2009

Dynamics, measure and dimension in the theory of computing

Satyadev Nandakumar
Iowa State University

Dynamics, measure and dimension in the theory of computing

by

Satyadev Nandakumar

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Computer Science

Program of Study Committee:
Jack H. Lutz, Major Professor
Pavan Aduri
Krishna Athreya
Roger Maddux
Giora Slutzki

Iowa State University

Ames, Iowa

2009

TABLE OF CONTENTS

Acknowledgments

My career has been decisively influenced by a few inspiring teachers, to whom I owe much. Mr. C. Balakrishnan first introduced me to the elegance of mathematical proofs. Dr. V. K. Govindan tried to impart to me the charm and immediacy of applied mathematics. Dr. Gary Leavens keeps the beacon of scholarship lit for students to follow. I would like to thank them, and I hope to imitate them in my academic career. From the beginning, Dr. Jack Lutz treated me as a peer when I hardly merited the status. He has given me problems which have been enlightening to work on. That my graduate life has been fulfilling, is due to his uncompromising standards. He taught me, foremost, to work out the concrete details fully. Along the way, I had great fellow-students in the complexity group at Iowa State - Dave Doty and Xiaoyang Gu have been insightful critics and energetic colleagues. Though our fields are largely distinct, Matt Patitz, and Scott Summers have lent a patient ear and generous company. Dr. Pavan Aduri, Dr. Shashi Gadia, Dr. Gurpur Prabhu, and Dr. Giora Slutzki have supported and encouraged me throughout my stay at Iowa State. I am indebted to them for their kindness.

I am fortunate for the bevy of my friends I am a part of – I would like to thank them immensely, especially Aakriti, Anu, Anumodh, Srini Neginhal, Vatsan, and Allada for keeping my spirits buoyant, for their pleasant company, for memorable trips, and for the needlessly frequent coffee breaks through the years of my stay in Ames :) Many idyllic summers and bleak winters later, if I have finally managed to pull through, it is because they have played an invaluable part. Their friendship is a treasured keepsake.

Above all, I thank my parents and my sister for their love, support, humor, and patience. I can never repay what they have done for me, but I hope my work is a fair token. I dedicate my thesis to them with fond gratitude.

LIST OF FIGURES

ABSTRACT

This thesis establishes some results in algorithmic information theory. Martin-Löf in 1966 formalized the notion of what it means for a given sequence to be random, using the theory of computation. Equivalent, alternate definitions justify that this definition is robust. A finite string is called random if there is no short program which outputs the string - that is, the Kolmogorov complexity of the string is within an additive constant of the length of the string itself. With the introduction of a suitable notion termed self-delimiting Kolmogorov complexity, an infinite sequence is called random if almost all its finite prefixes are incompressible in the above sense. Algorithmic information theory studies the properties of such random sequences.

It is natural to enquire whether a sequence random according to this definition satisfies the well-known probabilistic laws. Early results by van Lambalgen and Vovk established that all random sequences obey Borel-Cantelli Lemmas, the Strong Law of Large Numbers and the Law of the Iterated Logarithm. van Lambalgen conjectured that Birkhoff's ergodic theorem, however, will resist such effectivization. Nevertheless, V'yugin in a tour-de-force established an effective version of the Ergodic Theorem. V'yugin's version has a technical limitation, however, that restricts its applicability. Random variables are assumed to be computable, and hence continuous. In the first part of the thesis, we extend V'yugin's theorem to handle a restricted class of discontinuous functions, and prove that we can apply it to effectivize some results on continued fractions.

The concept of dimension gives a quantitative measurement of how non-random sequences or sets are. In subsequent chapters we deal with the concept of dimension, in various resource-bounded settings. We give an alternate characterization of constructive Hausdorff and constructive packing dimension using generalizations of tools used in establishing the effective

ergodic theorem.

The final two chapters deal with finite-state dimension, which can be seen as density of information as measured by finite-state machines. This ties to a classical notion of normality of numbers. We prove that multiplication by a rational number does not alter the finite-state dimension of the base-k expansion of a real number, generalizing and giving a new proof of Wall's classical theorem that multiplication by a non-zero rational does not alter the normality of a number.

In the last chapter, we give constructions of normal and non-normal Liouville numbers, a well-known class of transcendental numbers.

CHAPTER 1. Introduction

The subject of this thesis is algorithmic information theory, in an abstract setting as well as its application to the study of randomness of sequences and real numbers. In particular, we consider effective versions of limit laws in probability theory, and apply the theory of effective dimension to study properties of real numbers.

1.1 Motivation

Our most intuitive conception of randomness is irregularity. We often consider a sequence of 10 consecutive zeroes as a fairly simple sequence, where as the outcome of 10 tosses of a fair coin could produce strings like "0101011101"with no discernible pattern. Informally, we call the latter string to be "random".

We also expect most strings of data to be random. If we locate three stars on the celestial sphere and measure the angles of the triangle formed, we would consider a triangle with three right angles to be a surprising event. What we expect is the *lack* of a pattern. The presence of a pattern is considered rare, and would provoke an enquiry into its cause and repeatability.

In the practice of modern probability theory, however, we discard such primitive notions. We do not enquire into the origin or cause of probability – We take the notion of probability as given, and base the theory on self-evident axioms and study its consequences. The most influential axiomatization of probability is that of Kolmogorov [38], based on Lebesgue's measure theory. This axiomatization has proved to be remarkably productive, and led to the flourishing of probability theory in the last century.

However, there were alternate conceptions of the theory of probability that were prevalent at the time of Kolmogorov's axiomatization. One that particularly concerns us in this thesis, is that of the Austrian probability theorist Richard von Mises [80]. von Mises insisted that the concept of a random sequence was the primitive, and the concept of probability a consequence of this. Thus probability is not the given concept, but irregularity of data. He carefully developed the theory based on this notion, and argued that this approach was closer to the application of probability theory. Kolmogorov acknowledges this viewpoint in his book

> "There are other postulational systems of the theory of probability, particularly those in which the the concept of probability is not treated as one of the basic concepts, but is itself expressed by means of other concepts. However, in that case, the aim is different, namely, to tie up as close as possible the mathematical theory with the empirical development of the theory of probability." [38]

However, with the success of the measure-theoretic approach, the debate on the relevance of alternate formulations was quelled for a time. It was Kolmogorov himself who revived interest in this approach in the 1960s [39], [40]. He investigated the use of computability theory to formalize notions similar to that of von Mises. These efforts culminated in Martin-Löf's definition of a sequence random with respect to a given probability measure [60] , and to the concept of Kolmogorov complexity [74], [41], [14]. We discuss these definitions and their similarity to the notion of von Mises', in the next chapter.

Kolmogorov outlined a program for developing information theory along combinatorial lines independent of the notion of probability [42]. Indeed, he states:

> "Information Theory must precede probability theory, and not be based on it."

To this end, he proposed a combinatorial basis for information theory, for finite strings. He also outlined a program for the study of infinite sequences. Here, he made the remark that the theory of infinite random sequences may not have a bearing on the basis of probability, but investigating their properties in the light of the theory of randomness would bring out the algorithmic content of the theorems in probability.

This program was taken up by researchers like Martin-Löf[60], and Schnorr [70]. In particular, van Lambalgen's [78] thorough thesis proves effective versions of probabilistic laws like the Borel-Cantelli lemmas and the strong law of large numbers. This thesis takes this effort further.

1.2 Summary of Results

The thesis is organized as follows. Chapter 2 elaborates on some of the aspects of Kolmogorov's program to base information theory on computability theory. This chapter is expository, and does not contain new results. Chapter 3 gives the proof of an effective version of Birkhoff's ergodic theorem that deals with a class of discontinuous but computably approximable random variables. We then apply it to a specific dynamical system, that of continued fractions. The tool that we use to prove the effective ergodic theorem is a *measure of impossibility* [82]. In Chapter 4 we generalize this tool and show that we can characterize constructive Hausdorff and packing dimension with this generalization.

Chapters 5 and 6 investigate finite-state dimension, a particular specialization of Hausdorff-dimension. It generalizes of the classical concept of normality, and is a useful tool in the study of normality.

In Chapter 5, we use Finite-State dimension and Schur concavity to prove that multiplication by non-zero rationals does not alter the finite-state compressibility of base-k expansions of reals. This generalizes a classical theorem that multiplication by rationals does not alter the normality of a number. In the concluding chapter, we show that there are normal as well as non-normal numbers inside a well-known class of transcendental numbers.

CHAPTER 2. Kolmogorov's Program

In this expository chapter, we give a brief introduction to von Mises' definition of a random "collective" and Kolmogorov's definition of place selection inspired by this notion. Then we introduce Martin-Löf's definition of a sequence random with respect to a computable probability distribution, and give a brief overview of the Kolmogorov program. This provides the setting for the later chapters.

2.1 von Mises' Collectives

In von Mises' interpretation of probability, he adopted a strict form of a viewpoint called frequentism. His notion of a random sequence, called a *collective*, had the property of having stable limiting frequencies for their constituent alphabet symbols.

Let Σ be a discrete alphabet (not necessarily finite), and ω be an infinite sequence formed from that alphabet. Let n_i denote the frequency of the i^{th} symbol in the alphabet in the first n symbols of ω. von Mises' "first basic assumption" [80] is that a collective is an infinite sequence and the frequency of $\frac{n_i}{n}$ approaches a limit as n tends to infinity.

A binary sequence like "010101 . . . " satisfies the above assumption, but is very regular. To avoid such sequences, von Mises introduced the notion of "place selection".

The notion of a place selection on a collective was carefully described by von Mises.

> "A subsequence has been derived by place selection if the decision to retain or reject the n^{th} element of the original sequence depends on the number n and the labels (alphabet symbols) $x_1, x_2, \ldots, x_{n-1}$ of the $(n-1)$ preceding elements, and not on the n^{th} or any following element. The rule must be such that the selected subsequence is infinite." [80]

An allowed place selection rule is one where, for example, we choose all n such that n is a prime number, and $x_{n-1} = 0$.

A disallowed selection rule would be, "Choose i such that $x_i = 0$." This rule is guaranteed to select a sequence of all zeroes if there are infinitely many zeroes in x. The stated rule is disallowed, because it looks at the value at index i to retain or discard x_i. Every infinite sequence with an infinite number of 0s in it will have a subsequence of all zeroes, but it may be impossible for an admissible place selection rule to choose it. It may happen by accident that a given admissible selection rule chooses a subsequence consisting of zeroes alone, but the original sequence had positive limiting frequencies for all the alphabet symbols. In this case, the original sequence was not a collective.

von Mises then insisted that in probability theory, we shall deal (not exclusively) with particular sequences which satisfy the first basic assumption, for which it may be assumed that the limiting frequency of any symbol is invariant under place selections. If this holds true, then the limiting frequency will also be called a probability.

Definition. [80] Let $S = \{a_i\}$ be a discrete alphabet and $K = \{x_j\}$ be a sequence of elements of S. Let G be a system of denumberably many place selections. We assume that

- For all a_i, the limiting frequency p_i exists in K.
- $\sum_{a_i \in S} p_i = 1$, the sum extended over all elements in S.
- Any place selection Γ belonging to G applied to K produces an infinite subsequence of K in which, for all a_i, limiting frequency exists and is equal to p_i.

Then K, or more explicitly, $K(G, S)$ is called a *collective* and $p(a_i) = p_i$, the probability of a_i.

Thus collective is the assumed notion which abstracts some intuitive notions of randomness, and probability is defined based on collectives.

He then described four transformations on collectives - namely, place selection, mixing, partition and combining - and showed how if the input to the transformation is a collective, then so is its output. He showed how limit laws like the strong law of large numbers could be derived based on these transformations. Details of this can be found in either von Mises' textbook [80] or in van Lambalgen's thesis [78].

von Mises notably did not give an *a priori* notion of what was an admissible selection rule. Church in 1940 [16] tried to formalize this using computability theory. The notion that von Mises tried to describe is usually called a *stochastic sequence*, in the literature on algorithmic information theory. We have the following definition for binary sequences.

Definition. A sequence S is said to be *Stochastic* with respect to a given set of admissible selection rules if for any admissible selection rule, the sequence of bits selected from S is either finite, or the limiting frequency of 1 in the sequence is 1/2.

We will refer to this notion in the following section.

von Mises' definition was not widely accepted. The crucial attack came from Ville [79] who showed that there are binary collectives with $p_0 = p_1 = 1/2$ in von Mises' sense for which for all n,

$$\frac{1}{n}\sum_{i=1}^{n} x_i \geq 1/2,$$

implying that the frequency of ones converges to its limit strictly from above.

Ville's construction gives a collective x which violates the law of iterated logarithm in probability theory. This was seen as a death blow to the theory and it gradually became unpopular. He employed the more powerful notion of martingales, instead of place selection rules, to construct this collective. He characterized measure zero sets using martingales. This in fact proved that any probabilistic law can be characterized in this way. Martingales later proved to be an important tool in the theory.

2.2 Kolmogorov-Loveland Place Selection Rules

Kolmogorov revived the debate on the foundations of probability in a series of papers published in the 1960s. This eventually ended in the successful definition of randomness for finite strings through the notion of Kolmogorov complexity. In [39], he wrote:

> "The basis for the applicability of the results of the mathematical theory of probability to real 'random phenomena' must depend on some form of the frequency concept of probability, the unavoidable nature of which has been established by von Mises in a spirited manner."

Church considered only monotonic selection rules: once an algorithm made a decision on the admission of an i into a subsequence, it could not revert its decision at a later stage. Kolmogorov [39], [40] and Loveland[49], [48],initially independently proposed to generalize this notion by allowing some partial computable place selection functions that are not necessarily monotonic. If a sequence is stochastic under such place selection rules, then we call it a Kolmogorov-Loveland stochastic sequence.

Definition. A *selection rule* is a partial function $s : \{0,1\}^* \to \mathbb{N} \times \{0,1\}$ which maps a string w to the pair $(x(w), b(w))$ such that whenever $s(w)$ is defined, then for all proper prefixes v of w, the value $s(v)$ is defined, and $x(w)$ differs from $x(v)$.

Let T be a sequence which is acted on by s. The *sequence of scanned places* $x_0, x_1, \dots$ is defined inductively by

$$\begin{aligned} x_0(\lambda) =& x(\lambda) \\ x_{i+1} =& x(T(x_0)T(x_1)\dots T(n_i)). \end{aligned}$$

The *sequence of selected places* $z_0, z_1, \dots$ is the subsequence of scanned places that is obtained by canceling all elements where the corresponding value of b is equal to 0.

The sequence selected by s from T is $T(z_0)T(z_1)\dots$.

A selection rule is called *monotonic* if for all sequences, the sequence of scanned places is strictly ascending.

However, even this notion runs into problems. For example, Merkle [61] proved that there is a Kolmogorov-Loveland stochastic sequence S and even a computable monotonic selection rule which picks a subsequence T which is not Kolmogorov-Loveland stochastic. Thus T has a limiting frequency of 1 equal to 1/2, but further subsequence selection from T according to Kolmogorov-Loveland selection rules will lead to a non-stochastic sequence.

2.3 Martin-Löf's definition of a random sequence

The first successful definition of the notion of an *infinite random sequence* was given by Martin-Löf [60] who was a doctoral student of Kolmogorov. Martin-Löf's idea is based on constructive measure theory, and he defined a notion based on the definition of an effective null set.

We consider a sequence to be random if it satisfies all probabilistic laws with respect to that measure. So the set of random sequences should be the intersection of all properties of probability 1. However, this set is empty. So one way to work on this notion would be to restrict which probabilistic laws we consider, so that "most" strings are still random, and Martin-Löf followed this approach.

Every probabilistic law says that a property holds with probability 1. Thus it characterizes a measure 1 set, and such a set has a complement set of measure 0. It is hence sufficient to define the concept of an *effective* measure zero set.

Let P be a computable probability measure defined on the $\{0,1\}^\infty$. (We will define the notion precisely in Chapter 3.) For finite strings x, we consider cylinders C_x, the set of all infinite sequences with x as a prefix. A set S of sequences from the sample space of all sequences has P-measure zero if, for each $\varepsilon > 0$, there is a sequence of cylinders $C_{x_0}, C_{x_1}, \dots, C_{x_i}, \dots$ of cylinder sets such that

$$S \subseteq \cup_i C_{x_i} \text{ and } P(\cup_i C_{x_i}) < \varepsilon.$$

A set of sequences S has effective P-measure zero if there is a computable function $h(i, \varepsilon)$ such that $h(i, \varepsilon) = C_{x_i}$ for each i. Martin-Löf proved a universality property - that for every computable probability measure$P$, there is a unique largest effective P-measure zero set. The complement of this set is called the set of constructive random sequences wrt P.

Martin-Löf's definition stands von Mises' viewpoint on its head - randomness is defined in terms of the probability, and not as as the primitive. However, this notion proved to be successful in practice – not only did the random sequences obey limit laws in probability theory (with appropriate computability assumptions on the probability space), but it also proved robust: many seemingly approaches, based on different principles, of defining a random sequence defined exactly the same set of random sequences as Martin-Löf's. For example, recall that Ville characterized measure zero sets using martingales. Effective martingales, defined in a natural way, capture exactly the effective measure-zero sets of Martin-Löf [68], [69], [70]. In this thesis, we work with some of these such tools.

2.4 Kolmogorov Complexity and Kolmogorov's Program

The goal of both Kolmogorov and von Mises was to define a satisfactory notion of a random finite string. This was achieved using the notion of Kolmogorov Complexity. In this thesis, we do not use this notion much, but we add this to make the presentation self-contained, and to state Kolmogorov's program on founding information theory on the theory of computing.

2.4.1 Finite Strings

There is an apparent paradox in trying to settle the notion of irregularity for a finite data set, or a finite string: It is always possible to produce a (deterministic) law that fits the data at all the observed points. However, the crucial point is that this law need not be *simple.* Kolmogorov writes:

> Any results of observations can be registered in the form of a finite, though sometimes very long, entry. Therefore, when we speak of a lack of regularity in observational results, we have in mind merely the absence of a *fairly simple* regularity. [42]

Kolmogorov gives a specific illustration:

> For example, if we are given a polynomial of degree 999 whose values for $x = 1, 2, 3, \ldots, 1000$ yield a sequence of integers $p(x)$ between 0 and 9, obtained as a result of an honest random experiment like roulette play, then the presence of such a polynomial does not prevent us from continuing to regard the sequence as random.[42]

We have not achieved any economy of expression with the above law. The notion of Kolmogorov complexity, independently defined by Solomonoff [74], Kolmogorov [41] and Chaitin [14] is based on generalizing this insight.

We use the notion of an algorithm in the theory of computing, to give this definition. We fix any standard programming model.

Definition. The Kolmogorov Complexity of a finite string x is the length of the shortest program which prints x.

It is immediate from the presence of Universal Turing Machines, that is a program capable of running any other program(commonly called an "interpreter"), that this is an intrinsic property of the string (up to an additive constant), and not of the model of machines we choose. The theory is detailed in [47].

However, in the applications of probability theory, we do not merely study the lack of regularity, "but from the hypothesis of randomness of the observed phenomena, we draw definite positive conclusions."[42]

Kolmogorov outlines his program thus. von Mises' work substantiates the hypothesis that the stability of frequencies is often put as the basis of the *application* of probability theory. Given, for example, an n-bit string of 0s and 1s obtained as a result of independent tosses of a coin with probabilities p and q, the complexity of the string cannot be much more than

$$nH(p) = n\left(p \log p - (1-p)\log(1-p)\right),$$

where H is the Shannon entropy. Kolmogorov's initial paper [39] proves that the stability of frequencies in the sense of von Mises is automatically ensured if the complexity of the n-bit string is close to the above upper bound.

He then postulates that this claim is more general. For example, if the complexity of a given n-bit string with transition frequencies is close to the maximum complexity of a sequence of length n produced according to a Markov chain with that transition probabilities, then the conclusions of the probabilistic theory of Markov chains applies to the string.

He then proposes that

> "The application of probability theory can be put on a uniform basis. It is always a matter of hypotheses about the impossibility of reducing in one way or another the complexity of the description of objects in question." [39]

We call this Kolmogorov's program.

2.4.2 Infinite Sequences

Separate from the program, but related to it, Kolmogorov also recommends investigating the concept of infinite sequences. In his concluding remarks, Kolmogorov adds that the concepts of information theory as applied to infinite sequences, give rise to questions that have value in the investigation of the algorithmic side of mathematics, even if it has no bearing on the basis of probability theory.

Questions of this nature form the basis of this thesis.

CHAPTER 3. An Effective Ergodic Theorem and Some Applications

3.1 Introduction

In the context of Kolmogorov's program to base the theory of probability on the theory of computing, an early achievement was Martin-Löf's work establishing that there is a unique smallest constructive measure 1 set whose objects are individual random sequences [60]. In this program, we formulate probabilistic laws, i.e., laws of the form "Probability[$\{\omega : A(\omega)$ holds $\}] = 1$" for some property A, in their effective form, "If ω is random, then $A(\omega)$ holds."[1]

It is not known whether all such laws can be converted into this form: early work by Vovk on the Law of Iterated Logarithm [81] and van Lambalgen on the Strong Law of Large Numbers [78] were successes; but, it was conjectured that not all laws can be converted into the effective form. In particular, it was conjectured that two laws, the Ergodic Theorem of Birkhoff [6] and the Shannon-McMillan-Breiman Theorem [5] resist effectivization. Nevertheless, V'yugin in [83] converted a proof of a constructive version of the Ergodic Theorem by Bishop [7] to prove an effective version of the Ergodic Theorem.

The ergodic property is a weak form of independence obeyed by stochastic processes. If P is a finite measure, f is an integrable function and T is a transformation preserving the measure P, then Birkhoff's ergodic theorem states that the limit

$$\lim_{n\to\infty} \frac{1}{n} \left(f(\omega) + f(T\omega) + \cdots + f(T^{n-1}\omega) \right) \tag{3.1.1}$$

exists for almost all ω in the sample space (for instance, see [5]). Moreover, if T is ergodic (definitions in section 2), then the above said average is the same constant, $\int f dP$, almost

[1]This work appeared in [63].

everywhere. V'yugin's version establishes that if P is a computable measure, f is an integrable computable function and T is a computable measure-preserving transformation, the limit exists for all individual random ω. If T is ergodic, then the average (3.1.1) is the same constant for all individual random points. The convergence to the constant need not be effective - a computable function may not be able to predict the rate of convergence. [83]

We wish to explore applications of the effective version of the Ergodic Theorem in this paper. Classically, the Strong Law of Large Numbers can be proved to be a special case of the Ergodic Theorem. Moreover, the ergodic theory of continued fractions (see for example, Kraaikamp and Dajani [18]) provides some examples of non-trivial applications of the ergodic theorem to the metric theory of numbers. The celebrated theorems of Lévy-Kuzmin, and Khinchin, are examples. We would like to form effective versions of these theorems. These are known to hold effectively, and the proofs employ transfer operators [66].

Classically, the proofs of these properties fall out of the ergodic theorem. The lack of an effective ergodic theorem has hindered the proofs in their classical form being used to establish the theorems in their effective form. We find that V'yugin's version cannot be used for this purpose because of a technical limitation - computable functions are continuous, while most of the proofs employ functions which are discontinuous.

