of the expressions of energy conservation identified in the law of nature called Energy Conservation Equivalence. It is nature's way of packing huge energy in a physical system or carrying out a process most efficiently.

The New Real Number System

We now optimize the applications of qualitative mathematics to rectify the weakness of the real number system and build the new real number system as the new foundation of mathematics that retains all the interesting and desirable properties of the real numbers.

Our Strategy

Our strategy is not simply to build the contradiction-free mathematical space called the new real number system **R*** but also to meet the needs of natural science and practical affairs while retaining the valid interesting and useful properties of the real number system. Then the new real number system must contain mathematics that has worldwide applications. In particular, it must provide both continuous and discrete mathematics including the decimals whose physical model, the metric system, has worldwide applications that other systems of measures are converting to it. Any other useful mathematics that may arise we shall consider a bonus. Concretely, the new real number system must be a continuum since physical space is which pervades everything and cannot be split into disjoint nonempty subsets. The key is to choose the right consistent axioms upon which to build it. These are the parameters for our construction.

The Terminating Decimals

We first build our base space, the terminating decimals **R** under these axioms:

Axiom 1. 0 and 1 are elements of **R**.

Axioms 2 and 3. The addition and multiplication tables.

Axioms 2 and 3 initially well-define 0 and 1 then the digits or basic integers 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9 and the terminating decimals.

They are the elements of **R** as a mathematical space. The nonterminating decimals belong to the extension of **R** called the new real number system denoted by **R***. The elements 0 and 1 are called the additive and multiplicative identities of **R**, respectively. Initially, they are ill-defined until axioms 2 and 3 well-define them as well as their properties and relationship with the other integers and the terminating decimals.

We first define the digits or basic integers beyond 0 and 1:

$$1 + 1 = 2;\ 2 + 1 = 3;\ \ldots,\ 8 + 1 = 9. \qquad \textbf{(63)}$$

We omit the statements of the addition and multiplication tables which are familiar to everyone since primary school. Then we define the rest of the integers as base 10 place-value numerals:

$$a_n a_{n-1} \ldots a_1 = a_n 10^n + a_{n-1} 10^{n-1} + \ldots + a_1, \qquad \textbf{(64)}$$

where the a_ns are basic integers. The metric system models the system of decimals.

Now, we extend the integers to include the additive and multiplicative inverses −x and, if x is not 0, 1/x (reciprocal of x), respectively. Note that the reciprocal of an integer exists only if it has no prime factor other than 2 or 5. We also extend the operations + and × by re-stating associativity, commutativiy, distributativy, etc., and introduce something else that is new: the rules of sign that we take as part of the axioms of this extension (we need not write them as they are familiar). Then we define subtraction as a new operation: the difference between x and y or y subtracted from x. Then we define another new operation: division of an integer x by a nonzero integer y, or quotient, denoted by x/y and defined by:

$$x/y = x(1/y), \tag{65}$$

provided y is neither 0 nor a pime other than 2 and 5. We similarly extend distributivity of multiplication relative to addition and include them as axioms of the extension. We consider subtraction the inverse operation of addition and division that of multiplication as examples of duality that we shall consider in detail below. Formally, we define subtraction of y from x by the equation: $x - y = x + (-y)$ and division by: $x/y = x(1/y)$. We define a terminating decimal as follows:

$$a_na_{n-1}\ldots a_1.b_kb_{k-1}\ldots b_1 = a_n10^n + a_{n-1}10^{n-1} + \ldots + a_1 + b_1/10$$

$$+ b_2/10^2 + \ldots + b_k/10^k = a_n10^n + a_{n-1}10^{n-1} + \ldots + a_1$$

$$+ b_1(0.1) + b_2(0.1)^2 + \ldots + b_k(0.1)^k. \tag{66}$$

where $a_na_{n-1}\ldots a_1$ is the integral part, $b_1b_2\ldots b_k$ the decimal part and $0.1 = 1/10$. The terminating decimals are well-defined since the reciprocal of 10 has only the factors 2 and 5. If x and y are relatively prime integers, $y \neq 0$, then the quotient x/y of x by y exists only if y has no prime factor other than 2 or 5. Such quotient is called rational.