A recent work by Braverman [11] suggests a way of handling computability of discontinuous functions using a notion termed "graph-computability". Graph–computability cannot be adopted without modifications to prove the ergodic theorem. Indeed, we exhibit a graph–computable function for which the Constructive Ergodic Theorem fails. However, with suitable restrictions on the class of graph computable functions, we can prove a constructive version of the theorem. Our aim in this chapter is threefold - to prove a version of the effective ergodic theorem which handles discontinuous functions, to prove the effective strong law of large numbers as a consequence of the effective ergodic theorem, and to give new proofs of some classical results in continued fractions in their effective form, using the above.

3.2 Preliminaries

As usual, $\mathbb{R}$ denotes the set of real numbers, $\mathbb{N}$ denotes the set of natural numbers, and $\mathbb{Q}$ denotes the set of rationals. We denote the positive part of set Q by Q^+ and the negative part by Q^-. The notation $0^{\mathbb{N}}$ represents the unary notation of natural numbers. Now, to define the sample space, we consider an alphabet. Let Σ be the binary alphabet $\{0, 1\}$. We consider finite words over the alphabet, denoted by Σ^*, and infinite sequences over the alphabet, denoted by Σ^∞. For positive integer i, the i^{th} position of sequence or word ω is denoted as ω_i. The substring $\omega_i \dots \omega_{j-1}$ is denoted $\omega[i \dots j-1]$. For a word ω, the length of $\omega = \omega_0\omega_1\omega_2 \dots \omega_{n-1}$ is denoted as $|\omega|$ with value n. If x is a string and w is a word or a sequence, the symbol $x \sqsubseteq \omega$ denotes that x is a prefix of ω.

3.2.1 Basic Concepts in Ergodic Theory

Let $(\Omega, \mathcal{F}, P)$ denote a probability space, where $\Omega = \Sigma^\infty$ is the sample space, $\mathcal{F}$ denotes the Borel $\sigma-$ algebra generated by the cylinders $C_x = \{\omega \mid \omega \in \Omega, x \sqsubseteq \omega\}$, and $P : \mathcal{F} \to [0, 1]$ is the probability measure.

We introduce some basic concepts from Ergodic Theory.

Let $T : \Omega \to \Omega$ be a transformation, *i.e.*, a measurable function from Ω to itself. In particular we consider the case when the transformation T is a *measure-preserving transformation* with respect to the probability space $(\Omega, \mathcal{F}, P)$. That is, for every measurable set A, we have $P[T^{-1}A] = P[A]$. If, moreover, $T^{-1}A = A$ for only measure 0 and measure 1 sets, then T is called *ergodic*.

The intuition of a measure-preserving transformation is that T preserves the volume of every measurable set. Ergodicity may be understood by thinking about the contrapositive of the definition: if a measurable set A has measure neither 0 nor 1, then TA is necessarily different from A. This tries to capture the notion that T is a "thorough" stirring.

Ergodic systems are weakly independent systems. For further details, see, for instance, Walters [85].

Successive applications of T are denoted as follows: $T^0\omega = \omega$ and, for all n, $T^{n+1}\omega =$

$T(T^n\omega)$. Customarily, the set $\{T^n\omega : n \in \mathbb{N}\}$ is called the *orbit* of ω under T.

A *dynamical system* is a system $(\Omega, \mathcal{F}, P, T)$ where $(\Omega,\ \mathcal{F}, P)$ is the probability space, and $T : \Omega \to \Omega$ is the measure-preserving transformation (which need not be ergodic). Two examples of dynamical systems are given below.

1. The probability space $(\Omega, \mathcal{F}, \mu)$ where Ω is the set of binary sequences, $\mathcal{F}$ is the Borel σ-algebra generated by the cylinders $C_x = \{\omega \in \Omega \mid x \sqsubseteq \omega\}$, and μ is the uniform probability measure, Lebesgue measure. The transformation $T : \Omega \to \Omega$ is the left-shift transformation,

$$T[\omega_1\omega_2\omega_3\omega_4 \dots] = \omega_2\omega_3\omega_4 \dots,$$

identified with the numerical function

$$T\omega = 2\omega \mod 1;$$

T is seen to be measure-preserving and ergodic with respect to the probability space.

2. The probability space $(\mathbb{N}^{+\infty}, \mathcal{F}', \gamma)$ where $\mathbb{N}^{+\infty}$ is the set of positive integer sequences identified with the continued fraction expansion of reals in $(0, 1)$. (see section 6), $\mathcal{F}$ is the Borel σ-algebra generated by the cylinders $C_x = \{\omega \in \mathbb{N}^{+\infty} \mid x \sqsubseteq \omega\}$, and γ is the Gauss measure, *i.e.* for any measurable set A, the probability $P[A] = \int_A \frac{1}{1+x}dx$. The transformation T is the left-shift transformation, as above, now identified with the numerical function $T\omega = \frac{1}{\omega} \mod 1$. The Gauss measure is important, since T is measure-preserving with respect to the Gauss measure, but not with respect to the uniform measure. T is also ergodic with respect to the probability space. For details, see Billingsley [5].

3.2.2 Algorithms, Graph-Computability

We also consider algorithms which map finite objects to finite objects - for instance, of the type $(\mathbb{N} \to \mathbb{N})$, $(\mathbb{Q} \to \mathbb{Q})$ and $(\Sigma^* \to \mathbb{N})$. An element of $\mathbb{N}$, $\mathbb{Q}$ or Σ^* is a finite object. Any finite object is computable. A real number r is computable if there exists an algorithm $f : 0^{\mathbb{N}} \to \mathbb{Q}$ such that for any integer n presented in unary, $f(0^n)$ is a rational q such that $|r - q| \leq 2^{-n}$. The computable function f is called an *computability witness* for r. For convenience, we fix an encoding of a finite object as an element of Σ^*. Further, we define the following notions in computable analysis(Weihrauch [86]).

A function $f : \Omega \to [-\infty, \infty]$ is said to be *lower semicomputable* if the set

$$G_l = \{(w, q) \mid w \in \Omega, q \in \mathbb{Q}, q < f(\omega)\}$$

is the union of a computably enumerable sequence of cylinders in the natural topology on $\Omega \times \mathbb{Q}$ or $\Sigma^* \times \mathbb{Q}$. The natural topology on $\Sigma^* \times \mathbb{Q}$ is the discrete topology. The natural topology on $\Omega \times \mathbb{Q}$ is the topology generated by the cylinders of the form (x, q) where $x \in \Sigma^*$ and $q \in \mathbb{Q}$.

Analogously, the function f is said to be *upper semicomputable* if $-f$ is lower semicomputable. A function f is said to be computable if it is both lower and upper semicomputable. Equivalently, we can show that a real-valued f is computable if and only if there is a Turing machine M such that for every real r, if $\hat{r} : 0^N \to Q$ is a valid computability witness for r, then we have for all n, $|M^{\hat{r}}(0^n) - f(r)| < 2^{-n}$, where $M^{\hat{r}}$ is the machine M with oracle access to $\hat{r}$. It follows that every real-valued computable function on a bounded domain is necessarily continuous.

This introduces a problem: Many of the functions which come up in Ergodic proofs are not computable because they are not continuous. Graph–computability, introduced in Braverman [11], gives a framework for discussing computability of not necessarily continuous functions.

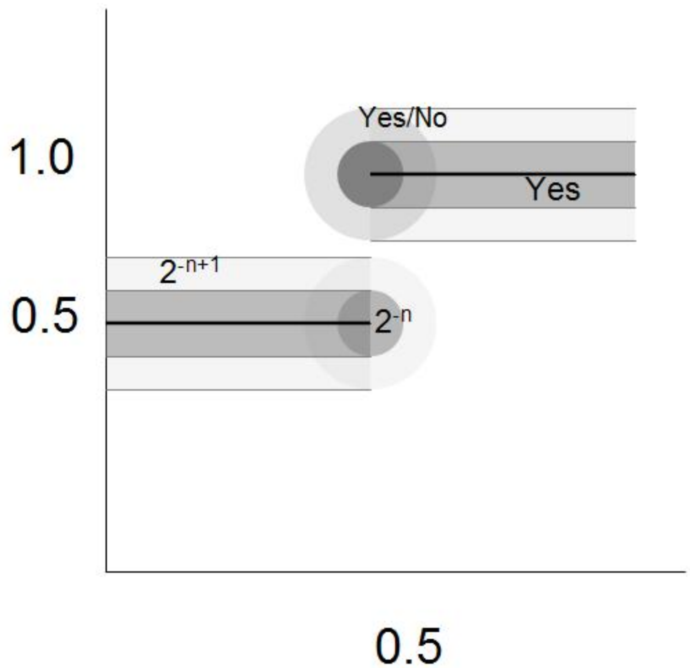
Definition ([11])**.** We say that a bounded subset S of $\mathbb{R}^n$ is bit-computable if there exists a computable function $f' : D^n \times \mathbb{N} \to \{0, 1\}$ such that

$$f'(d, 0^n) \quad = \begin{cases} 0 & \text{if } B(d, 2.2^{-n}) \cap S = \emptyset \\ 1 & \text{if } B(d, 2^{-n}) \text{ intersects } S \\ 0 \text{ or } 1 & \text{otherwise,} \end{cases}$$

where the neighborhood $B(d, r)$ represents the ball centered around d with radius r. A bounded real function on a bounded domain $D \subseteq \mathbb{R}$ is said to be *graph-computable* if the graph of $f = \{(x, f(x)) : x \in D\}$ is bit-computable as a set.

Every computable function over a bounded domain is graph computable. In addition to this, some step functions, which were not computable according to the definition above, are now shown to be graph-computable. For example, the step function on the unit interval ($f(x) =$ if $x > 0.5$ then 1, else 0.5) is graph-computable.

Figure 3.2.1 A graph computable function.



Now, we discuss computable transformations. We follow the definition in [83].

Transformations mapping Ω to itself are viewed as operating on the sequences themselves; the left-shift transformation is a simple case in point. Informally, a computable transformation is one which can be computed by an algorithm bit-by-bit. Formally, a *computable transformation* $T : \Omega \to \Omega$ is a transformation for which there is an algorithm which enumerates the set $\text{PREFIX}_T = \{(x, y) \mid x, y \in \Sigma^*\}$, such that

1. $(x, \lambda) \in \text{PREFIX}_T$, where λ is the empty string.

2. $(x, y) \in \text{PREFIX}_T \Rightarrow \forall x \sqsubseteq x', y' \sqsubseteq y, (x', y') \in \text{PREFIX}_T$.

3. $(x, y)$and$(x, y') \in \text{PREFIX}_T$ implies $y \sqsubseteq y'$ or $y' \sqsubseteq y$.

The transformation T is defined as

$$T\omega = \sup\{y \mid (x, y) \text{ such that } x \sqsubseteq \omega\}.$$

For example, the mappings $T_1, T_2 : \Omega \to \Omega$ defined as follows. The transformation $T_1(\omega) = \lambda$ is one that maps every sequence to the empty string. The transformation $T_2(\omega_1\omega_2 \ldots) = (\omega_2 \ldots)$ is the left-shift transformation. Both are computable transformations, though only T_2 is measure-preserving with respect to the Lebesgue measure on $[0, 1)$.

A *computable probability measure* P is one for which for every string $x \in \Sigma^*$, $P(x) = P(C_x)$ is computable.

3.3 Constructive Randomness

In this section, we define the notion of a constructively random sequence, and introduce the notion of a measure of impossibility, which we use to prove that a computable measure-preserving transformation conserves the randomness of a sequence.

Recall the discussion in Chapter 2 on the Martin-Löf definition of randomness. Let P be a computable probability measure defined on the $\{0, 1\}^\infty$. For finite strings x, we consider cylinders C_x, the set of all infinite sequences with x as a prefix. A set S of sequences from the sample space of all sequences has P-measure zero if, for each $\varepsilon > 0$, there is a sequence of cylinders $C_{x_0}, C_{x_1}, \ldots, C_{x_i}, \ldots$ of cylinder sets such that

$$S \subseteq \cup_i C_{x_i} \text{ and } P(\cup_i C_{x_i}) < \varepsilon.$$

A set of sequences S has effective P-measure zero if there is a computable function $h(i, \varepsilon)$ such that $h(i, \varepsilon) = C_{x_i}$ for each i. Martin-Löf proved a universality property - that for every computable probability measure$P$, there is a unique largest effective P-measure zero set. The complement of this set is called the set of constructive random sequences wrt P.

Another tool to study randomness is the concept of a measure of impossibility [82].

Gács in [25] extends the notion of Martin-Löf randomness to some non-compact spaces, one which he characterizes as spaces having recognizable boolean inclusions. We take the characterization and note that it works for Cantor Space, the space of infinite binary sequences, and Baire space, the space of infinite sequences of natural numbers.

Definition. A function $p : \Omega \to \mathbb{R}^+ \cup \{\infty\}$ is called a *measure of impossibility* with respect to the probability space $(\Omega, \mathcal{F}, P)$ if p is lower semicomputable and $\int p dP \leq 1$.

A *measure of impossibility* p of ω with respect to the computable probability distribution P denotes whether ω is random with respect to the given probability distribution or not. In particular, we can see that $p(\omega) < \infty$ if ω is random with respect to the computable probability distribution P [82], [25].

We now use this tool to give a proof of the fact that a computable, measure-preserving transformation conserves randomness. This extends, and gives a new proof of, Shen's [73] result on Cantor Space that a measure-preserving transformation conserves individual randomness.

Lemma 3.3.1. *Let ω be a Martin-Löf random real in Baire space or Cantor Space. Then for any computable measure-preserving transformation T, $T\omega$ is also Martin-Löf random.*

Proof. Let $T\omega$ be non-random. By assumption, there is a measure of impossibility p such that $p(T\omega) = \infty$. We define a new function $p' : \Omega \to \mathbb{R}^+ \cup \{\infty\}$ by $p'(\chi) = p(T\chi)$. p' is lower semicomputable by the lower semicomputability of p and the computability of T. Also, $\int p' dP = \int p dP \leq 1$ by the measure conservation property of T. Thus p' is a measure of impossibility such that $p'(\omega) = \infty$. □

This lemma implies that no point in the orbit of a sequence random wrt a computable measure, will be computable. This will be used in section 4.

3.4 Main Result

We would like to prove the following:

Ideal Theorem *If $(\Omega, \mathcal{F}, P)$ is a probability space where $\Omega = \Sigma^\infty$, with Borel σ-algebra generated by C_x, $x \in \Sigma^*$ and P is a computable probability measure, then for any function $f : \Omega \to \mathbb{R}$ which is graph-computable, $f \in L^1 P$, for any computable transformation T, and for any random ω wrt P, the ergodic average converges to $\int f dP$.*

However, there are graph-computable functions for which the ergodic average does not converge to the mean of the function.

Example 1. We construct a function $f : \Omega \to [0, 1]$ which is graph-computable, but is such that the effective ergodic theorem fails to hold. Consider f defined as follows.

Consider the uniform probability space. Let ω be an arbitrary Martin-Löf random real, e.g., the halting probability in binary notation, and $Tx = 2x \mod 1$. Then ω is normal: For all $n \in \mathbb{N}$ and $x \in \{0, 1\}^n$,

$$\lim_{i\to\infty} \frac{|\{m : 0 < m + n < i \text{ and } \omega[m \ldots m + n - 1] = x\}|}{i} = 2^{-n}.$$

In particular, the orbit of ω is dense in the unit interval. Define, for all $j \in \mathbb{N}$, $f(T^j\omega) = 1$, and $f(x) = 0$ for all other x.

This function is graph computable because both the sets $\{x : f(x) = 0\}$ and $\{x : f(x) = 1\}$ are dense in $[0, 1]$. The function is graph computable with a witness $B((q_1, q_2), 2^{-n}) = 1$ if and only if $|q_2| < 2^{-n-1}$ or $|1 - q_2| < 2^{-n-1}$.

We notice $\lim_{n\to\infty} \frac{\sum_{m=0}^{n-1} f(T^m\omega)}{n} = 1$.

However, $\int f(x)dx = 0$, since $\{x : f(x) = 0\}$ is a measure 1 set, so the effective ergodic theorem fails to hold for ω. (End Example)

This example serves to prove that graph-computability needs restriction in order to serve our purpose. One of the problems of the above example is the presence of a dense set of discontinuities. We posit the following class of graph-computable functions.

Definition. Let $\mathcal{G}_P$ be the class of graph-computable functions f continuous almost everywhere wrt P, with the property that f has only simple discontinuities which form a nowhere dense (one-dimensional) bit-computable set.

Note that a nowhere dense bit-computable set can contain only computable points, and hence there are at most countably many discontinuities. Hence by Lemma 3, for any constructively random ω, no point on its orbit $\{T^m\omega : m \in \mathbb{N}\}$ is a point of discontinuity. $\mathcal{G}_P$ is a superset of the class of computable functions. It is also large enough to subsume useful discontinuous functions used in some proofs of the metric theory of numbers. We have the following.

Lemma 3.4.1. *Let $f : \mathbb{R} \to \mathbb{R}$ be a graph-computable function in $\mathcal{G}_P$. Then there is a computable function $\overline{f} : \mathbb{Q} \times \mathbb{N} \to \mathbb{Q}$ such that for every point of continuity $r \in \mathbb{R}$ of f, every computability witness $\hat{r}$ of r the following hold:*

1. *For all natural numbers n, $\overline{f}(\hat{r}, 0^n) > \overline{f}(\hat{r}, 0^{n+1}) > f(r)$.*

2. $\lim_{n\to\infty} \overline{f}(\hat{r}, 0^n) = f(r)$.

We sketch the proof of this lemma in Section 5. We call any function which satisfies the above property as *essentially upper semicomputable.* A function f on a bounded domain is called *essentially lower semicomputable* if $-f$ is essentially upper semicomputable. A function f which has both properties is called *essentially computable.*

Main Theorem. *Let $(\Omega, \mathcal{F}, P)$ be a probability space where $\Omega = \Sigma^\infty$, $\mathcal{F}$ is the Borel σ-field generated by the cylinders C_x, and P is a computable probability measure. If $T : \Omega \to \Omega$ is a computable measure-preserving transformation, then for every essentially computable $f : \Omega \to \mathbb{R}$, $f \in L^1[P] \cap \mathcal{G}_P$, there is an integrable function $\tilde{f}$ such that*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j \omega) = \tilde{f}(\omega), \tag{3.4.1}$$

for every ω random wrt P, with $\tilde{f}(T\omega) = \tilde{f}(\omega)$ and $\int f dP = \int \tilde{f} dP$. Moreover, if T is ergodic, the abovementioned limit is a constant, for all individual random ω, $\tilde{f}(\omega) = \int f\, dP$.

The idea of the proof is as follows. For a graph-computable function f which obeys the restriction mentioned above, we prove that it is easy to obtain an essential upper semicomputation except at points of discontinuity; moreover, by the same argument, there is also an essential lower semicomputation of the function f. Using these essential upper and lower semicomputations, an upcrossing function is defined, which behaves reasonably well at points of continuity - namely, it converges if and only if the ergodic sum at the point converges. This function is shown to be semicomputable from below. We bound the integral of the upcrossing function over the whole space (this is essentially due to V'yugin [83]), thus we prove that the upcrossing function we defined measure of impossibility.

The upcrossing function attains ∞ only if the ergodic average diverges at the given point. This would imply that the ergodic sum converges at every individual random point. Details follow in the next sections.

3.5 Proof of Main Theorem

We first prove that for every function in $\mathcal{G}_P$, there exists an essential upper semicomputation and an essential lower semicomputation - i.e., functions which are semicomputations except at the points of discontinuity.

Proof Sketch for Lemma 3.4.1. Let g be the computability witness for the set of discontinuities. If x is a continuity point and the discontinuities are nowhere dense, then there is a least n_1 such that $q + 2.2^{-n_1} < x - 2.2^{-n_1} < x + 2.2^{-n_1} < r - 2.2^{-n_1}$ for the discontinuities q, r closest to x, $r > q$. Thus, x can be detected to be a continuity point at precision $n_1 + 1$, using g as a witness. We define the upper semicomputation as follows.

For all $n > n_1$, if $B((d_1, d_2), 2^{-n})$ intersects the graph of the function and $|d_1 - x| < 2^{-n}$, then $d_2 + 2.2^{-n}$ is an upper approximation of $f(x)$. So, given 0^k, we see whether we can determine that x is a continuity point at precision k. If no, we output ∞. If yes, we enumerate balls of radius 2^{-k} until we find k distinct upper approximants to $f(x)$, and output their minimum.

It is routine to verify that this process is an essential upper semicomputation of f. $\square$

Similarly, we establish that there is an essential lower semicomputation $\underline{f}$ of f, which at points of continuity, approximates f from below, and approaches f in the limit.

Now, we proceed to the proof of the main result.

Proof of Main Theorem

Let f be the graph computable function, as given, with essential upper semicomputation $\overline{f}$ and essential lower semicomputation $\underline{f}$. Let ω be a random point. Define the following:

$$a(\omega, f, n) = \sum_{i=0}^{n-1} \left[\lim_{n\to\infty} \overline{f}(T^i\omega, 0^n) - \alpha \right] \tag{3.5.1}$$

$$b(\omega, f, n) = \sum_{i=0}^{n-1} \left[\lim_{n\to\infty} \underline{f}(T^i\omega, 0^n) - \beta \right]$$

for any rational numbers α and β, natural numbers n and f given above. The set of functions can be undefined if for some f, the value at a point is $\pm\infty$. We assume that the functions are bounded, so this is impossible. We define the sets A and A' as follows:

$$\omega \in A'(u, v) \equiv \sum_{i=0}^{u-1} \left[f(T^i\omega) - \alpha \right] < \sum_{i=0}^{v-1} \left[f(T^i\omega) - \beta \right] \tag{3.5.2}$$

$$\omega \in A(u, v) \equiv a(\omega, f, n) < b(\omega, f, n). \tag{3.5.3}$$

Let n be a non-negative integer. A sequence of integers $s = \{u_1, v_1, \ldots, u_n, v_n\}$ is called an n-admissible sequence if $-1 \leq u_1 < v_1 \leq u_2 < v_2 \leq \cdots \leq u_N < v_N \leq n$. We use $m_s = N$. We define $a(\omega, f, -1) = 0$. We define Bishop's upcrossing function.

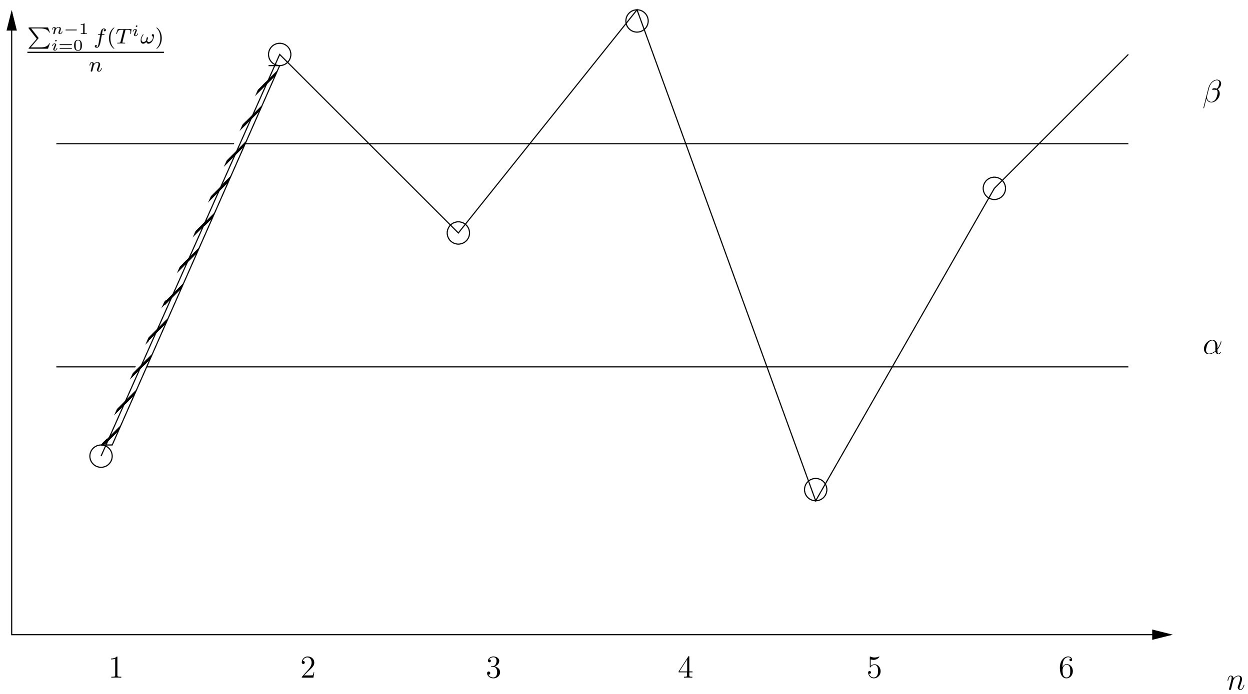
$$\begin{aligned} &\sigma'_n(\omega, \alpha, \beta) = \\ &\max(\{N \mid \omega \in \bigcap_{j=1}^{N} A'(u_j, v_j) \cap \bigcap_{j=1}^{N-1} A'(u_{j+1}, v_j) \cap \mathcal{C}_f \\ &\qquad \text{for some n-admissible } s = \{u_1, v_1, \ldots, u_N, v_N\}\} \cup \{0\}), \end{aligned} \tag{3.5.4}$$

We then define the function in its modified form:

$\sigma_n(\omega, \alpha, \beta)$ is defined analogous to σ'_n, with A replacing A' in the definition. It is easy to see that σ_n and σ'_n coincide on all points of continuity of f.

The function σ_n is lower semicomputable, since we can use the essential upper semicomputation $\overline{f}$ and the essential lower semicomputation $\underline{f}$ to compute σ_n, and at points of continuity ω, $\lim_{n\to\infty} \overline{f}(\omega, 0^n) = f(\omega) = \lim_{n\to\infty} \underline{f}(\omega, 0^n)$. The points of discontinuity are all non-random points. Define $\sigma' = \sup_n \sigma'_n$ and $\sigma = \sup_n \sigma$. Then σ is a lower semicomputable function.

Figure 3.5.1 Upcrossing Function: Only the shaded line is an upcrossing



We notice that $\sigma(\omega, \alpha, \beta)$ and $\sigma'(\omega, \alpha, \beta)$ differ only on a P-measure zero set. In particular, this means that

$$\int \sigma'(\omega, \alpha, \beta) dP = \int \sigma(\omega, \alpha, \beta) dP. \tag{3.5.5}$$

V'yugin proves that σ'_n is an integrable function. His proof assumes that f is computable and T is measure-preserving. We note that the proof that σ_n is integrable for graph-computable functions is identical to V'yugin's; it is reproduced here only to reflect the invariance in technical detail.

Since f is integrable, we have some constant M for which $\int (f(\omega) - \alpha)^+ dP \leq M + |\alpha|$ holds.

Lemma 3.5.1. [*V'yugin [83]*] *Let T be a measure-preserving transformation and $f : \Omega \to \mathbb{R}$ be an integrable, computable function. Then*

$$\int (M + |\alpha|)^{-1} (\beta - \alpha) \sigma(\omega, \alpha, \beta) \leq 1. \tag{3.5.6}$$

To define the measure of impossibility, we need to verify define an upcrossing function for every pair of rationals (q, r) where $q < r$. But we can enumerate all pairs of such rationals with a program, and let $\alpha(i)$ and $\beta(i)$ be the ith pair in this enumeration. The measure of impossibility is

$$p(\omega) = \frac{1}{2}\sum_{i=1}^{\infty} i^{-2}(M + |\alpha(i)|)^{-1}(\beta(i) - (\alpha(i))\sigma(\omega, \alpha(i), \beta(i)),$$

as in V'yugin [83]. It is easy to see that p is lower semicomputable and integrable.

This function attains ∞ only if the ergodic average diverges. It follows that for all individual random ω, the ergodic average converges.

Besides, $\tilde{f}(\omega) = \tilde{f}(T\omega)$ for any ω random wrt P, so $\tilde{f}$ is bounded almost everywhere P, and is therefore, integrable. Integrating (3.4.1) on both sides, by the measure preservation of T and the Dominated Convergence Theorem, we get $\int f dP = \int \tilde{f} dP$.

To see that for every individual random ω, the ergodic average converges to the same constant, we observe that every random ω is a point of continuity. Hence V'yugin's null cover can be used to prove that for every ω, if ω is random, then the ergodic average is the same constant - if the ergodic average at a random $\omega = d \neq \int f dP = c$, and r_1 and r_2 are rationals that separate c and d, then the sets $\{x \in \Omega : r_1 < \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) < r_2\}$ can be used to define an effective null cover to contain ω [83]. □

The class of essentially computable functions include simple step functions like the indicator random variables of the cylinders C_w. These functions, being discontinuous are not computable. These are the functions that we use to observe that the Strong Law of Large Numbers in probability theory is a special case of the Ergodic Theorem. We now are able have a simple proof that the Effective Strong Law of Large Numbers, originally due to van Lambalgen [78] is a special case of the modified Effective Ergodic Theorem.

Corollary 3.5.2 (Effective SLLN [78])**.** *Let* $(\Omega, \mathcal{F}, P)$ *be such that* P *is generated by a computable coin-toss probability measure. Then* $\lim_{n\to\infty}\sum_{i=0}^{n-1}\frac{\omega_i}{n} = P(1)$.

Proof. The system $(\Omega, \mathcal{F}, P, T)$ is a computable dynamical system. The indicator function $f = I_1$ is an effectively computable, integrable function. Substituting $f = I_1$ in (3.4.1), we get that $\lim_{n\to\infty}\sum_{i=0}^{n-1}\frac{\omega_i}{n} = P(1)$. □

3.6 Effective Notions in Ergodic Theory of Continued Fractions

We introduce some basic notions in the theory of continued fractions.

A real number $r \in (0,1)$ is said to have continued fraction expansion $[a_1, \dots, a_n, \dots]$ if r can be expressed as

$$r \quad = \quad \cfrac{1}{a_1 + \ddots + \cfrac{1}{a_n + \ddots}} \tag{3.6.1}$$

where $a_i \in \mathbb{N}$. The set of finite representations constitute exactly the set of rational numbers. We introduce some basic notations and inequalities.

Let $\frac{p_n}{q_n}$ be the representation of the rational $[a_1, \dots, a_n]$ obtained by truncating the representation of r to n places, written in lowest form (i.e., $\gcd(p_n, q_n) = 1$.). The fraction $\frac{p_n}{q_n}$ is called the n^{th} *convergent* of the real number r, the number p_n being the n^{th} *partial quotient* and q_n being the n^{th} *partial denominator* of the continued fraction. The following recurrence equality is well-known (see, for example, Dajani and Kraaikamp [18].), $\forall n p_N p_{n-1} - p_{n-1} q_n = (-1)^{n+1}$.

We prove the effective versions of two famous theorems in the ergodic theory of continued fractions, *viz.*, the Lévy-Kuzmin Theorem and Khinchin's Theorem.

We first note that the set of reals random wrt γ and μ are exactly the same, since the Radon-Nikodym derivative $\frac{d\gamma}{d\mu} = 1/x$ is a Lipshitz function, making every effective null-cover for μ, an effective null-cover for γ and conversely. We use Theorem 3.4 to prove the effective version of the Lévy-Kuzmin Theorem. The proof is an effective version of the classical proofs from the exposition in [18]. We only have to prove that the functions used in the proof are essentially computable.

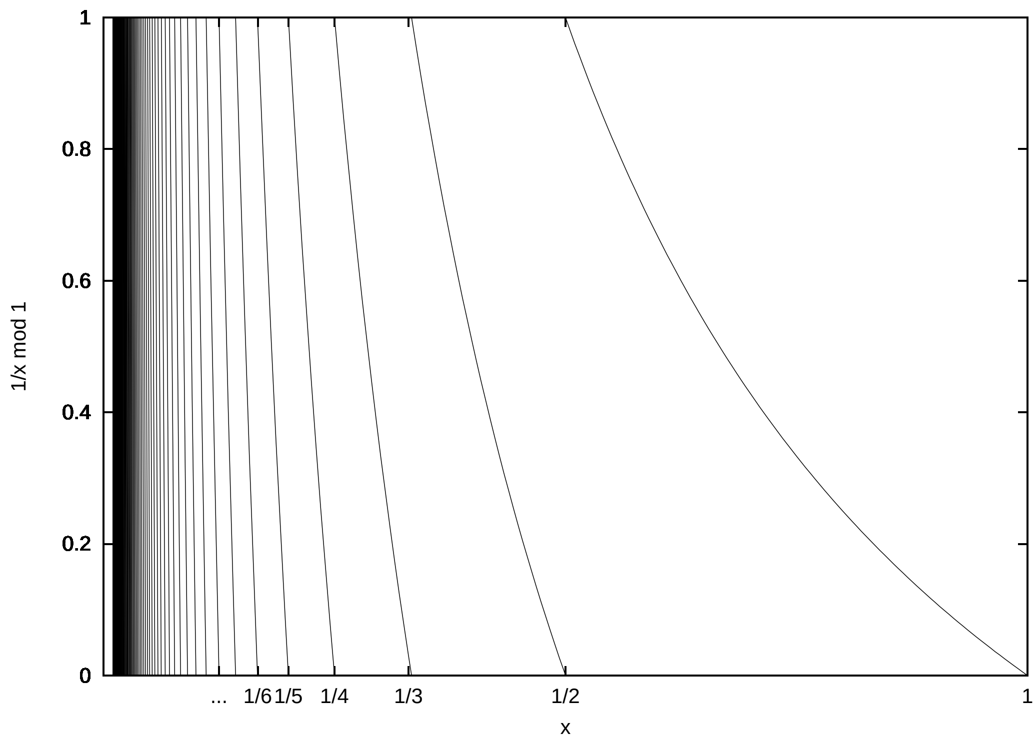
The left-shift transformation on the space of continued fractions with the Gauss measure is ergodic. (Fig 3.6.1)

I am grateful to David Kandathil for helping me in evaluating the integral in (3.6.3)

Theorem 3.6.1. *Let r be a real number in $(0,1)$. If r is constructively random, then*

$$\lim_{n \to \infty} \frac{\log\, q_n(r)}{n} = \frac{\pi^2}{12 \log 2}.$$

Figure 3.6.1 The left-shift map on continued fractions



Proof. We show that standard proofs using ergodic theory directly translate to the effective version. See [18] for the classical proof. It can be shown that

$$\log\, q_n(x) = \sum_{m=0}^{n-1} \log(T^m(x)) + R(n,x)$$

for all x, where the absolute value of the error $|R(n,x)|$ is bounded. The function $\log(x)$ is computable and monotone. The continued fraction map $g(x) = 1/x \mod 1$ on $(0,1)$ is a member of $\mathcal{G}_\mu$ and is a computable transformation.

Therefore, by the Effective Ergodic Theorem (Theorem 3.4), for all individual random x, we have

$$\lim_{n\to\infty} \frac{\sum_{m=0}^{n-1} f(T^m(x))}{n} = \int_0^1 \frac{\log x}{1+x} dx = \frac{1}{\ln 2} \int_0^1 \frac{\ln x}{1+x} dx \tag{3.6.2}$$

The integrand in (3.6.2) is unbounded in the region of integration, and this is an improper integral of type I. We have the following definition.

$$\int_0^1 \frac{\ln x}{1+x}dx = \lim_{\epsilon \to 0+} \int_\epsilon^1 \frac{\ln x}{1+x}dx \tag{3.6.3}$$

Integrating by parts, we get

$$\int_\epsilon^1 \frac{\ln x}{1+x}dx = \ln x \ln(1+x) \mid_\epsilon^1 - \int_\epsilon^1 \frac{\ln(1+x)}{x}dx = -\ln\epsilon\ln(1+\epsilon) - \int_\epsilon^1 \frac{\ln(1+x)}{x}dx. \tag{3.6.4}$$

The first term has the following evaluation as $\epsilon \to 0$:

$$\lim_{\epsilon\to 0} \frac{\ln\epsilon}{\frac{1}{\ln(1+\epsilon)}} \left[\frac{-\infty}{\infty} \text{ form}\right] = \lim_{\epsilon\to 0} \frac{\frac{1}{\epsilon}}{\frac{-\frac{1}{1+\epsilon}}{\ln^2(1+\epsilon)}} = \lim_{\epsilon\to 0} \frac{(1+\epsilon)\ln^2(1+\epsilon)}{\epsilon}$$

$$= \lim_{\epsilon\to 0} \frac{\ln^2(1+\epsilon)}{\epsilon} \left[\frac{0}{0} \text{ form}\right] = \lim_{\epsilon\to 0} \frac{2\ln(1+\epsilon)}{1+\epsilon} = \frac{2\times 0}{1} = 0.$$

The last integral term in the above expression can be evaluated using the Dilogarithm function.

$$Li_2(y) = -\int_0^y \frac{\ln(1-t)}{t}dt.$$

Substituting $t = -x$, we get

$$\int_0^1 \frac{\ln(1+x)}{x} = -\int_0^{-1} \frac{1-(-t)}{t}dt = -Li_2(-1) = -\left[-\frac{\pi^2}{12}\right],$$

so the quantity in (3.6.2) is equal to $\frac{\pi^2}{12\ln 2}$, which proves the result.

Alternately, using the logarithmic series,

$$\int_0^1 \frac{\ln(1+x)}{x}dx = \int_0^1 \sum_{i=0}^{\infty}(-1)^i \frac{x^i}{i+1}dx = \sum_{i=0}^{\infty}\int_0^1 (-1)^i \frac{x^i}{i+1}dx$$

$$= \sum_{i=0}^{\infty}(-1)^i \frac{1}{(i+1)^2} = \sum_{i=1}^{\infty}\frac{1}{i^2} - \frac{2}{4}\sum_{i=1}^{\infty}\frac{1}{i^2} = -\frac{\pi^2}{12}, \tag{3.6.5}$$

from the identity $\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$. The interchange of the summation and integration is justified by the fact that for all $0 \le x \le 1$ and all numbers n, $\sum_{i=1}^{n}(-1)^i \frac{1}{i+1} < \ln 2$. □

Now, we prove the effective version of Khinchin's theorem [18] where the additional result over graph computability is useful.

Theorem 3.6.2. *For every random $\omega \in [0,1)$ with the standard continued fraction expansion,*

$$\begin{aligned}\lim_{n\to\infty}(a_1a_2\dots a_n)^{1/n} &= \prod_{k=1}^{\infty}\left(1+\frac{1}{k(k+1)}\right)^{\frac{\log k}{\log 2}} \\ &= \mathcal{K} = 2.6854\dots.\end{aligned}$$

Proof. It suffices to show that for every random ω, the ergodic average

$$\lim_{n\to\infty}\frac{\log a_1(\omega)+\cdots+\log a_n(\omega)}{n} = \log\mathcal{K}.$$

The function $a_1(\omega) = \lfloor\frac{1}{\omega}\rfloor$ can be shown to be essentially upper (lower) semicomputable, even though it is not computable in $\mathcal{G}_\mu$, by approximating it from above (below) by a computable sequence of computable functions. Since log is a computable function, it follows that $f(x) = \log(a_1(x))$ is essentially computable. Hence by the effective ergodic theorem for random variables in $\mathcal{G}$, we have

$$\begin{aligned}\lim_{n\to\infty}\frac{\log a_1(\omega)+\cdots+\log a_n(\omega)}{n} &= \int\frac{\log a_1(\omega)}{(1+\omega)}d\omega \qquad (3.6.6)\\ = \sum_{k=1}^{\infty}\int_{\frac{1}{k+1}}^{\frac{1}{k}}\frac{\log a_1(\omega)}{1+\omega}d\omega &= \sum_{k=1}^{\infty}\frac{\log k}{k(k+2)},\end{aligned}$$

which is a convergent series with limit $\log\mathcal{K}$. □

Easily calculable analytical expressions and numerical evaluations of $\mathcal{K}$ are very tough to obtain, and it is not known whether this constant is irrational.

CHAPTER 4. A Characterization of Constructive Dimension

4.1 Introduction

The measure-theoretic approach to the definition of random sequences identifies a property of "typical" sets. A random sequence is one that belongs to every reasonable majority of sequences [43]. The notion of a reasonable majority is formulated as an effective version of measure 1. Each measure 1 set has a complement set of measure 0. It is hence sufficient to define the concept of the effective measure zero set. [1]

Let P be a computable probability measure defined on the Cantor Space. Martin-Löf proved a universality property – that for any computable measure P, there is a unique largest effective P-measure zero set. The elements in the complement of this set are the set of P-random sequences.

Another tool in the study of effective randomness is the concept of martingales. Introduced by Ville in the 1930s [79] (being implicit in the early work of Lévy [45], [46]), they were applied in theoretical computer science by Schnorr in the early 1970s [68], [69], [70] in his investigation of Martin-Löf randomness, and by Lutz [51], [52], [53] in the development of resource-bounded measure. A martingale is a betting strategy, which, for a probability measure P defined on the Cantor Space, obeys the conditions,

$$\begin{aligned} d(\lambda) &\leq 1 \\ d(w)P(w) &= d(w0)P(w0) + d(w1)P(w1). \end{aligned} \tag{4.1.1}$$

Intuitively, it can be seen as betting strategy on an infinite sequence, where, for each prefix w of the infinite sequence, the amount $d(w)$ is the capital that is in hand after betting. A

[1] This work appeared in [64].

martingale can be seen as a fair betting condition where the expected value after every bet is the same as the expected value before the bet is made. It is said to succeed on a sequence ω if

$$\limsup_{n\to\infty} d(\omega[0\ldots n-1]) = \infty.$$

The success set of a martingale, $S^{\infty}[d]$, is the set of all individual sequences on which it succeeds. In this chapter, we consider constructive martingales. There is a universal martingale $\tilde{d}$ which is constructive and for every ω, if there is a constructive martingale d which succeeds on ω, then $\tilde{d}$ succeeds on ω. The theory of martingales and their applications to the field of resource-bounded measure, complexity theory, and resource-bounded dimension has proved to be remarkably fruitful. In this work, we wish to establish connections between martingales and a third approach of defining randomness, *viz.*, that of a measure of impossibility.

This third approach to define a random sequence is to characterize a degree of disagreement between any sequence ω and the probability P. Following [82], a *measure of impossibility* is a positive function $p(\omega)$ which describes the quantitative level to which ω is impossible with respect to the probability measure P. A measure of impossibility is defined to be a lower semicomputable function p, which is integrable with respect to P. It can be seen that if ω is P-random, then $p(\omega) < \infty$. There is an optimal measure of impossibility $\tilde{p}$ such that a sequence ω is random if and only if $\tilde{p}(\omega) < \infty$ [82], [83]. This concept is a central one in V'yugin's proof of an effective version of the Ergodic Theorem [83].

The relation between martingales and Martin-Löf's definition of randomness was studied by Schnorr [68], [69], [70]. The proof that the notion of randomness defined by Martin-Löf corresponds to that of the ones defined via the measures of impossibility is due to Vovk and V'yugin [82] and V'yugin [83]. We establish a direct correspondence between the notions of constructive martingales and measures of impossibility.

We then apply this construction to come up with an analogous new definition of constructive dimension [56] in terms of a generalized version of the notion of a measure of impossibility. We show that this construction also generalizes to give an alternate definition of constructive scaled dimension [31].

The main difference between a proof based on martingales and one using a measure of impossibility is that a martingale is defined on the basis cylinders, and a measure of impossibility is a pointwise notion. Measure of impossibility seems to be an easier tool in dealing with theorems in which we have to reason about the convergence of general random variables defined on the points in the sample space. However, we show that at the constructive level, these tools are equivalent. Since there are universal objects available in both the cases, there exists an indirect conversion between the two such that the success set of a martingale can be converted into that of a measure of impossibility and conversely; this work contributes a direct constructive conversion of one into another. The theory of algorithmic randomness has been remarkably fruitful to date. (For a survey of the field, see [22].) Martingales have proved to have greater apparent utility in some cases than Martin-Löf tests in studying randomness, and measures of impossibility have been of use in establishing a remarkable result in the study of algorithmic randomness, the Effective Ergodic Theorem [83] . We hope that the explicit transformation of this work will improve the understanding, and perhaps the utility of measure of impossibility. Moreover, in the absence of universal objects which happens at computable and other levels a conversion of this nature may be useful.

4.2 Lower Semicomputability

We prove an equivalent notion of lower semicomputability (q.v):

Lemma 4.2.1. *The following hold.*

i. A function $f : \mathbf{C} \to [-\infty, \infty]$ is lower semicomputable if and only if there exists a computable function $\hat{f} : \Sigma^* \times 0^{\mathbb{N}} \to \mathbb{Q} \cup \{-\infty, \infty\}$ such that the following hold: For all $\omega \in \mathbf{C}$,

(a) Monotonicity: For all $m, n \in \mathbb{N}$, $\hat{f}(\omega[0 \ldots m-1], 0^n) \leq \hat{f}(\omega[0 \ldots m-1], 0^{n+1}) \leq f(\omega)$, and $\hat{f}(\omega[0 \ldots m-1], 0^n) \leq \hat{f}(\omega[0 \ldots m], 0^n) \leq f(\omega)$.

(b) Convergence: We have

$$\lim_{n \to \infty} \hat{f}(\omega[0 \ldots n-1], 0^n) = f(\omega).$$

ii. $f : \Sigma^* \to [-\infty, \infty]$ *is lower semicomputable if and only if there exists a function* $\hat{f}$: $\Sigma^* \times 0^{\mathbb{N}} \to \mathbb{Q} \cup \{-\infty, \infty\}$ *such that the following hold: For all* $x \in \Sigma^*$,

(a') Monotonicity: For all $n \in \mathbb{N}$, $\hat{f}(x, 0^n) \leq \hat{f}(x, 0^{n+1}) \leq f(x)$.

(b') Convergence: $\lim_{n\to\infty} \hat{f}(x, 0^n) = f(x)$.

Proof. The characterization for the case when f is defined on the domain Σ^* is standard in the literature (see [56]), and we prove the formulation for the case when the domain is $\mathbf{C}$.