The Nonterminating Decimals

We define the nonterminating decimals for the first time without contradiction and with contained ambiguity, *i.e.*, approximable by certainty. We build them on what we know: the terminating decimals, our point of reference for all its extensions.

A sequence of terminating decimals of the form,

$$N.a_1,\ N.a_1a_2,\ \ldots,\ N.a_1a_2\ldots a_n,\ \ldots. \tag{67}$$

where N is integer and a_n is called standard generating or g-sequence. Its nth g-term, $N.a_1a_2\ldots a_n$, defines and approximates its g-limit, the nonterminating decimal,

$$N.a_1a_2\ldots a_n\ldots, \tag{68}$$

at margin of error 10^{-n} provided each nth g-term is computable, *i.e.*, there is some algorithm or rule for computing the nth digit from the digits. For example, the nth digit can be the last digit of the sum of the squares of its preceding two digits. The digits of π can be computed from its infinite series expansion. A decimal is normal if every digit is chosen at random the digits [16]. The g-limit of (67) is the nonterminating decimal (68) provided the nth digits are not all 0 beyond a certain value of n; otherwise, it is terminating. As in standard analysis where a sequence converges, *i.e.*, tends to a number, in the standard norm, a standard g-sequence, converges to its g-limit in the g-norm where the g-norm of a decimal is itself. Note that a decimal consists of the integral part, the integer to the left of the decimal point, and the decimal part, the sequence of digits to the right of the decimal point which may be terminating or nonterminating. Then we alternatively define an integer as the integral part of a decimal.

We recall that in the real number system a rational is defined as nonterminating periodic, *i.e.*, the digits are periodic after a certain digit, and a real number is irrational if it is nonterminating and nonperiodic. Each of these concepts is ambiguous for it is impossible to verify if the digits are periodic or not. Thus, the concept *irrational* is ambiguous which we discard so that the decimals belong to two mutually exclusive classes, terminating or rational and nonterminating.

We define the nth distance d_n between two decimals a, b as the numerical value of the difference between their nth g-terms, a_n, b_n, *i.e.*, $d_n = |a_n - b_n|$ and their g-distance is the g-limit of d_n. We denote by **R*** the g-closure of **R**, *i.e.*, its closure in the g-norm.

A terminating decimal is degenerate nonterminating decimal, *i.e.*, the digits are all 0 beyond the nth decimal digit for some n. The nth g-term of a nonterminating decimal repeats every preceding digit at the same order so that if finite terms are deleted the nth g-term and g-limit are unaltered and the remaining terms comprise its g-sequence. Thus, a nonterminating decimal may have many g-sequences and we consider them equivalent for having the same g-limit.

Since addition and multiplication and their inverse operations subtraction and division are defined only on terminating decimals computing nonterminating decimals is done by approximating each term or factor by its nth g-term (called n-truncation) which is a terminating decimal and using their approximation to find the nth g-term of the result of addition or multiplication and its inverse operation as its approximation at the same margin of error. This is standard computation, *i.e.*, approximation by decimal segment at the nth digit. Thus, with our premises we have retained standard computation but avoided the contradictions and paradoxes of the real numbers. We have also avoided vacuous statement, e.g., vacuous approximation, because nonterminating decimals are g-limits of g-sequences which belong to **R***. Moreover, we have contained the inherent ambiguity of nonterminating decimals by approximating them by their nth g-terms which are not ambiguous being terminating decimals. In fact, the ambiguity of **R*** has been contained altogether by its construction on the additive and multiplicative identities 0 and 1.

As we raise n, the tail digits of the nth g-term of any decimal recedes to the right indefinitely, *i.e.*, it becomes steadily smaller until it is unidentifiable from the tail digits of the rest of the decimals. While it tends to 0 in the standard norm it never reaches 0 in the g-norm since the tail digits are never all equal to 0; it is also not a decimal since the digits are not fixed. Since none of the tail digits of a decimal is distinguishable from the rest the set of the tail digits of this set cannot be split into two distinct subsets which makes it a continuum in the algebraic sense.

In iterated computation to get closer and closer approximation of a decimal, e.g., calculating $f(n) = (2n^4+1)/3n^4$, n = 1, 2, …, the tail digits may vary but recede to the right indefinitely and become steadily smaller leaving fixed digits behind that define a decimal. We approximate the result by taking its initial segment, the nth g-term, to desired margin of error by choosing n suitably.