For the case when $f : \mathbf{C} \to [-\infty, \infty]$ is lower semicomputable, first assume that the set $S = \{(\omega, q) \mid \omega \in \mathbf{C}, q \in \mathbb{Q}, f(\omega) > q\}$ is the union of a computable enumeration $S : 0^{\mathbb{N}} \to \Sigma^* \times (\mathbb{Q} \cup \{-\infty\})$ of cylinders in the natural topology on $\mathbf{C} \times \mathbb{Q}$. The projection functions $\pi_1 : \Sigma^* \times \mathbb{Q} \to \Sigma^*$ and $\pi_2 : \Sigma^* \times \mathbb{Q} \to \mathbb{Q}$ are defined as $\pi_1(w, q) = w$ and $\pi_2(w, q) = q$. We design a witness function $\hat{f} : \Sigma^* \times 0^{\mathbb{N}} \to \mathbb{Q} \cup \{-\infty\}$ in the following algorithm.

procedure $\hat{f}(w,\, 0^n)$
    Set $\leftarrow \{-\infty\}$.
    i $\leftarrow 0$.
    while i $\leq n$ **do**
        if $\pi_1(S(i)) \sqsubseteq w$ **then**
            Set $\leftarrow$ Set $\cup \{\pi_2(S(i))\}$
        end if
        i $\leftarrow i + 1$.
    end while
    return max(Set)
end procedure

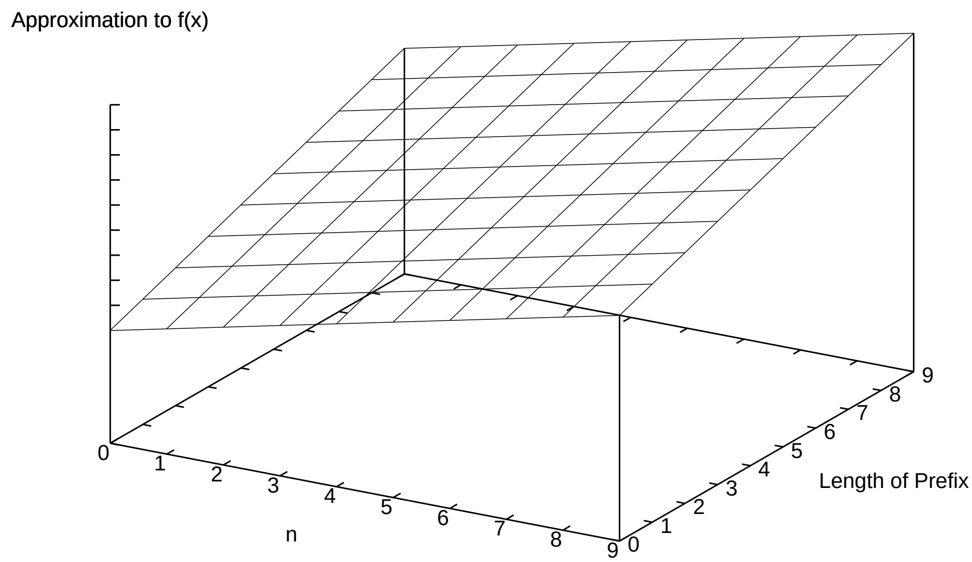
The monotonicity condition is satisfied, because in the algorithm, for every n, the sets have the following relationships:

$$\{\pi_2(S(i)) \mid \pi_1(S(i)) \sqsubseteq w, 0 \leq i \leq n\} \subseteq \{\pi_2(S(i)) \mid \pi_1(S(i)) \sqsubseteq w, 0 \leq i \leq n+1\},$$

and, for strings w' and w, if $w' \sqsubseteq w$, then

$$\{\pi_2(S(i)) \mid \pi_1(S(i)) \sqsubseteq w', 0 \leq i \leq n\} \subseteq \{\pi_2(S(i)) \mid \pi_1(S(i)) \sqsubseteq w, 0 \leq i \leq n\}.$$

Figure 4.2.1 $\hat{f}$ is monotone in both its arguments separately.



For the convergence, it is obvious that

$$\lim_{n \to \infty} f(\omega[0 \ldots n-1], 0^n) \tag{4.2.1}$$

exists, since it is a monotone bounded sequence in a compact space. To see that the limit is $f(\omega)$, assume that the limit (4.2.1) is a real $r < f(\omega)$. Then there exists a rational r', $r < r' < f(\omega)$, such that there is no prefix w of ω such that (w, r') occurs in the enumeration of S. This is a contradiction. Hence the condition is satisfied, and limit (4.2.1) is $f(\omega)$.

Conversely, let $f : \mathbf{C} \to \overline{\mathbb{R}}$ be a lower semicomputable function with witness $\hat{f} : \Sigma^* \times 0^{\mathbb{N}} \to \mathbb{Q}$ satisfying lower semicomputability conditions. We prove that the set $S = \{(\omega, r) : r \in \mathbb{Q}, r < f(\omega)\}$ is the union of a computable enumeration of cylinders in $\mathbf{C} \times \mathbb{Q}$.

We show that for every $r \in \mathbb{Q}$, $r < f(\omega)$, there is a prefix w of ω, such that (w, r) is accepted by an algorithm. This is routine to see, since, we can dovetail the execution of $\hat{f}$ on $\Sigma^* \times 0^{\mathbb{N}}$. If $r < f(\omega)$, there is an $r' > r$ such that $\hat{f}$ produces r' on some prefix w and some 0^m, and then we can accept (w, r). $\square$

A function f is called upper semicomputable if $-f$ is lower semicomputable. A function f is called computable if it is both lower and upper semicomputable. This may be seen to be equivalent to the following definition in the case of functions defined over Σ^*.

Definition. A function $f : \Sigma^* \to \mathbb{R}$ is said to be computable if there exists a function $\hat{f} : \Sigma^* \times 0^{\mathbb{N}} \to \mathbb{Q}$ such that for every $n \in \mathbb{N}$ and every $x \in \Sigma^*$,

$$|\hat{f}(x, 0^n) - f(x)| \leq 2^{-n}.$$

Remark. It is easy to show that if $\hat{f}$ is a witness function to the computability of f, then for all n, $\hat{f}(x, 0^n) - 2 \cdot 2^{-n}$ is a lower semicomputation, and $\hat{f}(x, 0^n) + 2 \cdot 2^{-n}$ is an upper semicomputation of f.

4.3 Effective Randomness

Recall the definition of a computable probability measure from section 3.2. A probability measure $P : \mathcal{F} \to [0, 1]$ is called computable if the probability measure $P : \Sigma^* \to [0, 1]$ is a computable function. The notation $P(w)$ for a string $w \in \Sigma^*$ stands for $P(C_w)$, the probability of the cylinder C_w.

Definition. A probability measure is called strongly positive if there is positive rational constant c ($0 < c < 1$) such that

$$c < \inf_{\substack{w \in \Sigma^* \\ b \in \Sigma}} \{P(C_{wb} \mid C_w)\}. \tag{4.3.1}$$

We limit ourselves to the class of computable, strongly positive probability measures.

Building on the definition of a measure of impossibility (section 3.2), we define the following.

Definition. Let $\omega \in \mathbf{C}$. Then ω is said to be P-impossible if there is a P-measure of impossibility p such that $p(\omega) = \infty$. A set $X \subseteq \mathbf{C}$ is said to be P-impossible, if

$$X \quad \subseteq \quad \{\omega : \omega \text{ is } P\text{-impossible}\}.$$

V'yugin and Vovk [82] proved that for every computable probability P, there is an optimal measure of impossibility $\tilde{p} : \mathbf{C} \to [0, \infty]$ such that a sequence ω is P-random if and only if $\tilde{p}(\omega) < \infty$. The set of Martin-Löf random sequences with respect to P is exactly the complement of the set of all $\omega \in \mathbf{C}$ that are P-impossible.

We also consider martingales.

Definition. A P-martingale $d : \Sigma^* \to [0, \infty]$, is a function which obeys the properties,

M1. $d(\lambda) \leq 1$.

M2. For all strings w, the following holds:

$$d(w)P(w) = d(w0)P(w0) + d(w1)P(w1).$$

A martingale is said to "succeed" on sequence ω if

$$\limsup_{n \to \infty} d(\omega[0 \ldots n-1]) = \infty.$$

A martingale is said to "strongly succeed" on sequence ω if

$$\liminf_{n \to \infty} d(\omega[0 \ldots n-1]) = \infty.$$

The success set of a martingale d, denoted $S^{\infty}[d]$, is defined to be the set of binary sequences on which d succeeds. The strong success set of a martingale d, denoted $S^{\infty}_{\text{str}}[d]$, is the set of binary sequences on which d strongly succeeds.

A constructive martingale is a lower semicomputable martingale.

We show, the concept of a measure of impossibility and that of a constructive martingale are equivalent, in that every measure of impossibility p corresponds to a martingale which wins on an ω if and only if $p(\omega) = \infty$. Since there is a universal martingale which succeeds on the set of non-random sequences, and there is a universal measure of impossibility which attains ∞ on the set of non-random sequences, it is indirectly known that there is a conversion between the success criteria of martingales and that of measures of impossibility. The new result here is a direct conversion of a martingale into a measure of impossibility and vice versa, such that the success sets of both are the same (under some assumptions on the probability measure).

4.4 Converting a Martingale into a Measure of Impossibility

Let P be a strongly positive computable probability measure. We wish to convert a lower semicomputable P-martingale which succeeds on a constructive P-measure-zero set, to a *measure of impossibility* $p : \mathbf{C} \to [0, \infty]$ with respect to P, such that $S^\infty[d]$ is P-impossible as witnessed by p. We show that if $d : \Sigma^* \to \mathbb{R}^+$ is a lower semicomputable P-martingale, then there exists a *measure of impossibility* $p : \Omega \to [0, \infty]$ such that $\forall \omega \in \Omega \ \limsup_{n\to\infty} d(\omega[0 \dots n-1]) = \infty$ if and only if $p(\omega) = \infty$.

We proceed in stages.

It is well-known that a sequence ω is non-random if and only if there is a martingale d which is such that $\liminf_n d(\omega[0 \dots n-1]) = \infty$. The following well-known lemma is stated here because we use the construction to prove results about the measure of impossibility.

Lemma 4.4.1. [*Schnorr [69]*] *If* $d : \Sigma^* \to [0, \infty]$ *is a constructive martingale, then there is a constructive martingale* $d' : \Sigma^* \to [0, \infty]$ *such that* $S^\infty[d] \subseteq S^\infty_{str}[d']$. *Moreover, there is a monotone function* $\mathtt{sa} : \Sigma^* \to [0, \infty]$ *such that* $\lim_{n\to\infty} \mathtt{sa}(\omega[0 \dots n-1]) = \infty$ *if and only if* $\omega \in S^\infty_{str}[d']$.

Proof sketch:

The construction follows the construction of the martingale in Theorem 2.6 of [69]. Let the constant c witness the strong positivity of the computable probability measure P, as in Definition 4.3. For each n, let $A_n = \{w \in \Sigma^* \mid d(w) > [1/c]^n\}$. Let us denote, for any set of strings A, by $P(A)$, the probability $\cup_{w\in A} P(C_w)$. It follows that $P(A_n) \leq c^n$. Since d is constructive, the A_n's are uniformly computably enumerable. For each computably enumerable set A_n, there is a recursive, prefix-free set B_n such that the set of infinite sequences with prefixes in B_n is exactly the same as those with prefixes in A_n.

For each n, the function $d_n : \Sigma^* \to \mathbb{R}^+$ is defined as

$$d_n(x) = \sum_{xy \in B_n} P(xy|x) + \mathtt{sa}_n(x)$$

where $\mathtt{sa}_n(x) = 1$ if some prefix of x is in B_n, and 0 otherwise. Then

$$d_n(\lambda) = \sum_{w \in B_n} P(w) \leq c^{-n}.$$

Each d_n is a P-martingale, so $d' = \sum_{i=0}^{\infty} d_n$ is a P-martingale. The A_ns can be uniformly generated by an algorithm, hence so can the d_ns, it follows that d' is constructive.

Also, if $d(\omega[0 \ldots k-1]) > [1/c]^i$, then $\sum_{j=0}^{i} \mathtt{sa}_j(\omega[0 \ldots k-1]) > i$, so $d'(w[0 \ldots m-1]) > i$ for all $m > k$. This proves the lemma. □

Let d' be a P-martingale as defined above. The measure of impossibility p is defined as follows.

$$p(\omega) = \lim_{n \to \infty} \sum_{i=0}^{n-1} \mathtt{sa}(\omega[0 \ldots n-1]). \tag{4.4.1}$$

We prove that p is a measure of impossibility which attains ∞ on all sequences on which d' strongly succeeds.

Lemma 4.4.2. *p defined in (4.4.1) is a measure of impossibility.*

Proof. Let $\hat{\mathtt{sa}}$ be the witness of the lower semicomputability of $\mathtt{sa}$. Then $\hat{\mathtt{sa}}$ witnesses that p is lower semicomputable.

Moreover,

$$\begin{aligned} E_P\, p &= \int p(\omega) dP \\ &\leq \liminf_{n \to \infty} \sum_{w' \in \{0,1\}^n} \int_{C_{w'}} \mathtt{sa}(w') dP \\ &= \lim_{n \to \infty} \sum_{w' \in \{0,1\}^n} \int_{C_{w'}} \mathtt{sa}(w') dP, \end{aligned}$$

by Fatou's Lemma. But, since $\mathtt{sa}(w') \leq d(w')$ for all w', it follows by Kraft's inequality that

$$\sum_{w' \in \{0,1\}^n} \int \mathtt{sa}(w') P(w') dP \leq \sum_{w' \in \{0,1\}^n} \int d'(w') P(w') dP \leq 1,$$

for all n. Hence it follows that $\int p(\omega) dP \leq 1$. Therefore p defines a *measure of impossibility.* □

Let $\omega \in S^{\infty}[d]$. It is clear that, since $\liminf_{n \to \infty} d'(\omega[0 \ldots n-1]) = \infty$ implies that $\sup_n \mathtt{sa}(\omega[0 \ldots n-1]) = \infty$, we have $p(\omega) = \infty$. Thus p satisfies the conditions of being a measure of impossibility which attains ∞ on ω.

4.5 Converting a Measure of Impossibility into a Martingale

We assume $P : \mathcal{F} \to [0,1]$ is a computable probability measure. If $p : \Omega \to [0,\infty]$ is a *measure of impossibility* with respect to P, we prove: there exists a constructive P-martingale $d : \Sigma^* \to \mathbb{R}^+$ such that d succeeds on every ω on which p assumes ∞– *i.e.* $\{\omega : \limsup_{n\to\infty} d(\omega[0 \ldots n-1]) = \infty\} \supseteq \{\omega : p(\omega) = \infty\}$, with equality if P is a measure which assigns positive probability to every cylinder.

We make the following restrictions: We ensure that P is not just a computable probability measure, but also *very strongly* positive: if $\hat{P}$ testifies to the fact that P is computable, then there exists a program $f : \Sigma^* \times 0^{\mathbb{N}} \to \mathbb{Q}$ such that for every cylinder C_x, the probability of the cylinder $P(C_x) > 0$ if and only if for all positive integers n, we have $\hat{P}(x, 0^n) > f(x, 0^n)$. Note that if a probability measure is strongly positive, then it is very strongly positive.

We define the P-martingale d.

For the empty string, $d(\lambda) = E_P[p]$. For all strings w, and $b \in \{0,1\}$,

$$d(wb) = \begin{cases} E_P[p \mid C_{wb}] & \text{if } P(C_{wb}) > 0 \\ 2 \cdot d(w) & \text{otherwise.} \end{cases} \tag{4.5.1}$$

Lemma 4.5.1. *If p is a measure of impossibility with respect to a very strongly positive, computable probability measure P, then d defined in (4.5.1), is a lower semicomputable P-martingale.*

Proof. We show that d is a P-martingale, and d is lower semicomputable,

d is a P-martingale:

We have $d(\lambda) = \frac{\int_\Omega p(\omega)dP}{P(\Omega)} \leq 1$.

For a string w, if all of C_w, C_{w0} and C_{w1} have non-zero probability, then the stipulation (M2) is satisfied by linearity of conditional expectation. If, $P(C_w) \neq 0$, but one of $P(C_{w0})$,

$P(C_{w1})$ is zero, without loss of generality, say $P(C_{w0}) = 0$, then

$$\begin{aligned} d(w0)P(C_{w0}) + d(w1)P(C_{w1}) &= 2d(w) \times 0 + E[p|C_{w1}]P[C_{w1}] \\ &= E[p|C_{w1}]P(C_w) = E[p|C_w]P(C_w) \\ &= d(w)P(C_w). \end{aligned}$$

d is lower semicomputable:

Consider the following program: Algorithm for $\hat{d} : \Sigma^* \times 0^{\mathbb{N}} \to \mathbb{Q}$: Let $x \in \Sigma^*, b \in \Sigma$.

procedure $\hat{d}(xb,\, 0^n)$

Input $xb \in \Sigma^*(b \in \Sigma)$ and $n \in \mathbb{N}$.

if $f(xb, 0^n) > \hat{P}(xb, 0^n)$ **then**

$\hat{d}(xb, 0^n) = 2\,\hat{d}(x, 0^n)$.

else

$$\hat{d}(xb, 0^n) = \frac{\sum_{w\in\{0,1\}^n}\ \max_{y\sqsubseteq xbw}\{\hat{p}(y, 0^n)\} \times (\hat{P}(xbw, 0^{2n+1}) - 2\cdot 2^{-2n-1})}{P(xb, 0^{2n+1}) + 2\cdot 2^{-2n-1}}.$$

end if

end procedure

To show that $\hat{d}(xb, \cdot)$ is a lower semicomputation of $d(xb)$, we proceed as follows. We prove that the numerator in else statement converges to the appropriate limit. From this, it follows that the output of the program converges to the value of d for the given string xb from below in a lower semicomputable way.

Define the following:

$$\begin{aligned} \forall xb \in \Sigma^*, m \in \mathbb{N}\ f_m^{xb0^\infty} &= \sum_{w\in\{0,1\}^m} \max_{y\sqsubseteq xbw} \{\hat{p}(y, 0^m)\} \left[\hat{P}(xbw, 0^{2m+1}) - 2\cdot 2^{-2m-1}\right] \\ \forall xb \in \Sigma^*, m \in \mathbb{N}\ S_m^{xb0^\infty} &= \sum_{w\in\{0,1\}^m} \max_{y\sqsubseteq xbw} \{\hat{p}(y, 0^m)\} P(xbw) \end{aligned}$$

The following claims suffice to prove that $\hat{d} : \mathbb{N} \times \Sigma^* \to \mathbb{Q}$ is a lower semicomputation of d.

Lemma 4.5.2. $\forall m \in \mathbb{N}\ f_m^{xb0^\infty} \ \leq S_m^{xb0^\infty} \ \leq \int_{C_{xb}} p(\omega)\, dP.$

Proof sketch: Lower semicomputability of P with witness $\hat{P}(\cdot, 0^m) - 2.2^{-m}$ for all $m \in \mathbb{N}$ implies the first inequality. The second is by the lower semicomputability of $\hat{p}$ with respect to p, and the fact that each $S_m^{x0^\infty}$ is the integral of a step function defined on Ω, $\hat{p} < p$ (everywhere), and by the definition of the Lebesgue integral. $\square$

Now, we show that the sum converges as $n \to \infty$ to the required integral:

Lemma 4.5.3. *The series $f_m^{xb0^\infty}$ converges uniformly to the same limit as of the sum series $S_m^{xb0^\infty}$ as $m \to \infty$.*

Proof. By the computability witness $\hat{P}$ of P, we have, for any xbw, $m \in \mathbb{N}$,

$$P(xbw) - \hat{P}(xbw, 0^{2(m+1)}) < \frac{1}{2^{(2m+1)}},$$

whereby

$$|f_m^{xb0^\infty} - S_m^{xb0^\infty}| < \frac{1}{2^{2m+1}} \times 2^m = \frac{1}{2^{m+1}}$$

$\square$

The fact that $S_m^{xb0^\infty} \to \int_{C_{xb}} p(\omega)dP$ as $m \to \infty$, follows due to the fact that $\hat{p}$ is a lower semicomputation of p. Property (1) of lower semicomputability ensures that p dominates the step function summed in $S_m^{xb0^\infty}$. The convergence property of lower semicomputability ensures that the the function $S_m^{xb0^\infty}$ converges to the integral $\int pdP$.

These claims suffice to establish the condition that $\hat{d}$ has to satisfy to be a lower semicomputation of d.

$\square$

Lemma 4.5.4. *For any $\omega \in \mathbf{C}$, $p(\omega) = \infty$ implies $\limsup_{n\to\infty} d(\omega[0 \to n-1]) = \infty$.*

Proof. First, if $P(C_{\omega[0\ldots n-1]}) = 0$ for some n, this is routine to prove.

So, assume for all n, $P(C_{\omega[0\ldots n-1]}) > 0$, so that for all n, $d(\omega[0\ldots n-1]) = E[p|C_{\omega[0\ldots n-1]}]$. We show that d succeeds on ω. It is enough to show that for every rational q, there is some $x \sqsubseteq \omega$ such that $d(x) > q$.

Let S be the graph of p, described by the union of a computable enumeration $\mathcal{C}$ of cylinders in $\mathbf{C} \times \mathbb{Q}$. If $p(\omega) = \infty$, then for every $q \in \mathbb{Q}$, there is an $(x, q) \in \mathcal{C}$ such that x is some prefix of ω. If this is so, then for every $\sigma \in C_x$, $p(\sigma) > q$, whence $E[p|C_x] > q$, which proves the result. $\square$

4.6 A New Characterization of Constructive Dimension

In this section, we generalize the construction of the previous sections, to come up with an alternate definition of constructive Hausdorff and constructive packing dimension.

For $s \in [0, \infty)$, we introduce the notion of a set being s-improbable with respect to a measure of impossibility.

Definition. Let $X \subseteq \mathbf{C}$. We say that X is *s-improbable with respect to a P-measure of impossibility* $p : \mathbf{C} \to [0, \infty]$ if for every infinite binary sequence $\omega \in X$, we have

$$\limsup_{n\to\infty} \frac{\int_{C_{\omega[0\ldots n-1]}} p(\omega)dP}{P^s(C_{\omega[0\ldots n-1]})} = \infty. \tag{4.6.1}$$

X is *strongly s-improbable with respect to p* if

$$\liminf_{n\to\infty} \frac{\int_{C_{\omega[0\ldots n-1]}} p(\omega)dP}{P^s(C_{\omega[0\ldots n-1]})} = \infty. \tag{4.6.2}$$

The concept of s-improbability generalizes the concept of impossibility.

Lemma 4.6.1. *Let $p : \mathbf{C} \to [0, \infty]$ be a P-measure of impossibility.*

1. *For any $\omega \in \mathbf{C}$, if $p(\omega) = \infty$, then $\limsup_{n\to\infty} \frac{\int_{C_{\omega[0\ldots n-1]}} p(\omega)dP}{P(C_{\omega[0\ldots n-1]})} = \infty$.*
2. *For any $\omega \in \mathbf{C}$, if $\limsup_{n\to\infty} \frac{\int_{C_{\omega[0\ldots n-1]}} p(\omega)dP}{P(C_{\omega[0\ldots n-1]})} = \infty$, then there exists a P-measure of impossibility $p' : \mathbf{C} \to [0, \infty]$ such that $p'(\omega) = \infty$.*

Proof. 1.) Let $\omega \in \mathbf{C}$ be such that $p(\omega) = \infty$. Since p is lower semicomputable and $P(C_w) > 0$ for all strings w, the function $d : \Sigma^* \to \mathbb{R}^+$ defined by $d(w) = E_P[p|C_w]$ is a martingale, such that $\omega \in S^\infty[d]$. Thus, $\limsup_{n\to\infty} \frac{\int_{C_{\omega[0\ldots n-1]}} p(\omega)dP}{P(C_{\omega[0\ldots n-1]})} = \infty$.

2.) Since $\int pdP \leq 1$ and $P(C_w) > 0$ for all w, we take the equivalent characterization that $\sup_n \frac{\int_{C_{\omega[0\ldots n-1]}} p(\omega)dP}{P(C_{\omega[0\ldots n-1]})} = \infty$. Since p is lowersemicomputable and P is computable, $f' : \mathbf{C} \to [0, \infty]$ defined by $f'(\chi) = \sup_n \frac{\int_{C_{\chi[0\ldots n-1]}} p(\chi)dP}{P(C_{\chi[0\ldots n-1]})} = \infty$ is a measure of impossibility. $\square$

We now review the notion of a lower semicomputable s-P-gale, which, following Lutz [56], we use to give a definition of constructive Hausdorff (or constructive Billingsley) dimension.

Definition. (Lutz [56]) Let $s \in [0, \infty)$. An *s-P-gale* $d : \Sigma^* \to \mathbb{R}^+$ is a function that satisfies the condition for all $w \in \Sigma^*$,

$$d(w)P^s(w) \quad = \quad [d(w0)P^s(w0) + d(w1)P^s(w1)] \tag{4.6.3}$$

Definition. (Lutz [56]) Let d be an s-P-gale, where $s \in [0, \infty)$.

- We say that d succeeds on $\omega \in \mathbf{C}$ if

$$\limsup_{n\to\infty} d(\omega[0 \ldots n-1]) = \infty.$$

- The success set of d is

$$S^{\infty}[d] = \{\omega \in \mathbf{C} \mid d \text{ succeeds on } \omega\}.$$

- We say that d strongly succeeds on $\omega \in \mathbf{C}$ if

$$\liminf_{n\to\infty} d(\omega[0 \ldots n-1]) = \infty.$$

- The strong success set of d is

$$S^{\infty}_{str}[d] = \{\omega \in \mathbf{C} \mid d \text{ strongly succeeds on } \omega\}.$$

The notion of constructive Hausdorff dimension, a constructive analogue of the classical Hausdorff dimension, is defined using the notion of constructive s-gales [56].

Remark.[Lutz [56]] For every $s_1, s_2 \in [0, \infty)$, the function $d : \Sigma^* \to [0, \infty]$ is a P-s_1-gale if and only if the function $d' : \Sigma^* \to [0, \infty]$ defined by $d'(w) = P^{(s_1 - s_2)}(w)d(w)$ is an s_2 gale.

Definition. (Lutz [56], Lutz and Mayordomo [50]) The constructive *Hausdorff dimension* of a set $X \subseteq \mathbf{C}$ is

$$\dim_H^P(X) \quad = \quad \inf\{s \in [0, \infty) \mid \text{ There is a constructive } s\text{-}P\text{-gale for which } X \subseteq S^{\infty}[d].\}$$

Analogously, the notion of constructive packing dimension is defined as the constructive analogue of the classical packing dimension [77], [76]. We use the following equivalent notion defined using strong success of s-P-gales.

Definition. (Athreya et al. [2], [50]) The constructive *packing dimension* of a set $X \subseteq \mathbf{C}$ is

$$\mathrm{Dim}_H^P(X) \quad = \quad \inf\{s \in [0,\infty) \mid \text{ There is a constructive } s\text{-}P\text{-gale for which } X \subseteq S^{\infty}_{\mathrm{str}}[d].\}$$

Notation. For $X \subseteq \mathbf{C}$, let $\mathcal{G}(X)$ be the set of all $s \in [0,\infty)$ such that there is an s-P-gale d for which $X \subseteq S^{\infty}[d]$ and $\mathcal{P}(X)$ be the set of all $s \in [0,\infty)$ such that X is s-improbable with respect to some P-measure of impossibility. Similarly, let $\mathcal{G}_{\mathrm{str}}(X)$ be the set of all $s \in [0,\infty)$ such that there is an s-P-gale d for which $X \subseteq S^{\infty}_{\mathrm{str}}[d]$ and $\mathcal{P}_{\mathrm{str}}(X)$ be the set of all $s \in [0,\infty)$ such that X is strongly s-improbable with respect to some P-measure of impossibility.

The following lemma asserts that for every strongly positive computable probability measure P, every positive rational s, for every $X \subseteq \mathbf{C}$, there is a P-s-gale which succeeds on X if and only if X is s-improbable with respect to the probability measure, and analogously, there is an P-s-gale which strongly succeeds on X if and only if X is strongly s-improbable with respect to the probability measure.

The construction uses an analogue of the "savings account" method used in the proof of Lemma 7. In Lemma 7, the method was used to prove that the set of sequences on which there is a P-martingale succeeds, is the same as that on which there is a P-martingale which strongly succeeds. In the case of s-gales, we cannot prove this: there are sequences and some $s \in [0,\infty)$ such that there is some constructive s-gale which succeeds on them, but there are no s-gales which succeed strongly. However, we used the construction in Lemma 7 to extract a monotone behavior from the martingale, which was used in the definition of the measure of impossibility. The construction in the next lemma shows that it is possible to extract a monotone behavior from an s-gale to define a measure of impossibility which has desirable properties.

Lemma 4.6.2. *Let P be a strongly positive computable probability measure. Let $X \subseteq \mathbf{C}$ and $s \in [0,\infty)$ be a rational. Then $s \in \mathcal{G}(X)$ if and only if $s \in \mathcal{P}(X)$, and $s \in \mathcal{G}_{str}(X)$ if and only*

Figure 4.6.1 Algorithm 2

procedure $d^{(n)}(w)$
    $\alpha_n = 2c^{(s_2-s_1)n/2}$
    if $|w| < n$ **then**
        $\mathtt{bc}'_n(w) \leftarrow \alpha_n d_2(w)$
        $\mathtt{sa}'_n(w) \leftarrow 0$
    else
        $val_w \leftarrow d_2(w[0 \ldots n-1])P^{(1-s_2)}(C_w | C_{w[0 \ldots n-1]})$
        if $d_2(w[0 \ldots n-1]) > P^{s_1-s_2}(C_{w[0 \ldots n-1]})$ **then**
            $\mathtt{bc}'_n(w), \mathtt{sa}'_n(w) \leftarrow \frac{\alpha_n val_w}{2}$
        else
            $\mathtt{bc}'_n(w) \leftarrow \alpha_n val_w$
            $\mathtt{sa}'_n(w) \leftarrow 0$
        end if
    end if
    return $\mathtt{bc}'_n(w) + \mathtt{sa}'_n(w)$
end procedure

if $s \in \mathcal{P}_{str}(X)$.

Proof. Let c be as in (4.3.1). Let $\epsilon > 0$ be arbitrary and $s_1 < 1$ be rational, and let d_1 be an s_1-P-gale. Let s_2 be a rational number such that $s_2 \in (s_1, \min\{1, s_1 + \epsilon\})$. Then $d_2 : \Sigma^* \to R$ defined by, for all strings w, $d_2(w) = d_1(w)P(w)^{(s_1-s_2)}$ is an s_2-P-gale such that $S^\infty[d_1] \subseteq S^\infty[d_2]$ and $S^\infty_{\text{str}}[d_1] \subseteq S^\infty_{\text{str}}[d_2]$. As in the case of constructive martingales, we can form another s_2-P-gale d' consisting of $\mathtt{bc}'$ and $\mathtt{sa}'$ such that the following hold.

1. For all strings w, $d'(w) = \mathtt{bc}'(w) + \mathtt{sa}'(w)$.

2. $\mathtt{bc}'(\lambda) = d_2(\lambda)$.

3. $\mathtt{sa}'(\lambda) = 0$.

The construction is an adaptation of the construction in [24, 33]. Using this, we can build a P-measure of impossibility p which witnesses that X is s_2-improbable wrt P.

Consider $d^{(n)} : \Sigma^* \to \mathbb{Q}$ defined in Algorithm 2.

In Algorithm 2, we have, for every string $w \in \{0,1\}^{<n}$, $d^{(n)}$ behaves exactly the same as $\alpha_n d_2$.

For $wb \in \Sigma^{=n}$, we have $val_{wb} = d_2(wb)$. We have that half of the betting capital at wb is transferred to the the savings account of wb if $d_2(wb) > P^{(s_1-s_2)}(wb)$. This transfer maintains the property $\sum_{b\in\{0,1\}} d^{(n)}(wb)P^{s_2}(wb) = d^{(n)}(w)P^{s_2}(w)$. This transfer preserves the s_2-P-gale property of $d^{(n)}$.

Also, for $w \in \{0,1\}^{\geq n}$ and $b \in \{0,1\}$, in lines 8–13,

$$\sum_{b\in\{0,1\}} \mathtt{bc}'_n(w)P^{1-s_2}(C_{wb}|C_w)P^{s_2}(C_{wb}) = \mathtt{bc}'_n(w)\frac{P(C_w)}{P^{1-s_2}C_w}$$

$$= \mathtt{bc}'_n(w)P^{s_2}(C_w),$$

and

$$\sum_{b\in\{0,1\}} \mathtt{sa}'_n(w)\left[\frac{P^{1-s_2}(C_{wb})}{P^{1-s_2}(C_w)}\right]P^{s_2}(C_{wb}) = \sum_{b\in\{0,1\}} \mathtt{sa}'_n(w)\left[\frac{P(C_{wb})}{P^{1-s_2}(C_w)}\right]$$

$$= \mathtt{sa}'_n(w)P^{s_2}(C_w),$$

so the s_2-P-gale condition for $d^{(n)}$ is satisfied. Hence, each $d^{(n)}$ is an s_2-P-gale.

Denote $\alpha = \sum_{n=0}^{\infty} \alpha_n$. This is finite since $c < 1$. Define $d' = 1/\alpha \sum_{i=0}^{\infty} d^{(n)}$. Then $d'(\lambda) = d_2(\lambda)$. It follows that d' is an s_2-P-gale.

We note that P is a computable probability measure, s_2 is a rational number and d_2 is a constructive s_2-P-gale. For each n, the condition in line 8 can be verified by a lower semicomputation. It follows that each d_n is a constructive s_2-P-gale, and hence, so is d'.

We define $\mathtt{bc}'(w) = \sum_{i=0}^{|w|-1} \mathtt{bc}'_i(w)$. Let $\mathtt{sa}'(w) = \sum_{i=0}^{|w|-1} \mathtt{sa}'_i(w)$. Each $\mathtt{sa}_n$ is lower semicomputable by a computation that is 0 until the condition in line 8 is true, and converges from below to $\frac{\alpha_n val_w}{2}$ when the condition is true. Hence $\mathtt{sa}'$ is also lower semicomputable. Each $\mathtt{sa}_n$ is lower semicomputable by a computation that is 0 until the condition in line 8 is true, and converges from below to $\frac{\alpha_n val_w}{2}$ when the condition is true. Hence $\mathtt{sa}'$ is also lower semicomputable. Unlike the martingale case, we cannot say that $\mathtt{sa}'(w)$ is monotone increasing in the length of w. However, we can use the remark previously noted, to construct a martingale which for every w is defined as $\mathtt{bc}'(w)P^{s_2-1}(C_w) + \mathtt{sa}'(w)P^{s_2-1}(C_w)$.

We prove that for each n, $\mathtt{sa}'_n(w)P^{(s_2-1)}(C_w)$ is constant for all w with at least n bits. To see this, let w be a string such that $|w| \geq n$. If $d_2(w[0\ldots n-1]) > P^{s_1-s_2}(C_{w[0\ldots n-1]})$, then we have

$$\begin{aligned}
\mathtt{sa}'_n(w)P(C_w) &= \frac{\alpha_n}{2} val_w P^{(s_2-1)}(C_w) \\
&= \frac{\alpha_n}{2} d_2(w[0\ldots n-1]) P^{(1-s_2)}(C_w|C_{w[0\ldots n-1]}) P^{(s_2-1)}(C_w) \\
&= \frac{\alpha_n}{2} d_2(w[0\ldots n-1]) \frac{1}{P^{(1-s_2)}(C_{w[0\ldots n-1]})},
\end{aligned}$$

which is a constant. If, $d_2(w[0\ldots n-1]) \leq P^{s_1-s_2}(C_{w[0\ldots n-1]})$, then $\mathtt{sa}'_n$ is zero, so $\mathtt{sa}'_n(w)P^{(s_2-1)}(C_w)$ is zero.

We can see that $\mathtt{sa}'(w)P^{s_2-1}(C_w)$ is monotone increasing with the length of w. This is because for each i, $\mathtt{sa}'_i$ appears in the summation of $\mathtt{sa}'$ only on strings which are at least i bits long.

We define $\overline{p} : \Sigma^* \to \mathbb{R}^+$ by

$$\overline{p}(w) = \mathtt{sa}'(w)P^{s_2-1}(C_w),$$

for finite strings w.

Now, we define $p : \mathbf{C} \to [0, \infty]$ by

$$p(\omega) = \lim_{n\to\infty} \overline{p}(\omega[0\ldots n-1]).$$

Since $\mathtt{sa}'$ is lower semicomputable, it follows that p is lower semicomputable. Then we have $\int_{C_\lambda} pdP \leq d_2(\lambda)P(\mathbf{C}) = 1$. Thus p is a measure of impossibility.

Also,

$$\frac{\int_{C_w} p(\omega)dP}{P^{s_2}(C_w)} \geq \mathtt{sa}'(w)$$

for all strings w.

We have to analyze the behavior of $\mathtt{sa}'$ on sequences where d_1 succeeds. We can observe the following fact. If the s_1-gale d_1 succeeds on ω, then for every i, there is a least n such that $d_1(\omega[0\ldots n-1]) \geq 2[1/c]^i$. Then, at this length n, $d_2(\omega[0\ldots n-1]) \geq 2[1/c]^i P(\omega[0\ldots n-$

$1])^{(s_1-s_2)}$. This quantity is at least $[1/c]^{i+(s_2-s_1)n}$. Then $\mathtt{sa}'_n(\omega[0\dots n-1]) > [1/c]^{i+(s_2-s_1)n/2}$, so $\mathtt{sa}'(\omega[0\dots n-1]) > [1/c]^{\Omega(n)}$.

So we have that

$$\limsup_{n\to\infty} d_1(\omega[0\dots n-1]) = \infty \Leftrightarrow \limsup_{n\to\infty} \mathtt{sa}'(\omega[0\dots n-1]) = \infty.$$

If d_1 strongly succeeds on ω, then for each i, for all large enough n, we have $d_1(\omega[0\dots n-1]) > 2[1/c]^i$, which implies $d_2(\omega[0\dots n-1]) \geq 2[1/c]^i P(\omega[0\dots n-1])^{(s_1-s_2)}$, and hence $\mathtt{sa}'_n(\omega[0\dots n-1]) > [1/c]^{i+(s_2-s_1)n-0.5(s_2-s_1)n} = [1/c]^{\Omega(n)}$. This implies that, $\liminf_{n\to\infty} \mathtt{sa}'(\omega[0\dots n-1]) = \infty$.

Thus,

$$\liminf_{n\to\infty} d_1(\omega[0\dots n-1]) = \infty \Leftrightarrow \liminf_{n\to\infty} \mathtt{sa}'(\omega[0\dots n-1]) = \infty.$$

Hence

$$\limsup_{n\to\infty} d(\omega[0\dots n-1]) = \infty \quad \Longrightarrow \quad \limsup_{n\to\infty} \frac{\int_{C_{\omega[0\dots n-1]}} p(\omega)dP}{P^{s_2}(C_{\omega[0\dots n-1]})} = \infty.$$

Similarly,

$$\liminf_{n\to\infty} d(\omega[0\dots n-1]) = \infty \quad \Longrightarrow \quad \liminf_{n\to\infty} \frac{\int_{C_{\omega[0\dots n-1]}} p(\omega)dP}{P^{s_2}(C_{\omega[0\dots n-1]})} = \infty.$$

Thus $s_2 \in \mathcal{P}_{\rm str}(X)$ if $s_2 \in \mathcal{G}_{\rm str}(X)$, and $s_2 \in \mathcal{P}(X)$ if $s_2 \in \mathcal{G}(X)$.

If $s = 1$, the condition that $\limsup_{n\to\infty} \frac{\int_{C(\omega[0\dots n-1])} pdP}{P^s(C_{\omega[0\dots n-1]})} = \limsup_{n\to\infty} E_P[p|C_{\omega[0\dots n-1]}]$ is similar to the case of conversion of the martingale case.

If $s > 1$, and let $\epsilon > 0$ be arbitrary. then we pick a rational number s_2 such that $s < s_2 < s+\epsilon$. The s_2-gale d_2 which bets, for every string w and every symbol b, in the proportion $P(C_{wb}|C_w)^{1-s_2}$ succeeds on the Cantor Space $\mathbf{C}$. We define $\overline{p}(\omega[0\dots n-1]) = d_2(\omega[0\dots n-1])P^{s_2-1}(\omega[0\dots n-1])$, a monotone function in n. Then $p(\omega) = \lim_{n\to\infty} p(\omega[0\dots n-1])$ is a measure of impossibility. Then

$$\frac{\int_{\omega[0\dots n-1]} pdP}{P^{s_2}(C_{\omega[0\dots n-1]})} = d_2(\omega[0\dots n-1]),$$

so ω is s_2-improbable with respect to P if $\omega \in S^\infty[d_2]$, and ω is strongly s_2-improbable with respect to P if $\omega \in S^\infty_{\rm str}[d_2]$.

Conversely, let $s \in \mathcal{P}(X)$. Then there exists a measure of impossibility p such that X is s-improbable with respect to p. We define a martingale $d : \Sigma^* \to \mathbb{R}^+$ as follows. For finite strings w,

$$d(w) \quad = \quad E_P[p|C_w]P^{1-s}(C_w).$$

It is routine to see that d is a lowersemicomputable s-P-gale. Moreover,

$$\limsup_{n\to\infty} \frac{\int_{C_{\omega[0\ldots n-1]}} pdP}{P^s(C_{\omega[0\ldots n-1]})} = \infty \quad \Longrightarrow \limsup\nolimits_{n\to\infty} d(\omega[0\ldots n-1]) = \infty.$$

Thus, we can see that $s \in \mathcal{G}(X)$. Similarly, we can establish that if $s \in \mathcal{P}_{\mathrm{str}}(X)$, then $s \in \mathcal{G}_{\mathrm{str}}(X)$.

□

Using this, we can characterize effective Hausdorff and packing dimensions in the following way.

Corollary 4.6.3. *(Alternate Characterization of Constructive Dimension) For any set* $X \subseteq \mathbf{C}$*, the constructive* Hausdorff P-dimension *of* $X = \inf \mathcal{P}(X)$*, and the* constructive P-packing *dimension of* $X = \inf \mathcal{P}_{str}(X)$*.*

4.7 Scaled Dimension

In this section, we give an alternate characterization of constructive scaled dimension, introduced by Hitchcock, Lutz and Mayordomo [31]. The discussion in this section focuses on Lebesgue measure, since the theory of scaled dimension has been developed so far for the Lebesgue measure (or equivalently, uniform probability distribution).

Scaled dimension is a generalization of classical Hausdorff and classical packing dimension. Just as Hausdorff dimension and packing dimension allows us to make distinctions between measure 0 sets, scaled dimension allows us to make distinctions between dimension 0 sets. The motivation for studying scaled dimension comes from complexity theory and its applications, like cryptography, where many interesting classes have resource-bounded dimension 0 in the appropriate resource-bounded setting (for example, ESPACE) [34].

However, we give a characteracterization of scaled dimension at the *constructive* level, using a suitable generalization of the characterization for constructive dimension, given in the previous section. We first give the notion of a scale, and then introduce the notion of a scaled improbability, and use this notion to give a new characterization of scaled dimension.

Definition. (Hitchcock, Lutz, Mayordomo, 2004 [31]) A *scale* is a continuous function $g : H \times \mathbb{R} \to \mathbb{R}$ with the following properties.

1. $H = (a, \infty)$ for some $a \in \mathbb{R} \cup \{\infty\}$.

2. $g(m, 1) = m$ for all $m \in H$.

3. $g(m, 0) = g(m', 0) \geq 0$ for all $m, m' \in H$.

4. For every sufficiently large $m \in H$, the function $s \mapsto g(m, s)$ is nonnegative and strictly increasing.

5. For all $s' > s \geq 0$, $\lim_{m\to\infty}[g(m, s') - g(m, s)] = \infty$.

Notation Let g be a scale. Then for any $j \in H$, we define

$$\delta_j(s_1, s_2) \quad = \quad 2^{g(j,s_2)-g(j,s_1)}.$$

If $s_1 < s_2$, then $\delta_j(s_1, s_2) > 1$.

We deal with computable rapidly growing scales in this section. We give the definition of a scaled gale.

Definition. (Hitchcock, Lutz and Mayordomo [31]) Let $g : H \times \mathbb{R} \to \mathbb{R}$ be a computable scale, and $s \in [0, \infty)$.

A g-scaled s-gale (briefly, an $s^{(g)}$-*gale*) is a function $d : \Sigma^* \to [0, \infty)$ such that for all $w \in \Sigma^*$ with $|w| \in H$,

$$d(w)2^{-g(|w|,s)} \quad = \quad d(w0)2^{-g(|w0|,s)} + d(w1)2^{-g(|w1|,s)}. \tag{4.7.1}$$

An s-gale (whose definition is given in section 6) is a g-scaled s-gale where $g(n, s) = ns$ for all s and n.

The notion of success and strong success of a g-scaled s-gale are exactly the lim sup and the lim inf success defined in section 6. Therefore, we do not reproduce it here.

We now introduce a notion of scaled improbability, based on the notion of the scale functions defined above.

Definition. Let $X \subseteq \mathbf{C}$. We say that X is $s^{(g)}$*-improbable* with respect to a measure of impossibility[2] $p : \mathbf{C} \to \mathbb{R}^+$ if for every binary sequence $\omega \in X$, we have

$$\limsup_{n\to\infty} \frac{\int_{C\omega[0\ldots n-1]} p(\omega)d\omega}{2^{-g(n,s)}} \quad = \quad \infty. \tag{4.7.2}$$

We say X is *strongly* $s^{(g)}$*-improbable* with respect to p if for every binary sequence $\omega \in X$, we have

$$\liminf_{n\to\infty} \frac{\int_{C\omega[0\ldots n-1]} p(\omega)d\omega}{2^{-g(n,s)}} \quad = \quad \infty. \tag{4.7.3}$$

We introduce some notation which is used in the definition of scaled dimension. We recall the standard definitions of scaled dimension, and scaled strong dimension, and then give an alternate characterization of scaled dimension using the concept of scaled improbability.