The Dark Number d*

Consider the sequence of decimals,

$$(\delta)^n a_1 a_2 \ldots a_k,\ n = 1, 2, \ldots, \qquad \textbf{(69)}$$

where δ is any of the decimals, 0.1, 0.2, 0.3, …, 0.9, a_1, …, a_k, basic integers (not all 0 simultaneously). We call the nonstandard sequence (68) d-sequence and its nth term nth d-term. For fixed combination of δ and the a_js, j = 1, …, k, in (68) the nth d-term is a terminating decimal and as n increases indefinitely it traces the tail digits of some nonterminating decimal and becomes smaller and smaller until it is indistinguishable from the tail digits of the other decimals. As $n \to \infty$ the nth d-term recedes to the right and tends to some number d, its d-limit in the d-norm, which is never 0 (since the a_js are not simultaneously 0 and each d-term is not 0). It is called dark number d which is indistinguishable from the rest of the d-limits of (68) for all other choices of δ and a_js. Therefore, the set of all dark numbers for all choices of δ and a_js is a countable continuum (since any set of sequences is countable) denoted by d*. Thus, d* is set-valued and a continuum (negation of discrete) of dark numbers; the decimals are joined by the continuum d* at their tails. The dark number d* is a continuum in the algebraic sense since no notion of disjoint open set is involved. Note that while the nth d-term of (69) becomes smaller and smaller with indefinitely increasing n it is greater than 0 no matter how large n is so that if x is a decimal, $0 < d < x$. If an equation or function is satisfied by every dark number d we may substitute d* for d in it so that we can write $0 < d^* < x$ in the above inequality.

At the same time, since the tail digits of all the nonterminating decimals form a countable combination of the basic digits 0, 1, …, 9 they are countably infinite, *i.e.*, in one-one correspondence with the integers but their d-limits, being a continuum, have no cardinality (which applies only to discrete set). Any set whose elements can be labeled by integers or there is a scheme for establishing one-one correspondence between them and the integers is countably infinite. It follows that the countable union of countable set is countable.

Observation

Cantor's diagonal method proves neither the existence of nondenumerable set nor that of a continuum; it proves only the existence of countably infinite set, *i.e.*, the off-diagonal elements consisting of countable union of countably infinite sets. The off diagonal elements are not even well-defined because we know nothing about their digits (a decimal is determined by its digits). Therefore, we have the following:

Corollary

(1) Nondenumerable set does not exist; (2) Only discrete set has cardinality; a continuum has none.

Corollary (1) follows from the fact that a well defined set can be constructed only from at most countable union of finite set. Thus, the continuum hypothesis of set theory collapses. In view of the requirements of a mathematical space that it must be well defined by consistent set of axioms, it is not necessary to develop set theory as a kind of universal language for mathematics since its axioms are not valid in any mathematical space anyway unless there is a set of consistent axioms that well defines it in which case it becomes a mathematical space.

Like a nonterminating decimal, an element of d* is unaltered if finite d-terms are altered or deleted from its d-sequence. When $\delta = 1$ and $a_1a_2\ldots a_k = 1$ (69) is called the basic or principal d-sequence of d*, its d-limit the basic element of d*; basic because all its d-sequences can be derived from it. The principal d-sequence of d* is,

$$(0.1)^n,\ n = 1, 2, \ldots, \quad \textbf{(70)}$$

obtained from the iterated difference,

$N - (N - 1).99\ldots = 1 - 0.99\ldots = 0$, excess remainder of 0.1;

$0.1 - 0.09 = 0$, excess remainder of 0.01;

$0.01 - 0.009 = 0$, excess remainder of 0.001; **(71)**

Taking the nonstandard g-limits of the extreme left side of (70) and recalling that the g-limit of a decimal is itself and denoting by d_p the d-limit of the principal d-sequence on the rightmost side we have,

$$N - (N - 1).99\ldots = 1 - 0.99\ldots = d_p. \quad \textbf{(72)}$$

Since all the elements of d* share its properties then whenever we have a statement "an element d of d* has property P" we may write "d* has property P", meaning, this statement is true of every element of d*. This applies to any equation involving an element of d*. Therefore, we have,

$$d^* = N - (N - 1).99\ldots = 1 - 0.99\ldots. \quad \textbf{(73)}$$

Like a decimal, we define the d-norm of d* as d* and $d^* > 0$.