Notation. For $X \subseteq \mathbf{C}$, let $\mathcal{G}^{(g)}(X)$ be the set of all $s \in [0, \infty)$ such that there is an $s^{(g)}$-gale d for which $X \subseteq S^\infty[d]$ and $\mathcal{P}(X)$ be the set of all $s \in [0, \infty)$ such that X is $s^{(g)}$-improbable with respect to some measure of impossibility. Similarly, let $\mathcal{G}^{(g)}_{\text{str}}(X)$ be the set of all $s \in [0, \infty)$ such that there is an $s^{(g)}$-gale d for which $X \subseteq S^\infty_{\text{str}}[d]$ and $\mathcal{P}_{\text{str}}(X)$ be the set of all $s \in [0, \infty)$ such that X is strongly $s^{(g)}$-improbable with respect to some P-measure of impossibility.

Definition. (Hitchcock, Lutz, Mayordomo [31]) If g is a scale, then the g-scaled dimension of a set $X \subseteq \mathbf{C}$ is $\dim^{(g)}(X) = \inf \mathcal{G}^{(g)}(X)$. The g-scaled strong dimension of X is $\mathrm{Dim}^{(g)}(X) = \inf \mathcal{G}^{(g)}(X)$.

We introduce the notion of a rapidly growing scale.

Definition. Let p be the smallest natural number in H. A rapidly growing scale is a scale such that for any s, s' where $s' > s$, there is a constant $0 < k < 1$ such that $\sum_{m=p}^{\infty} 2^{-k\delta_m(s',s)}$ is finite.

---

[2]The probability measure is the uniform probability measure.

The above notion is reminiscent of the notion of strong positivity of the probability measure dealt with in the preceding sections, since strong positivity also ensures convergence of an infinite series necessary in the construction of the countable family of martingales.

The following lemma is crucial in the alternate characterization of scaled dimension. The constructions we use in the proof of the lemma are exactly analogous to those in the proof of Lemma 4.6.2 (in fact, in the case of the uniform probability measure, the constructions in Lemma 4.6.2 are special cases of the following constructions when the scale is $g(n, s) = ns$.).

Lemma 4.7.1. *Let g be a computable rapidly growing scale, $s \in [0, \infty)$ and $X \subseteq \mathbf{C}$. Then $s \in \mathcal{G}^{(g)}(X)$ if and only if $s \in \mathcal{P}^{(g)}(X)$, and $s \in \mathcal{G}^{(g)}_{str}(X)$ if and only if $s \in \mathcal{P}^{(g)}_{str}(X)$.*

Proof. Let g be a computable rapidly growing scale.

Let $s_1 \in [0, \infty)$ be a rational such that $g(n, s_1) < n$ for all $n \in \mathbb{N}$, when defined, and $\epsilon > 0$ be arbitrary. Let $s_2 \in [0, \infty)$ be such that for every n where both are defined, $\epsilon f(n) < g(n, s_2) - g(n, s_1)$, and $g(n, s_2) < n$. Let p be the smallest natural number such that both $g(p, s_1)$ and $g(p, s_2)$ are defined, and $M > 2^{-g(p,s_2)-g(p,s_1)}$. Let $0 < k < 1$ be such that $\sum_{m=p}^{\infty} 2^{-k\delta_n(s_2,s_1)} < \infty$. Let $d_1 : \Sigma^* \to \mathbb{R}$ be a g-scaled s_1-gale. Then $d_2 : \{0, 1\}^* \to \mathbb{R}$ defined by $d_2(w) = d_1(w)2^{g(|w|,s_2)-g(|w|,s_1)}$ is a g-scaled s_2-gale.

We will construct the following g-scaled s_2-gale, which transfers some capital into a "savings account". We will use this gale to build the $s_2^{(g)}$-measure of improbability. The gale $d^{(n)} : \Sigma^* \to \mathbb{R}$ is defined in Algorithm 3.

Let $\alpha = \sum_{i=p}^{\infty} 2^{-k\delta_i(s_2,s_1)}$. Then, as in the proof of Lemma 20, that $d' = \frac{1}{\alpha}\sum_{i=0}^{\infty} d^{(n)}$ is a constructive g-scaled s_2-gale.

Let $\mathsf{sa}'(w) = \sum_{i=0}^{|w|-1} \mathsf{sa}^{(n)\prime}(w)$. We define $\overline{p} : \Sigma^* \to [0, \infty]$ by

$$\overline{p}(w) = \begin{cases} \mathsf{sa}'(w)2^{-g(|w|,s_2)+|w|} & \text{if g is defined for } |w| \\ \mathsf{sa}'(w) & \text{otherwise.} \end{cases}$$

for strings w.

We define the function $p : \mathbf{C} \to [0, \infty]$ by

$$p(\omega) = \lim_{n\to\infty} \overline{p}(\omega[0 \ldots n-1]).$$

Figure 4.7.1 Algorithm 3

procedure $d^{(n)}(w)$
  $\alpha_n = 2^{-k\delta_n(s_2,s_1)}$
  if $g(n, s_2)$ is not defined **then**
    $\mathtt{bc}^{(n)'}(w) \leftarrow \alpha_n d_2(w)$
    $\mathtt{sa}^{(n)'}(w) \leftarrow 0$
    return $\mathtt{bc}^{(n)'}(w) + \mathtt{sa}^{(n)'}(w)$
  end if
  if $|w| < n$ **then**
    $\mathtt{bc}^{(n)'}(w) \leftarrow \alpha_n d_2(w)$
    $\mathtt{sa}^{(n)'}(w) \leftarrow 0.$
  else
    $val_w = d_2(w[0 \dots n-1])2^{g(j,s_2)-g(j-1,s_2)-1}$
    if $d_2(w[0 \dots n-1]) > 2^{\delta_n(s_2,s_1)}$ **then**
      $\mathtt{bc}^{(n)'}(w), \mathtt{sa}^{(n)'}(w) \leftarrow \alpha_n val_w/2$
    else
      $\mathtt{bc}^{(n)'}(w) \leftarrow \alpha_n val_w$
      $\mathtt{sa}^{(n)'}(w) \leftarrow 0$
    end if
  end if
  return $\mathtt{bc}^{(n)'}(w) + \mathtt{sa}^{(n)'}(w)$
end procedure

As in Lemma 4.6.2, if d is lower semicomputable, it is routine to verify that p is a measure of impossibility.

Also,

$$\frac{\int_{C(\omega[0\ldots n-1])} pdP}{2^{-g(n,s_2)}} \quad \geq \quad \mathtt{sa}'(\omega[0\ldots n-1]).$$

If d_1 succeeds on a binary sequence ω, then there is an n for which $d_2(\omega[0\ldots n-1]) > 2^{\delta_n(s_2,s_1)}$. Then, $\mathtt{sa}^{(n)\prime}(\omega[0\ldots n-1]) \geq M^{-1}2^{(1-k)\delta_n(s_2,s_1)}$. Hence,

$$\limsup_{n\to\infty} d_1(\omega[0\ldots n-1]) = \infty \Leftrightarrow \limsup_{n\to\infty} \mathtt{sa}'(\omega[0\ldots n-1]) = \infty.$$

If d_1 strongly succeeds on ω, then for all large enough lengths n, we have $\mathtt{sa}^{(n)\prime}(\omega[0\ldots n]) \geq M^{-1}2^{(1-k)\delta_n(s_2,s_1)}$, so

$$\liminf_{n\to\infty} d_1(\omega[0\ldots n-1]) = \infty \Leftrightarrow \liminf_{n\to\infty} \mathtt{sa}'(\omega[0\ldots n-1]) = \infty.$$

If, $g(n,s) = n$ for all n, then any g-scaled s-gale is a martingale. If there is a martingale that succeeds on ω, there is a measure of impossibility p to witness that ω is $s^{(g)}$-improbable with respect to P, and if there is a martingale that strongly succeeds on ω, then there is a measure of impossibility p to witness that ω is $s^{(g)}$-improbable.

If $g(n,s) > n$ for all large enough n, then the gale $d(w) = 2^{g(|w|,s)-g(|w|-1,s)}$ strongly succeeds on $\mathbf{C}$. Then, if we define $p(\omega) = \lim_{n\to\infty} d(\omega[0\ldots n-1])2^{-g(n,s)+g(n-1,s)-1}$, we can prove that p is a measure of impossibility such that $\mathbf{C}$ is strongly $s^{(g)}$-improbable.

Hence, if $s \in \mathcal{G}^{(g)}(X)$, we can conclude that

$$\limsup_{n\to\infty} \frac{\int_{C_{\omega[0\ldots n-1]}} p(\omega)d\omega}{2^{-g(n,s)}} = \infty,$$

and if $s \in \mathcal{G}^{(g)}_{str}(X)$, we can conclude that

$$\liminf_{n\to\infty} \frac{\int_{C_{\omega[0\ldots n-1]}} p(\omega)d\omega}{2^{-g(n,s)}} = \infty.$$

Conversely, let $s \in \mathcal{P}^{(g)}(X)$. Then there is a measure of impossibility p such that X is $s^{(g)}$-improbable with respect to p. Define a function $d : \Sigma^* \to \mathbb{R}$ as follows. For finite strings w,

$$d(w) = E[p|C_w]2^{g(|w|,s)-|w|}$$

It is routine to show that d is a lower semicomputable $s^{(g)}$-gale. Moreover, as in Lemma 4.6.2,

$$\limsup_{n\to\infty} \frac{\int_{C_{\omega[0\dots n-1]}} p(x)dx}{2^{-g(n,s)}} = \limsup_{n\to\infty} d(\omega[0\dots n-1]).$$

Thus, we can see that $s \in \mathcal{G}^{(g)}(X)$. Similarly, we can establish that if $s \in \mathcal{P}^{(g)}_{str}(X)$, then $s \in \mathcal{G}^{(g)}_{str}(X)$. □

Corollary 4.7.2. *(Alternate Characterization of Constructive Scaled Dimension) For any scale g, for any set $X \subseteq \mathbf{C}$, the constructive g-scaled dimension of X is* $\inf \mathcal{P}^{(g)}(X)$*, and the constructive g-scaled strong dimension of X is* $\inf \mathcal{P}^{(g)}_{str}(X)$*.*

CHAPTER 5. Finite-state dimension and real arithmetic

"Anyone who attempts to generate random numbers by deterministic means is, of course, living in a state of sin" - *von Neumann*

5.1 Summary

This is joint work with Dave Doty and Jack Lutz [21]. Much of the content of this chapter is taken from that work. The finite-state dimension of an infinite sequence S over a finite alphabet Σ is an asymptotic measure of the density of information in S as perceived by finite-state automata. This quantity, denoted $\dim_{\mathrm{FS}}(S)$, is a finite-state effectivization of classical Hausdorff dimension introduced by Dai, Lathrop, Lutz, and Mayordomo. A dual quantity, the finite-state *strong* dimension of S, denoted $\mathrm{Dim}_{\mathrm{FS}}(S)$, is a finite-state effectivization of classical packing dimension introduced by Athreya, Hitchcock, Lutz, and Mayordomo . The study of finite-state dimension is related to the study of normal numbers, because the set of normal numbers is exactly the set of numbers with finite-state dimension 1.

This chapter initiates the study of operations that preserve finite-state dimension and finite-state strong dimension. This study is related to, but distinct from, the study of operations that preserve normality. It is clear that every operation that preserves finite-state dimension must also preserve normality, but the converse does not hold. For example, a subsequence selected from a sequence according an arithmetical progression need not have the same finite-state dimension as the original sequence. This is because a sequence with finite-state dimension less than 1 may have its information content distributed heterogeneously. Specifically, given a normal sequence S over the alphabet $\{0,1\}$, define a sequence T whose n^{th} bit is the $\frac{n}{2}^{\mathrm{th}}$ bit of S if n is even and 0 otherwise. Then the sequence S and the constant sequence 0^∞

are both selected from T according to arithmetic progressions, but it is easy to verify that $\dim_{\mathrm{FS}}(T) = \mathrm{Dim}_{\mathrm{FS}}(T) = \frac{1}{2}, \dim_{\mathrm{FS}}(0^\infty) = \mathrm{Dim}_{\mathrm{FS}}(0^\infty) = 0$, and $\dim_{\mathrm{FS}}(S) = \mathrm{Dim}_{\mathrm{FS}}(S) = 1$. Hence, Wall's first above-mentioned theorem does not extend to the preservation of finite-state dimension. Of course, this holds *a fortiori* for the stronger results by Agafonov, Kamae, and Weiss.

Our main theorem states that Wall's second above-mentioned theorem, unlike the first one, does extend to the preservation of finite-state dimension. That is, we prove that, for every integer $k \geq 2$, every nonzero rational number q, and every real number α, the base-k expansions of $\alpha, q + \alpha$, and $q\alpha$ all have the same finite-state dimension and the same finite-state strong dimension.

In this, we use Schur concavity and entropy rates to show that the finite-state dimension of a number is unchanged by the addition with, or multiplication by, a non-zero rational number. Since it is known that the set of *Borel-normal numbers* is exactly the set of sequences with finite state dimension 1, this result extends, and gives a new proof of, Wall's result that multiplication by a non-zero rational number preserves normality.

5.2 Introduction

The finite-state dimension of an infinite sequence S over a finite alphabet Σ is an asymptotic measure of the density of information in S as perceived by finite-state automata. This quantity, denoted $\dim_{\mathrm{FS}}(S)$, is a finite-state effectivization of classical Hausdorff dimension [29, 23] introduced by Dai, Lathrop, Lutz, and Mayordomo [17]. A dual quantity, the finite-state *strong* dimension of S, denoted $\mathrm{Dim}_{\mathrm{FS}}(S)$, is a finite-state effectivization of classical packing dimension [77, 76, 23] introduced by Athreya, Hitchcock, Lutz, and Mayordomo [2]. (Explicit definitions of $\dim_{\mathrm{FS}}(S)$ and $\mathrm{Dim}_{\mathrm{FS}}(S)$ appear in section 5.3.) In fact both $\dim_{\mathrm{FS}}(S)$ and $\mathrm{Dim}_{\mathrm{FS}}(S)$ are asymptotic measures of the density of finite-state information in S, with $0 \leq \dim_{\mathrm{FS}}(S) \leq \mathrm{Dim}_{\mathrm{FS}}(S) \leq 1$ holding in general. The identity $\dim_{\mathrm{FS}}(S) = \mathrm{Dim}_{\mathrm{FS}}(S)$ holds when S is sufficiently "regular", but, for *any* two real numbers $0 \leq \alpha \leq \beta \leq 1$, there exists a sequence S with $\dim_{\mathrm{FS}}(S) = \alpha$ and $\mathrm{Dim}_{\mathrm{FS}}(S) = \beta$ [27].

Although finite-state dimension and finite-state strong dimension were originally defined in terms of finite-state gamblers [17, 2] (following the gambling approach used in the first effectivizations of classical fractal dimension [55, 56]), they have also been shown to admit equivalent definitions in terms of information-lossless finite-state compressors [17, 2], finite-state predictors in the log-loss model [32, 2], and block-entropy rates [10]. In each case, the definitions of $\dim_{\rm FS}(S)$ and $\mathrm{Dim}_{\rm FS}(S)$ are exactly dual, differing only in that a limit inferior appears in one definition where a limit superior appears in the other. These two finite-state dimensions are thus, like their counterparts in fractal geometry, robust quantities and not artifacts of a particular definition.

The sequences S satisfying $\dim_{\rm FS}(S) = 1$ are precisely the *(Borel) normal* sequences, i.e., those sequences in which each nonempty string $w \in \Sigma^*$ appears with limiting frequency $|\Sigma|^{-|w|}$. (This fact was implicit in the work of Schnorr and Stimm [71] and pointed out explicitly in [10].) The normal sequences, introduced by Borel in 1909 [8], were extensively investigated in the twentieth century [65, 44, 87, 19, 28]. Intuitively, the normal sequences are those sequences that are random relative to finite-state automata. This statement may seem objectionable when one first learns that the Champernowne sequence

$$0100011011000001010011100\ldots,$$

obtained by concatenating all binary strings in standard order, is normal [15], but it should be noted that a finite-state automaton scanning this sequence will spend nearly all its time in the middle of long strings that are random in the (stronger) sense of Kolmogorov complexity [47] and, having only finite memory, will have no way of "knowing" where such strings begin or end. This perspective is especially appropriate when modeling situations in which a data stream is truly massive relative to the computational resources of the entity processing it.

An informative line of research on normal sequences concerns operations that preserve normality. For example, in his 1949 Ph.D. thesis under D.H. Lehmer, Wall [84] proved that every subsequence that is selected from a normal sequence by taking all symbols at positions occurring in a given arithmetical progression is itself normal. Agafonov [1] extended this by showing that every subsequence of a normal sequence that is selected using a regular language is itself normal; Kamae [35] and Kamae and Weiss [36] proved related results; and Merkle and Reimann [62] proved that a subsequence selected from a normal sequence using a context-free language need not be normal (in fact, can be constant, even if selected by a one-counter language). For another example, again in his thesis, Wall [84] (see also [44, 9]) proved that, for every integer $k \geq 2$, every nonzero rational number q, and every real number α that is normal base k (i.e., has a base-k expansion that is a normal sequence), the sum $q + \alpha$ and the product $q\alpha$ are also normal base k. (It should be noted that a real number α may be normal in one base but not in another [13, 67].)

This chapter initiates the study of operations that preserve finite-state dimension and finite-state strong dimension. This study is related to, but distinct from, the study of operations that preserve normality. It is clear that every operation that preserves finite-state dimension must also preserve normality, but the converse does not hold. For example, a subsequence selected from a sequence according an arithmetical progression need not have the same finite-state dimension as the original sequence. This is because a sequence with finite-state dimension less than 1 may have its information content distributed heterogeneously. Specifically, given a normal sequence S over the alphabet $\{0, 1\}$, define a sequence T whose n^{th} bit is the $\frac{n}{2}^{\text{th}}$ bit of S if n is even and 0 otherwise. Then the sequence S and the constant sequence 0^∞ are both selected from T according to arithmetic progressions, but it is easy to verify that $\dim_{\text{FS}}(T) = \text{Dim}_{\text{FS}}(T) = \frac{1}{2}, \dim_{\text{FS}}(0^\infty) = \text{Dim}_{\text{FS}}(0^\infty) = 0$, and $\dim_{\text{FS}}(S) = \text{Dim}_{\text{FS}}(S) = 1$. Hence, Wall's first above-mentioned theorem does not extend to the preservation of finite-state dimension. Of course, this holds *a fortiori* for the stronger results by Agafonov, Kamae, and Weiss.

Our main theorem states that Wall's second above-mentioned theorem, unlike the first one, does extend to the preservation of finite-state dimension. That is, we prove that, for every integer $k \geq 2$, every nonzero rational number q, and every real number α, the base-k expansions of $\alpha, q + \alpha$, and $q\alpha$ all have the same finite-state dimension and the same finite-state strong dimension.

The proof of our main theorem does not, and probably cannot, resemble Wall's uniform distribution argument. Instead we use Bourke, Hitchcock, and Vinodchandran's block-entropy rate characterizations of $\dim_{\mathrm{FS}}$ and $\mathrm{Dim}_{\mathrm{FS}}$ [10], coupled with the Schur concavity of the entropy function [72, 59, 4], to prove that finite-state dimension and finite-state strong dimension are contractive functions with respect to a certain "logarithmic block dispersion" pseudometric that we define on the set of all infinite k-ary sequences. (A function is contractive if the distance between its values at sequences S and T is no more than the pseudodistance between S and T.) This gives a general method for bounding the difference between the finite-state dimensions, and the finite-state strong dimensions, of two sequences. We then use this method to prove our main theorem. In particular, this gives a new proof of Wall's theorem on the sums and products of rational numbers with normal numbers.

In summary, our main result is a fundamental theorem on finite-state dimension that is a quantitative extension of a classical theorem on normal numbers but requires a different, more powerful proof technique than the classical theorem.

5.3 Preliminaries

Throughout this chapter, $\Sigma = \{0, 1, \ldots, k-1\}$, where $k \geq 2$ is an integer.

A *base-k expansion* of a real number $\alpha \in [0, 1]$ is a sequence $S \in \Sigma^\infty$ such that

$$\alpha = \sum_{n=0}^{\infty} S[n]k^{-(n+1)}.$$

A sequence $S \in \Sigma^\infty$ is *(Borel) normal* if, for every nonempty string $w \in \Sigma^+$

$$\lim_{n\to\infty} \frac{1}{n} \left|\left\{ u \in \Sigma^{<n} \middle| uw \sqsubseteq S \right\}\right| = |\Sigma|^{-|w|},$$

i.e., if each string w appears with asymptotic frequency $k^{-|w|}$ in S.

If Ω is a nonempty finite set, we write $\Delta(\Omega)$ for the set of all (discrete) probability measures on Ω, i.e., all functions $\pi : \Omega \to [0, 1]$ satisfying $\sum\limits_{w\in\Omega} \pi(w) = 1$. We write $\Delta_n = \Delta(\{1, \ldots, n\})$.

All logarithms in this paper are base 2. The *Shannon entropy* of a probability measure $\pi \in \Delta(\Omega)$ is

$$H(\pi) = \sum_{w\in\Omega} \pi(w) \log \frac{1}{\pi(w)},$$

where $0 \log \frac{1}{0} = 0$.

We briefly define finite-state dimension and finite-state strong dimension. As noted in the introduction, several equivalent definitions of these dimensions are now known. In this paper, it is most convenient to use the definitions in terms of block-entropy rates, keeping in mind that Bourke, Hitchcock, and Vinodchandran [10] proved that these definitions are equivalent to earlier ones.

For nonempty strings $w, x \in \Sigma^+$, we write

$$\#_\Box(w,x) = \left| \left\{ m \leq \frac{|x|}{|w|} - 1 \;\middle|\; w = x[m|w| \ldots (m+1)|w|-1] \right\} \right|$$

for the number of *block occurrences* of w in x (when x is divided into consecutive blocks of length $|w|$). Note that $0 \leq \#_\Box(w,x) \leq \frac{|x|}{|w|}$.

For each sequence $S \in \Sigma^\infty$, positive integer n, and string $w \in \Sigma^{<n}$, the n^{th} *block frequency* of w in S is

$$\pi_{S,n}(w) = \frac{\#_\Box(w, S[0 \ldots n|w|-1])}{n}.$$

Note that, for all $S \in \Sigma^\infty$ and $0 < l < n$,

$$\sum_{w \in \Sigma^l} \pi_{S,n}(w) = 1,$$

i.e., $\pi^{(l)}_{S,n} \in \Delta(\Sigma^l)$, where we write $\pi^{(l)}_{S,n}$ for the restriction of $\pi_{S,n}$ to Σ^l.

For each sequence $S \in \Sigma^\infty$ and positive integer l, the l^{th} *normalized lower and upper block entropy rates* of S are

$$H^-_l(S) = \frac{1}{l \log k} \liminf_{n \to \infty} H\left(\pi^{(l)}_{S,n}\right)$$

and

$$H^+_l(S) = \frac{1}{l \log k} \limsup_{n \to \infty} H\left(\pi^{(l)}_{S,n}\right),$$

respectively.

Definition. Let $S \in \Sigma^\infty$.

1. The *finite-state dimension* of S is

$$\dim_{\text{FS}}(S) = \inf_{l \in \mathbb{Z}^+} H^-_l(S).$$

2. The *finite-state strong dimension* of S is

$$\mathrm{Dim}_{\mathrm{FS}}(S) = \inf_{l \in \mathbb{Z}^+} H_l^+(S).$$

More discussion and properties of these dimensions appear in the references cited in the introduction, but this material is not needed to follow the technical arguments in the present paper.

5.4 Logarithmic Dispersion and Finite-State Dimension

In this section we prove a general theorem stating that the difference between two sequences' finite-state dimensions (or finite-state strong dimensions) is bounded by a certain "pseudodistance" between the sequences. Recall that $\Delta_n = \Delta(\{1, \ldots, n\})$ is the set of all probability measures on $\{1, \ldots, n\}$.

We now define the log-dispersion between two probability measures π and μ to be the minimum "complexity" of a stochastic matrix mapping π to μ.

Definition. Let n be a positive integer. The *logarithmic dispersion* (briefly, the *log-dispersion*) between two probability measures $\pi, \mu \in \Delta_n$ is

$$\delta(\pi, \mu) = \log m,$$

where m is the least positive integer for which there is an $n \times n$ nonnegative real matrix $A = (a_{ij})$ with the following three properties.

(i) A is *stochastic*: each column of A sums to 1, i.e., $\sum_{i=1}^{n} a_{ij} = 1$ holds for all $1 \leq j \leq n$.

(ii) $A\pi = \mu$, i.e., $\sum_{j=1}^{n} a_{ij}\pi(j) = \mu(i)$ holds for all $1 \leq i \leq n$.

(iii) No row or column of A contains more than m nonzero entries.

It is clear that $\delta : \Delta_n \times \Delta_n \to [0, \log n]$. We now extend δ to a normalized function $\delta^+ : \Sigma^\infty \times \Sigma^\infty \to [0, 1]$. Recall the block-frequency functions $\pi_{S,n}^{(l)}$ defined in section 5.3.

Definition. The *normalized upper logarithmic block dispersion* between two sequences $S, T \in \Sigma^\infty$ is

$$\delta^+(S,T) = \limsup_{l\to\infty} \frac{1}{l \log k} \limsup_{n\to\infty} \delta\left(\pi_{S,n}^{(l)}, \pi_{T,n}^{(l)}\right).$$

Recall that a *pseudometric* on a set X is a function $d : X \times X \to \mathbb{R}$ satisfying the following three conditions for all $x, y, z \in X$.

(i) $d(x,y) \geq 0$, with equality if $x = y$. (nonnegativity)

(ii) $d(x,y) = d(y,x)$. (symmetry)

(iii) $d(x,z) \leq d(x,y) + d(y,z)$. (triangle inequality)

(A pseudometric is a *metric*, or *distance function*, on X if it satisfies (i) with "if" replaced by "if and only if".) The following fact must be known, but we do not know a reference at the time of this writing.

Lemma 5.4.1. *For each positive integer n, the log-dispersion function δ is a pseudometric on Δ_n.*

It is easy to see that S is not a metric on Δ_n for any $n \geq 2$. For example, if π is any nonuniform probability measure on $\{1, \ldots, n\}$ and μ obtained from π by permuting the values of π nontrivially, then $\pi \neq \mu$ but $\delta(\pi, \mu) = 0$.

Lemma 5.4.1 has the following immediate consequence.

Corollary 5.4.2. *The normalized upper log-block dispersion function δ^+ is a pseudometric on Σ^∞.*

If d is a pseudometric on a set X, then a function $f : X \to \mathbb{R}$ is *d-contractive* if, for all $x, y \in X$,

$$|f(x) - f(y)| \leq d(x,y),$$

i.e., the distance between $f(x)$ and $f(y)$ does not exceed the pseudodistance between x and y.

We prove the following lemma at the end of this section.

Lemma 5.4.3. *For each positive integer n, the Shannon entropy function $H : \Delta_n \to [0, \log n]$ is δ-contractive.*

The following useful fact follows easily from Lemma 5.4.3.

Theorem 5.4.4. *Finite-state dimension and finite-state strong dimension are δ^+-contractive. That is, for all $S, T \in \Sigma^\infty$,*

$$|\mathrm{dim}_{\mathrm{FS}}(S) - \mathrm{dim}_{\mathrm{FS}}(T)| \leq \delta^+(S, T)$$

and

$$|\mathrm{Dim}_{\mathrm{FS}}(S) - \mathrm{Dim}_{\mathrm{FS}}(T)| \leq \delta^+(S, T).$$

In this paper, we only use the following special case of Theorem 5.4.4.

Corollary 5.4.5. *Let $S, T \in \Sigma^\infty$. If*

$$\limsup_{n \to \infty} \delta\left(\pi_{S,n}^{(l)}, \pi_{T,n}^{(l)}\right) = o(l)$$

as $l \to \infty$, then

$$\mathrm{dim}_{\mathrm{FS}}(S) = \mathrm{dim}_{\mathrm{FS}}(T)$$

and

$$\mathrm{Dim}_{\mathrm{FS}}(S) = \mathrm{Dim}_{\mathrm{FS}}(T).$$

The proof of Lemma 5.4.3 uses Schur concavity [72, 59, 4], which we now review. We say that a vector $\vec{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is *nonincreasing* if $x_1 \geq \ldots \geq x_n$. If $\vec{x}, \vec{y} \in \mathbb{R}^n$ are nonincreasing, then we say that $\vec{x}$ *majorizes* $\vec{y}$, and we write $\vec{x}y$, if the following two conditions hold.

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

(ii) For all $1 \leq t \leq n$, $\sum_{i=1}^t x_i \geq \sum_{i=1}^t y_i$.

Given a vector $\vec{x} \in \mathbb{R}^n$ and a permutation π of $\{1, \ldots, n\}$, write $\pi(\vec{x}) = (x_{\pi(1)}, \ldots, x_{\pi(n)})$. Call a set $D \subseteq \mathbb{R}^n$ *symmetric* if $\pi(\vec{x}) \in D$ holds for every $\vec{x} \in D$ and every permutation π of $\{1, \ldots, n\}$. For $D \subseteq \mathbb{R}^n$, a function $f : D \to \mathbb{R}$ is then *symmetric* if D is symmetric and $f(\vec{x}) = f(\pi(\vec{x}))$ holds for every $\vec{x} \in D$ and every permutation π of $\{1, \ldots, n\}$.

Definition. Let $D \subseteq \mathbb{R}^n$ and $f : D \to \mathbb{R}$ be symmetric. Then f is *Schur-concave* if, for all $\vec{x}, \vec{y} \in \mathbb{R}^n$,

$$\vec{x}y \implies f(\vec{x}) \leq f(\vec{y}).$$

The set Δ_n of all probability measures on $\{1, \ldots, n\}$ can be regarded as the $(n-1)$-dimensional simplex

$$\Delta_n = \left\{ \vec{p} \in [0,1]^n \;\middle|\; \sum_{i=1}^{n} p_i = 1 \right\} \subseteq \mathbb{R}^n.$$

This set Δ_n is symmetric, as is the Shannon entropy function $H : \Delta_n \to [0, \log n]$. In fact, the following fundamental property of Shannon entropy is well known [4].

Lemma 5.4.6. *The Shannon entropy function $H : \Delta_n \to [0, \log n]$ is Schur-concave.*

We now use Lemma 5.4.6 to prove Lemma 5.4.3.

Proof of Lemma 5.4.3. Fix a positive integer n, and let $\vec{p}, \vec{q} \in \Delta_n$. By the symmetry of δ (established in Lemma 5.4.1), it suffices to prove that

$$H(\vec{p}) \leq H(\vec{q}) + \delta(\vec{p}, \vec{q}). \tag{5.4.1}$$

Without loss of generality, assume that $\vec{p}$ and $\vec{q}$ are nonincreasing. Let m be the positive integer such that $\delta(\vec{p}, \vec{q}) = \log m$, and let $A = (a_{ij})$ be an $n \times n$ matrix testifying to the value of $\delta(\vec{p}, \vec{q})$. Define an $n \times n$ matrix $B = (b_{ij})$ by

$$b_{ij} = \begin{cases} 1 & \text{if } (i-1)m < j \leq \min\{im, n\}; \\ 0 & \text{otherwise.} \end{cases}$$

That is, the first block of m entries in the first row of B are 1's, the second block of m entries in the second row of B are 1's, and so on, until the last $n - m\left(\left\lceil \frac{n}{m} \right\rceil - 1\right)$ entries in the $\left\lceil \frac{n}{m} \right\rceil^{\text{th}}$ row of B are 1's.

Let $\vec{r} = B\vec{p}$. Intuitively, B represents the "worst-case" matrix with no more than m nonzero entries in each row and column, in the sense that, of all matrices with no more than m nonzero entries in each row and column, it produces from $\vec{p}$ the vector $\vec{r}$ with the lowest possible entropy. More formally, we show that $\vec{r}$ majorizes the vector $\vec{q}$, and thus $\vec{r}$ has entropy at most that of $\vec{q}$. However, since B is limited to m nonzero entries in each row and column, it cannot redistribute the values in $\vec{p}$ by too much, so the entropy of $\vec{r}$ will be close to that of $\vec{p}$.

Since B is stochastic (because each column contains exactly one 1) and $\vec{p} \in \Delta_n$, we have $\vec{r} \in \Delta_n$. Clearly, $\vec{r}$ is nonincreasing. For each $1 \leq i \leq n$, let $C_i = \{j \mid a_{ij} > 0\}$, noting that $|C_i| \leq m$. Then, for all $1 \leq t \leq n$,

$$
\begin{aligned}
\sum_{i=1}^{t} r_i &= \sum_{i=1}^{t}\sum_{j=1}^{n} b_{ij}p_j = \sum_{i=1}^{t}\sum_{j=(i-1)m+1}^{\min\{im,n\}} p_j = \sum_{i=1}^{\min\{tm,n\}} p_i \\
&\geq \sum_{j\in C_1\cup\ldots\cup C_t} p_j \geq \sum_{i=1}^{t}\sum_{j\in C_i} a_{ij}p_j = \sum_{i=1}^{t}\sum_{j=1}^{n} a_{ij}p_j = \sum_{i=1}^{t} q_i.
\end{aligned}
$$

(The first inequality holds because $\vec{p}$ is nonincreasing and each $|C_i| \leq m$. The second inequality holds because $\sum_{i=1}^{n} a_{ij} = 1$ holds for each $1 \leq j \leq n$, whence a single p_j's appearances in various C_i's collectively contribute at most p_j to the sum on the right.) This shows that $\vec{r}q$, whence Lemma 5.4.6 tells us that $H(\vec{r}) \leq H(\vec{q})$. It follows by Jensen's inequality and the (ordinary) concavity of the logarithm that

$$
\begin{aligned}
H(\vec{p}) &\leq H(\vec{p}) + \log\sum_{i=1}^{n} p_i \frac{1}{p_i} r_{\lceil \frac{i}{m}\rceil} - \sum_{i=1}^{n} p_i \log\left(\frac{1}{p_i} r_{\lceil \frac{i}{m}\rceil}\right) \\
&= H(\vec{p}) + \log\sum_{i=1}^{n} r_{\lceil \frac{i}{m}\rceil} - \sum_{i=1}^{n} p_i \log\left(\frac{1}{p_i} r_{\lceil \frac{i}{m}\rceil}\right) \\
&\leq H(\vec{p}) + \log m - \sum_{i=1}^{n} p_i \log\left(\frac{1}{p_i} r_{\lceil \frac{i}{m}\rceil}\right) \\
&= \sum_{i=1}^{n} p_i \log\frac{1}{r_{\lceil \frac{i}{m}\rceil}} + \log m \\
&= \sum_{i=1}^{n} r_i \log\frac{1}{r_i} + \log m \\
&= H(\vec{r}) + \delta(\vec{p},\vec{q}) \\
&\leq H(\vec{q}) + \delta(\vec{p},\vec{q}),
\end{aligned}
$$

i.e., (5.4.1) holds. $\square$

5.5 Finite-State Dimension and Real Arithmetic

Our main theorem concerns real numbers rather than sequences, so the following notation is convenient. For each real number α and each integer $k \geq 2$, write

$$\dim_{\mathrm{FS}}^{(k)}(\alpha) = \dim_{\mathrm{FS}}(S)$$

and

$$\mathrm{Dim}_{\mathrm{FS}}^{(k)}(\alpha) = \mathrm{Dim}_{\mathrm{FS}}(S),$$

where S is a base-k expansion of $\alpha - \lfloor \alpha \rfloor$. Note that this notation is well-defined, because a real number α has two base-k expansions if and only if it is a k-adic rational, in which case both expansions are eventually periodic and hence have finite-state strong dimension 0. It is routine to verify the following.

Observation 5.5.1. *For every integer $k \geq 2$, every positive integer m, and every real number α,*

$$\dim_{\mathrm{FS}}^{(k)}(m + \alpha) = \dim_{\mathrm{FS}}^{(k)}(-\alpha) = \dim_{\mathrm{FS}}^{(k)}(\alpha)$$

and

$$\mathrm{Dim}_{\mathrm{FS}}^{(k)}(m + \alpha) = \mathrm{Dim}_{\mathrm{FS}}^{(k)}(-\alpha) = \mathrm{Dim}_{\mathrm{FS}}^{(k)}(\alpha).$$

The following lemma contains most of the technical content of our main theorem.

Lemma 5.5.2 (main lemma)**.** *For every integer $k \geq 2$, every positive integer m, and every real number $\alpha \geq 0$,*

$$\dim_{\mathrm{FS}}^{(k)}(m\alpha) = \dim_{\mathrm{FS}}^{(k)}(\alpha)$$

and

$$\mathrm{Dim}_{\mathrm{FS}}^{(k)}(m\alpha) = \mathrm{Dim}_{\mathrm{FS}}^{(k)}(\alpha).$$

Proof. Let k, m, and α be as given, let $S, T \in \Sigma^\infty$ be the base-k expansions of $\alpha - \lfloor \alpha \rfloor$, $m\alpha - \lfloor m\alpha \rfloor$, respectively, and write

$$\pi_{\alpha,n}^{(l)} = \pi_{S,n}^{(l)} \quad , \quad \pi_{m\alpha,n}^{(l)} = \pi_{T,n}^{(l)}$$

for each $l, n \in \mathbb{Z}^+$. By Corollary 5.4.5, it suffices to show that

$$\limsup_{n\to\infty} \delta\left(\pi^{(l)}_{\alpha,n}, \pi^{(l)}_{m\alpha,n}\right) = o(l) \tag{5.5.1}$$

as $l \to \infty$.

Let $r = \lfloor \log_k m \rfloor$, let

$$m = \sum_{i=0}^{r} m_i k^i$$

be the base-k expansion of m, and let

$$s = \sum_{i=0}^{r} m_i.$$

The first thing to note is that, in base k, $m\alpha - \lfloor m\alpha \rfloor$ is the sum, modulo 1, of s copies of $\alpha - \lfloor \alpha \rfloor$, with m_i of these copies shifted i symbols to the left, for each $0 \le i \le r$.

For each $l \in \mathbb{Z}^+$ and $j \in \mathbb{N}$, let

$$\begin{aligned} u^{(l)}_j &= S[jl \dots (j+1)l - 1], \\ v^{(l)}_j &= T[jl \dots (j+1)l - 1] \end{aligned}$$

be the j^{th} l-symbol blocks of $\alpha - \lfloor \alpha \rfloor$, $m\alpha - \lfloor m\alpha \rfloor$, respectively. If we let

$$\tau^{(l)}_j = \sum_{i=0}^{r} m_i \sum_{t=(j+1)l}^{\infty} S[t+i] k^{-(t+1)}$$

be the sum of the tails of the above-mentioned s copies of $\alpha - \lfloor \alpha \rfloor$ lying to the right of the j^{th} l-symbol block, then the block $v^{(l)}_j$ of $m\alpha - \lfloor m\alpha \rfloor$ is completely determined by $u^{(l)}_j$, the "carry"

$$c^{(l)}_j = \left\lfloor k^{(j+1)l} \tau^{(l)}_j \right\rfloor,$$

and the longest string of symbols shifted from the right, which is the string $u^{(l)}_{j+1}[0 \dots r-1]$. To be more explicit, note that

$$0 \le c^{(l)}_j \le k^{(j+1)l} \tau^{(l)}_j \le k^{(j+1)l} \sum_{i=0}^{r} m_i \sum_{t=(j+1)l}^{\infty} (k-1) k^{-(t+1)} = s;$$

define the "advice"

$$h^{(l)}_j = \left(c^{(l)}_j, u^{(l)}_j[0 \dots r-1]\right) \in \{0, \dots, s\} \times \Sigma^r;$$

and define the function

$$f^{(l)} : \Sigma^l \times \{0, \ldots, s\} \times \Sigma^r \to \Sigma^l$$

by letting $f^{(l)}(x, c, z)$ be the base-k expansion of the integer

$$mn_x^{(k)} + c + \sum_{i=0}^{r} m_i \sum_{t=0}^{i-1} z[t]k^t \ \mathbf{mod}\ k^l,$$

where $n_x^{(k)}$ is the nonnegative integer of which x is a base-k expansion, possibly with leading 0's. (Intuitively, the three terms here are the "block product", the "carry", and the "shift", respectively.) Then, for all integers $l > 0$ and $j \geq 0$,

$$v_j^{(l)} = f^{(l)}(u_j^{(l)}, h_j^{(l)}).$$

For positive integers l and n, define the $k^l \times k^l$ matrix $A^{(l,n)} = \left(a_{y,x}^{(l,n)}\right)$ by

$$a_{y,x}^{(l,n)} = \begin{cases} \dfrac{\left|\left\{ j < n \;\middle|\; u_j^{(l)} = x \text{ and } f^{(l)}\left(x, h_j^{(l)}(j)\right) = y \right\}\right|}{n\pi_{\alpha,n}^{(l)}(x)} & \text{if } \pi_{\alpha,n}^{(l)}(x) > 0 \\ 1 & \text{if } \pi_{\alpha,n}^{(l)}(x) = 0 \text{ and } x = y \\ 0 & \text{otherwise} \end{cases}$$

for all $x, y \in \Sigma^l$. It is routine to verify that

$$\sum_{y \in \Sigma^l} a_{y,x}^{(l,n)} = 1$$

for all $x \in \Sigma^l$, i.e., $A^{(l,n)}$ is stochastic, and that

$$\sum_{x \in \Sigma^l} a_{y,x}^{(l,n)} \pi_{\alpha,n}^{(l)}(x) = \pi_{m\alpha,n}^{(l)}(y)$$

for all $y \in \Sigma^l$, i.e., $A^{(l,n)}\pi_{\alpha,n}^{(l)} = \pi_{m\alpha,n}^{(l)}$. We complete the proof by bounding the number of nonzero entries in each row and column of $A^{(l,n)}$.

Fix a column x of $A^{(l,n)}$. If $\pi_{\alpha,n}^{(l)}(x) = 0$, then there is exactly one nonzero entry in column x of $A^{(l,n)}$. If $\pi_{\alpha,n}^{(l)}(x) > 0$, then the number of nonzero entries in column x is bounded by

$$|\{0, \ldots, s\} \times \Sigma^r| = (s+1)k^r \leq (s+1)m.$$

Hence there are at most $(s+1)m$ nonzero entries in column x of $A^{(l,n)}$.

Fix a row y of $A^{(l,n)}$. Let g be the greatest common divisor of m and k^l. Note that, for all $n_1, n_2 \in \mathbb{Z}^+$,

$$\begin{aligned} mn_1 \equiv mn_2 \mod k^l &\implies k^l \mid m(n_2 - n_1) \\ &\implies \left.\frac{k^l}{g} \,\right|\, \frac{m}{g}(n_2 - n_1) \\ &\implies \left.\frac{k^l}{g} \,\right|\, n_2 - n_1 \\ &\implies n_1 \equiv n_2 \mod \frac{k^l}{g}. \end{aligned}$$

This implies that each string $y \in \Sigma^l$ has at most g preimages x under the mapping that takes x to the base-k expansion of $mn_x^{(l)} \textbf{ mod } k^l$. This, in turn, implies that there are at most $g|\{0, \ldots, s\} \times \Sigma^r| \leq g(s+1)m$ nonzero entries in row y of $A^{(l,n)}$.

We have shown that, for each $l, n \in \mathbb{Z}^+$, the matrix $A^{(l,n)}$ testifies that

$$\delta\left(\pi^{(l)}_{\alpha,n}, \pi^{(l)}_{m\alpha,n}\right) \leq \log(g(s+1)m) \leq \log(m^2(s+1)).$$

Since this bound does not depend on l or n, this proves (5.5.1). $\square$

We now prove that addition and multiplication by nonzero rationals preserve finite-state dimension and finite-state strong dimension.

Theorem 5.5.3 (main theorem)**.** *For every integer $k \geq 2$, every nonzero rational number q, and every real number α,*

$$\dim^{(k)}_{\mathrm{FS}}(q+\alpha) = \dim^{(k)}_{\mathrm{FS}}(q\alpha) = \dim^{(k)}_{\mathrm{FS}}(\alpha)$$

and

$$\mathrm{Dim}^{(k)}_{\mathrm{FS}}(q+\alpha) = \mathrm{Dim}^{(k)}_{\mathrm{FS}}(q\alpha) = \mathrm{Dim}^{(k)}_{\mathrm{FS}}(\alpha).$$

Proof. Let k, q, and α be as given, and write $q = \frac{a}{b}$, where a and b are integers with $a \neq 0$ and $b > 0$. By Observation 5.5.1 and Lemma 5.5.2,

$$\begin{aligned} \dim^{(k)}_{\mathrm{FS}}(q\alpha) &= \dim^{(k)}_{\mathrm{FS}}\left(\frac{|a|}{b}\alpha\right) &= \dim^{(k)}_{\mathrm{FS}}\left(b\frac{|a|}{b}\alpha\right) \\ &= \dim^{(k)}_{\mathrm{FS}}(|a|\alpha) &= \dim^{(k)}_{\mathrm{FS}}(\alpha), \end{aligned}$$

and

$$\begin{aligned}
\dim_{\mathrm{FS}}^{(k)}(q+\alpha) &= \dim_{\mathrm{FS}}^{(k)}\left(\frac{a}{b}+\alpha\right) &&= \dim_{\mathrm{FS}}^{(k)}\left(\frac{a+b\alpha}{b}\right) \\
&= \dim_{\mathrm{FS}}^{(k)}\left(b\frac{a+b\alpha}{b}\right) &&= \dim_{\mathrm{FS}}^{(k)}(a+b\alpha) \\
&= \dim_{\mathrm{FS}}^{(k)}(b\alpha) &&= \dim_{\mathrm{FS}}^{(k)}(\alpha).
\end{aligned}$$

Similarly, $\mathrm{Dim}_{\mathrm{FS}}^{(k)}(q\alpha) = \mathrm{Dim}_{\mathrm{FS}}^{(k)}(\alpha)$, and $\mathrm{Dim}_{\mathrm{FS}}^{(k)}(q+\alpha) = \mathrm{Dim}_{\mathrm{FS}}^{(k)}(\alpha)$. □

Finally, we note that Theorem 5.5.3 gives a new proof of the following classical theorem.