Theorem

The d-limits of the indefinitely receding to the right nth d-terms of d* is a continuum that coincides with the g-limits of the tail digits of the nonterminating decimals traced by those nth d-terms as the a_ks vary along the basic digits.

If x is nonzero decimal, terminating or nonterminating, there is no difference between $(0.1)^n$ and $x(0.1)^n$ as they become indistinguishably small, *i.e.*, as n increases indefinitely. This is analogous to the sandwich theorem of calculus that says, lim(x/sinx) = 1, as $x \to 0$; in the proof, it uses the fact that sinx < x < tanx or 1 < x/sinx < secx where both extremes tend to 1 so that the middle term tends to 1 also. In our case, if $0 < x < 1$, $0 < x(0.1)^n < (0.1)^n$ and both extremes tend to 0 so must the middle term and they become indistinguishably small as n increases indefinitely. If x > 1, we simply reverse the inequality and get the same conclusion. Therefore, we may write, $xd_p = d_p$ (where d_p is the principal element of d*) and since the elements of d* share this property we may write xd* = d*, meaning, that xd = d for every element d of d*. We consider d* the equivalence class of its elements. In the case of $x + (0.1)^n$ and x, we look at the nth g-terms of each and, as n increases indefinitely, $x + (0.1)^n$ and x become indistinguishable. Now, since $(0, 1)^n > ((0.1)^m)^n > 0$ and the extreme terms both tend to 0 as n increases indefinitely, so must the middle term tend to 0 so that they become indistinguishably small (the reason d* is called dark for being indistinguishable from 0 yet greater than 0). We summarize our discussion as follows: if x is not a decimal integer (a decimal integer has the form, x = N.99…, N = 0, 1, …) then,

x + d* = x; otherwise, if x = N.99…,

$x + d^* = N+1$, $x – d^* = x$; if $x \neq 0$, $xd^* = d^*$; $(d^*)^n = d^*$, n = 1, 2, …, N = 0, 1, …; 1 – d* = 0.99…;

N – (N – 1).99…; 1 – 0.99… = d*, N = 1, 2, …. **(74)**

It follows that the g-closure of **R**, *i.e.*, its closure in the g-norm, is **R*** which includes the additive and multiplicative inverses and d*. We also include in **R*** the upper bounds of the divergent sequences of terminating decimals and integers (a sequence is divergent if the nth terms are unbounded as n increases indefinitely, e.g., the sequence 9, 99, …) called unbounded number u* which is countably infinite since the set of sequences is. We follow the same convention for u*: whenever we have a statement "u has property P for every element u of u*" we can simply say "u* has property P). Then u* satisfies these dual properties: for all x,

$x + u^* = u^*$; for $x \neq 0$, $xu^* = u^*$. **(75)**

Neither d* nor u* is a decimal and their properties are solely determined by their sequences. Then d* and u* have the following dual or reciprocal properties and relationship:

$0d^* = 0$, $0/d^* = 0$, $0u^* = 0$, $0/u^* = 0$, $1/d^* = u^*$, $1/u^* = d^*$. **(76)**

Numbers like u* – u*, d*/d* and u*/u* are still indeterminate but indeterminacy is avoided by computation with the g- or d-terms.

It is clear that d* and u* are the counterparts of the infinitesimal and infinity of calculus; the only difference is that both d* and u* are well defined.

The decimals are linearly ordered by the lexicographic ordering "<" defined as follows: two elements of **R** are equal if corresponding digits are equal. Let

$N.a_1a_2..., M.b_1b_2... \in \mathbf{R}$. **(77)**

Then,

$N.a_1a_2... < M.b_1b_2$ if N < M or if N

= M, $a_1 < b_1$; if $a_1 = b_1$, $a < b_2$; …, **(78)**

and, if x is any decimal we have,

$0 < d^* < x < u^*$. **(79)**

The trichotomy axiom follows from lexicographic ordering. This is the natural ordering mathematicians sought among the real numbers but it does not exist there because it contradicts the trichotomy axiom.