Corollary 5.5.4. *(Wall [84]) Let $k \geq 2$. For every nonzero rational number q and every real number α that is normal base k, the sum $q+\alpha$ and the product $q\alpha$ are also normal base k.*

CHAPTER 6. Liouville Numbers and Normality

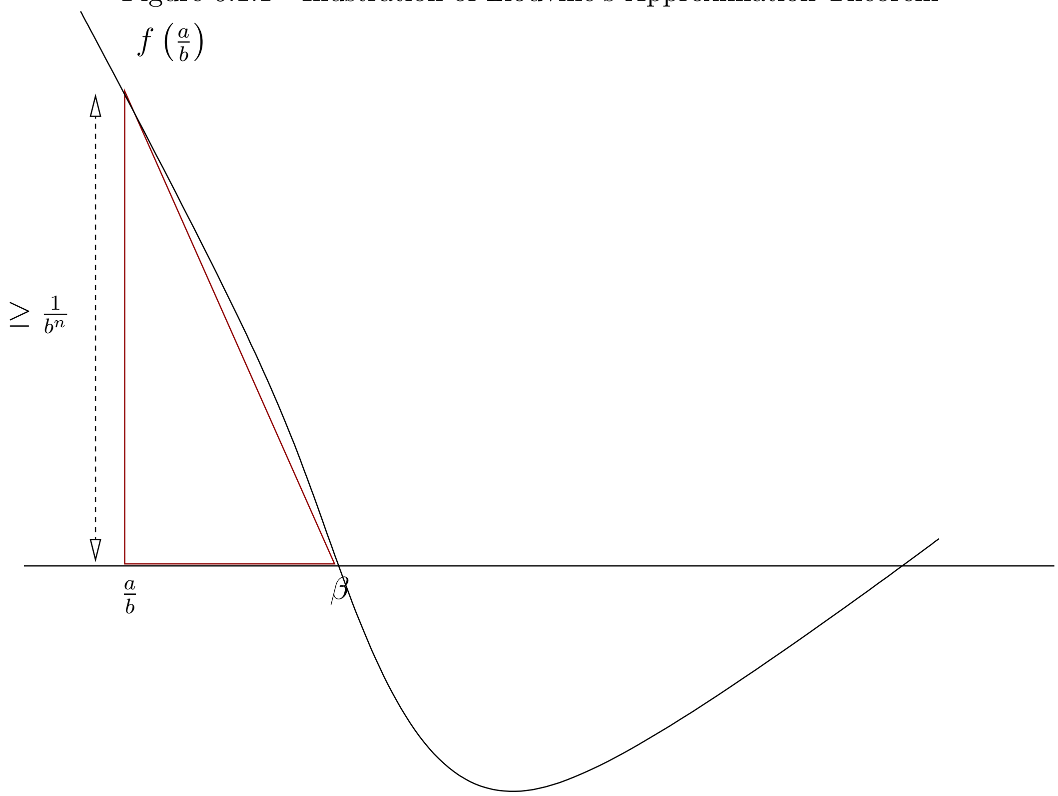
One of the important open questions in the study of normality is whether any algebraic irrational number is normal. On the other hand, it is known that there are normal transcendentals as well as non-normal transcendentals. For example, Mahler has proved that the Champernowne constant [58] as well as the Thue-Morse [57] constant are transcendental. However, the Champernowne constant is normal [15], whereas the Thue-Morse constant has finite-state dimension 0 [3].

Liouville numbers were the first class of numbers which were proven transcendental. In this chapter, we show that there are Liouville numbers which are non-normal, and others which are normal. Examples of non-normal Liouville numbers are well-known.

Normal Liouville numbers are harder to construct, and there are works by Kano [37] and Bugeaud [12] which construct such numbers. Kano constructs, for any bases a and b, Liouville numbers which are normal in base a but not in base ab. Bugeaud gives a non-constructive proof using Fourier analytic techniques that there are Liouville numbers which are absolutely normal – that is, normal in all bases. In this chapter, we give a combinatorial construction of a number that is normal to a given base b. The construction is elementary, but the result is weaker.

Thus the Liouville numbers forms a class of numbers whose transcendence is easy to establish, and which contain simple examples of normal and non-normal numbers. We begin with a survey briefly explaining Liouville's approximation theorem and defining the class of Liouville numbers. Section 6.2 constructs a Liouville number which is disjunctive – that is to say, has all strings appearing in its base b expansion– but is still not normal. The subsequent section gives the construction of a normal Liouville number.

Figure 6.1.1 Illustration of Liouville's Approximation Theorem



6.1 Liouville's Constant and Liouville Numbers

Liouville's approximation theorem says that algebraic irrationals are inapproximable by rational numbers to arbitrary precisions.

Theorem 6.1.1 (Liouville's Theorem)**.** *Let β be a root of $f(x) = \sum_{j=0}^{n} a_j x^j \in \mathbb{Z}[x]$. Then there is a constant C_β such that for every pair of integers a and b, $b > 0$, we have*

$$\left|\beta - \frac{a}{b}\right| > \frac{C_\beta}{b^n}.$$

The proof may be motivated by the following discussion. α is the root of a polynomial of degree n. This puts an upper bound on the absolute value of the slope of the polynomial curve in a region around α. Let $\frac{a}{b}$ be a rational. Again using the fact that the polynomial has integer coefficients, we see that $f(\frac{a}{b}) \geq \frac{1}{b^n}$. Liouville's theorem then states the fact that the triangle in Figure 6.1.1 has positive area.

Liouville then constructed the following provably irrational number

$$\psi_1 = \sum_{i=0}^{\infty} 10^{-i!},$$

and showed that it had arbitrarily good rational approximations, and therefore is a transcendental number.

Definition. A real number α in the unit interval is called a *Liouville number* if for all numbers n, there are numbers $p > 0$ and $q > 1$ such that

$$\left|\alpha - \frac{p}{q}\right| < \frac{1}{q^n}.$$

For every n, we have

$$\left|\psi_1 - \sum_{i=1}^{n} \frac{1}{2^{i!}}\right| = \sum_{i=n+1}^{\infty} \frac{1}{2^{i!}} = \sum_{i=(n+1)!}^{\infty} \frac{1}{2^i} = \frac{1}{2^{(n+1)!-1}} < \frac{1}{2^{n!n}} = \frac{1}{q_n^n},$$

where q_n is the denominator of the finite sum $\sum_{i=1}^{n} \frac{1}{2^{i!}}$. Thus ψ_1 is a Liouville number.

It is easy to see that the Liouville constant ψ_1 is not a normal number - the sequence 111 never appears in the decimal expansion of ψ_1. It is natural to investigate whether all Liouville numbers are non-normal.

This requires sharper observations than the one above.

6.2 Disjunctive Liouville Sequences

Hertling [30] has showed that there are disjunctive Liouville numbers - that is, there are Liouville numbers whose base r expansions have all possible r-alphabet strings. Staiger strengthened this result to show that there are Liouville numbers which are disjunctive in any base [75]. This shows that we cannot rely on the above argument of absent strings to show non-normality.

Here, to motivate the construction of a normal number in the next section, we give a different construction of a different disjunctive Liouville sequence. Consider

$$\psi_2 = \sum_{i=3}^{\infty} \frac{i}{2^{i!}}.$$

For any binary string w, we know that $1w$ (1 concatenated with w) is the binary representation of a number, hence it appears in the binary expansion of ψ_2. Thus ψ_2 is a disjunctive sequence.

It is easy to see that ψ_2 is not a normal number. At all large enough prefix lengths of the form $n!$, there are at most $n(\lfloor \log_2 n \rfloor + 1)$ ones - this follows from the fact that at most n unique non-zero numbers have appeared in the binary expansion of ψ_2, and each of the numbers can be represented with at most $\lfloor \log_2 n \rfloor + 1$ bits. Hence

$$\liminf_{n\to\infty} \frac{|\{i : 0 \le i \le n-1 \text{ and } \psi_2[i] = 1\}|}{n} \le \lim_{n\to\infty} \frac{n\left(\lfloor \log_2 n \rfloor + 1\right)}{n!} = 0,$$

which proves that ψ_2 is not normal.

However, ψ_2 is a Liouville number: For every n, there are rationals with denominators of size $2^{n!}$ which satisfy the Liouville criterion, as follows:

$$\begin{aligned}
\left|\psi - \sum_{i=3}^{n} \frac{i}{2^{i!}}\right| &= \sum_{i=n+1}^{\infty} \frac{i}{2^{i!}} \\
&< \sum_{k=1}^{\infty} \frac{n+k}{2^{(n+1)!\cdot k}} \\
&= \frac{[n+1]\cdot 2^{(n+1)!-1} + 1}{2^{((n+1)!-1)\cdot 2}} \qquad \text{[Summation of an arithmetico-geometric series]} \\
&< \frac{[n+2]}{2^{(n+1)!-1}} \\
&< \frac{1}{2^{(n!)\cdot n}}.
\end{aligned}$$

6.3 A Normal Liouville Number

Though the Liouville numbers constructed above were non-normal, there are normal Liouville numbers. We give such a construction below, which depends on DE BRUIJN SEQUENCES introduced by de Bruijn [20] and Good [26], a standard tool in the study of normality.[1]

Definition. Let Σ be an alphabet with size k. A *k-ary de Bruijn sequence* $B(k,n)$ of order n, is a finite string for which every possible string in Σ^n appears exactly once.

de Bruijn proved that such sequences exist for all k and all orders n. Since each de Bruijn sequence $B(k,n)$ contains each n-length string exactly once, it follows that the length of $B(k,n)$ is exactly k^n.

[1] There are several historical forerunners of this concept in places as varied as Sanskrit prosody and poetics. However, the general construction for all bases and all orders is not known to be ancient.

6.3.1 Construction

If w is any string, we write w^i for the string formed by repeating w, i times. In this section, we limit ourselves to the binary alphabet $\Sigma = \{0, 1\}$ even though the construction generalizes to all alphabets.

Consider $\alpha \in [0, 1)$ with binary expansion defined as follows.

$$\alpha = 0\ \ .\ \ B(2,1)^{1^1} B(2,2)^{2^2} B(2,3)^{3^3} \ldots B(2,i)^{i^i} \ldots .$$

Informally, we can explain why this construction defines a normal Liouville number, as follows. The Liouville numbers ψ_1 and ψ_2 that we considered before, have prefixes that are mostly zeroes. The density of 1s go asymptotically tends to zero as we consider longer and longer prefixes. So it is fairly easy for a finite state compressor to compress the data in a prefix. However, in this construction, the repeating patterns employed are those which are eventually hard for any given finite state compressor. This is why the sequence could be normal.

Moreover, the transition in the patterns occur at prefix lengths of the form k^k. By Stirling's approximation,

$$k! \approxeq k^k e^{-k} \sqrt{2\pi k}.$$

So the transitions in the pattern occur at prefix lengths similar to that of ψ_1 and ψ_2, so it is reasonable to expect a sequence of rationals approximating α that obeys the Liouville criterion. We now make this argument more precise.

For any i, let $n_i = \sum_{m=1}^{i} m^m 2^m$. We call the part of $\alpha[n_{i-1} \ldots n_i - 1]$ as the $i^t h$ stage of α, which consists of i^i copies of $B(2, i)$. Thus n_i denotes the length of the prefix of α which has been defined at the end of the i^{th} stage. We have the following estimate for n_i.

$$
\begin{aligned}
n_i &= \sum_{m=1}^{i} m^m 2^m \\
&< i^i \sum_{m=1}^{i} 2^m \\
&= i^i \left[2^{i+1} - 2\right] \\
&= i^i 2^{i+1} - 2i^i \\
&< 2\left(i^i 2^i\right).
\end{aligned}
$$

Thus $n_i = O([2i]^i)$.

Lemma 6.3.1. *α is a Liouville number.*

Proof. Consider the rational number $\frac{p_i}{q_i}$ [2] with a binary expansion which coincides with α until the $i-1^{\text{st}}$ stage, followed by a recurring block of $B(2,i)$. This rational number is

$$
\frac{\alpha[0\ldots n_{i-1}]}{2^{n_{i-1}}} + \frac{B(2,i)}{(2^{2^i}-1)2^{n_{i-1}}},
$$

obtained by evaluating the binary expansion as a geometric series. The *exponent* of the denominator of this rational number is

$$
2^i + n_{i-1} = 2^i + O\left([2(i-1)]^{i-1}\right) = O(n_{i-1}),
$$

so the denominator of the rational is $2^{O(n_{i-1})}$.

We add i^i copies of $B(2,i)$ in the i^{th} stage. Thus the expansion of α and that of r_i coincide for the first n_i positions. So, α and $\frac{p_i}{q_i}$ are in the same dyadic interval of length at most 2^{-n_i}, and hence are within $\frac{1}{2^{n_i}}$ of each other.

We have also that $n_i > iO\left([2(i-1)]^{(i-1)}\right)$, so that

$$
\frac{1}{2^{n_i}} < \frac{1}{\left(2^{O(n_{i-1})}\right)^i}.
$$

Thus,

$$
\left|\alpha - \frac{p_i}{q_i}\right| < \frac{1}{q_i^i}.
$$

Since this is true of any stage i, we can see that α is a Liouville number. $\square$

[2] not necessarily in the lowest form

Lemma 6.3.2. *α is normal to the base 2.*

Proof. Let us define COUNT : $\Sigma^* \times \Sigma^* \to \mathbb{N}$ by

$$\text{COUNT}(w; x) = |\{n \mid x[n \ldots n + |w| - 1] = w\}|\,,$$

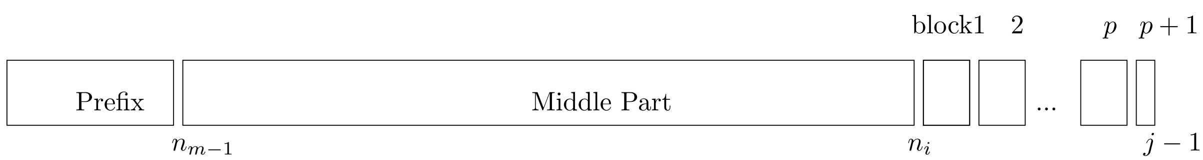
that is, the number of times w occurs in x, counting in a sliding block fashion. For example, 00 occurs twice in 1000. It is enough to show that for an arbitrary binary string w of length m, for all large enough indices j,

$$\text{COUNT}(w; \alpha[0 \ldots j-1]) = 2^{-m} j + o(j).$$

Let j be a number greater than n_m. Every such index j has a number i such that $n_i < j \leq n_{i+1}$.

We split the analysis into three phases, that of the prefix $\alpha[0 \ldots n_{m-1} - 1]$, of the middle region $\alpha[n_{m-1} \ldots n_i]$, and of the suffix $\alpha[n_i \ldots j-1]$.

Figure 6.3.1 Schema of Analysis of $\alpha[0 \ldots j-1]$



The prefix $\alpha[0 \ldots n_{m-1}]$ has a constant length that depends on w but not on j. Hence the discrepancy in the count of w due to this prefix, which is at most n_{m-1}, is $o(j)$.

Since the number of times w occurs in $B(2, m)$ is exactly 1,

$$\text{COUNT}(w; B(2, m)) = 2^{-m}\ |B(2, m)|.$$

Similarly, it is easy to see that for any $M > m$, by the properties of the de Bruijn sequences,

$$\text{COUNT}(w; B(2, M)) = 2^{-m}\ |B(2, M)|.$$

This observation is used in the following analysis of the middle part and the suffix.

The part of of α in the stretch $n_{m-1} \dots n_m - 1$ parses exactly into m^m disjoint blocks of $B(2, m)$. Consequently,

$$\text{COUNT}(w; \alpha[n_{m-1} \dots n_m]) = 2^{-m}(n_m - n_{m-1}).$$

For all stages k between $m-1$ and i,

$$\text{COUNT}(w; \alpha[n_k \dots n_{k+1} - 1]) = 2^{-m}(n_{k+1} - n_k),$$

hence by a telescoping sum,

$$\text{COUNT}(w; \alpha[n_{m-1} \dots n_i - 1]) = 2^{-m}(n_i - n_{m-1}).$$

The suffix is formed during the $i + 1^{\text{st}}$ stage of construction of α, and hence consists of $(i+1)^{(i+1)}$ copies of $B(2, i+1)$. Let j be such that

$$p.2^{(i+1)} < j - n_i < (p+1)2^{i+1} - 1.$$

That is, j falls within the $p + 1^{\text{st}}$ copy of $B(2, i+1)$. Then,

$$j = n_i + p2^{i+1} + o(j),$$

since the last term is at most $2^{i+1} - 1$ and $n_i = \Omega(i^i)$.

Since w is normally distributed in $\alpha[0 \dots n_i]$ and in each of the p preceding copies of $B(2, i+1)$, we have

$$\text{COUNT}(w; \alpha[0 \dots j-1]) = 2^{-m}n_i + 2^{-m}p2^{i+1} + o(j) = 2^{-m}j + o(j),$$

showing that α is normal. $\square$

Bibliography

[1] V. N. Agafonov. Normal sequences and finite automata. *Soviet Mathematics Doklady*, 9:324–325, 1968.

[2] K. B. Athreya, J. M. Hitchcock, J. H. Lutz, and E. Mayordomo. Effective strong dimension, algorithmic information, and computational complexity. In *Proceedings of the 21st International Symposium on Theoretical Aspects of Computer Science*, pages 632–643. Springer-Verlag, 2004.

[3] M. Bachan. Finite State Dimension of the Kolakoski Sequence. Master's thesis, Iowa State University, Ames, U. S. A., 2005.

[4] R. Bhatia. *Matrix Analysis*. Springer, 1997.

[5] P. Billingsley. *Ergodic theory and information*. John Wiley and Sons, New York, 1965.

[6] G. D. Birkhoff. Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences, USA*, 17:656–660, 1931.

[7] E. Bishop. *Foundations of constructive analysis*. McGraw Hill, New York, 1967.

[8] É. Borel. Sur les probabilités dénombrables et leurs applications arithmétiques. *Rend. Circ. Mat. Palermo*, 27:247–271, 1909.

[9] J. Borwein and D. Bailey. *Mathematics by experiment: plausible reasoning in the 21st century*. A. K. Peters Ltd., Natick, MA, 2004.

[10] C. Bourke. Effective randomness and finite state dimension, 2004.

[11] M. Braverman. On the complexity of real functions. In *Proceedings of the forty sixth IEEE Annual Symposium on the Foundations of Computer Science*, pages 155–164, 2005.

[12] Yann Bugeaud. Nombres de liouville aux nombres normaux. *Comptes Rendus Mathématique. Académie des Sciences. Paris*, 335:117, 2002.

[13] J. W. S. Cassels. *Colloquium Mathematicum*, 7:95–101, 1959.

[14] G. J. Chaitin. On the length of programs for computing finite binary sequences: statistical considerations. *Journal of the ACM*, 16:145–159, 1969.

[15] D. G. Champernowne. Construction of decimals normal in the scale of ten. *J. London Math. Soc.*, 2(8):254–260, 1933.

[16] A. Church. On the concept of a random sequence. *Bulletin of the American Mathematical Society*, 46:130–135, 1940.

[17] J. J. Dai, J. I. Lathrop, J. H. Lutz, and E. Mayordomo. Finite-state dimension. *Theoretical Computer Science*, 310:1–33, 2004.

[18] K. Dajani and C. Kraaikamp. *Ergodic Theory of Numbers.* Mathematical Association of America, 2004.

[19] Karma Dajani and Cor Kraaikamp. *Ergodic Theory of Numbers.* The Mathematical Association of America, 2002.

[20] N. G. de Bruijn. A combinatorial problem. *Koninklijke Nederlandse Akademie v. Wetenschappen*, 49, 1946.

[21] D. Doty, J. H. Lutz, and S. Nandakumar. Finite state dimension and real arithmetic. *Information and Computation*, 205:1640–1651, 2007.

[22] R. Downey and D. Hirschfeldt. *Algorithmic randomness and complexity.* Book Draft, 2006.

[23] K. Falconer. *Fractal Geometry: Mathematical Foundations and Applications.* John Wiley & Sons, 1990.

[24] S. A. Fenner. Gales and supergales are equivalent for defining constructive hausdorff dimension. Technical Report cs.CC/0208044, Computing Research Repository, 2002.

[25] P. Gács. Unifrom test of algorithmic randomness over a general space. *Theoretical Computer Science*, 341:91–137, 2005.

[26] I. J. Good. Normal recurring decimals. *Journal of the London Mathematical Society*, 21:167, 1946.

[27] X. Gu, J. H. Lutz, and P. Moser. Dimension of copeland-erdös sequences. *Information and Computation*, 205:1640–1651, 2007.

[28] G. Harman. One hundred years of normal numbers. In *Surveys in Number Theory: Papers from the millennial conference on Number theory*, pages 57–74, 2003.

[29] F. Hausdorff. Dimension und äußeres Mass. *Mathematische Annalen*, 79:157–179, 1919.

[30] Peter Hertling. Disjunctive omega-words and real numbers. *J. UCS*, 2(7):549–568, 1996.

[31] J. Hitchcock, J. Lutz, and E. Mayordomo. Scaled dimension and nonuniform complexity. *Journal of Computer and System Sciences*, 69:97–122, 2004.

[32] J. M. Hitchcock. Fractal dimension and logarithmic loss unpredictability. *Theoretical Computer Science*, 304(1–3):431–441, 2003.

[33] J. M. Hitchcock. Gales suffice for constructive dimension. *Information Processing Letters*, 86(1):9–12, 2003.

[34] J. M. Hitchcock, J . H. Lutz, and E. Mayordomo. The fractal geometry of complexity classes. *SIGACT News*, 36:24–38, 2005.

[35] T. Kamae. Subsequences of normal sequences. *Israel Journal of Mathematics*, 16:121–149, 1973.

[36] T. Kamae and B. Weiss. Normal numbers and selection rules. *Israel Journal of Mathematics*, 21:101–110, 1975.

[37] H. Kano. General constructions of normal numbers of the Korobov type. *Osaka Journal of Mathematics*, 4:909, 1993.

[38] A. N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeishrechnung*. Julius Springer, 1933.

[39] A. N. Kolmogorov. On tables of random numbers. *Sankhyā, Series A*, 25:369–376, 1963.

[40] A. N. Kolmogorov. Three approaches to the quantitative definition of 'information'. *Problems of Information Transmission*, 1:1–7, 1965.

[41] A. N. Kolmogorov. Logical basis for information theory and probability theory. *IEEE Transactions on Information Theory*, IT-14:662–664, 1968.

[42] A. N. Kolmogorov. Combinatorial foundations of information theory and calculus of probabilities. *Russian Mathematical Surveys*, 38:29–40, 1983.

[43] A. N. Kolmogorov and V. A. Uspenksy. Algorithms and randomness. *Theory of Probability and Its Applications*, 32:389, 1987.

[44] L. Kuipers and H. Niederreiter. *Uniform Distribution of Sequences*. Wiley-Interscience, 1974.

[45] P. Lévy. Propriétés asymptotiques des sommes de variables indépendantes ou enchainées. *Journal des mathématiques pures et appliquées. Series 9.*, 14(4):347–402, 1935.

[46] P. Lévy. *Théorie de l'Addition des Variables Aleatoires*. Gauthier-Villars, 1937 (second edition 1954).

[47] M. Li and P. M. B. Vitányi. *An Introduction to Kolmogorov Complexity and its Applications*. Springer-Verlag, Berlin, 1997. Second Edition.

[48] D. W. Loveland. The Kleene hierarchy classification of recursively random sequences. *Transactions of the American Mathematical Society*, 125:497–510, 1966.

[49] D. W. Loveland. A new interpretation of von Mises' concept of a random sequence. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 12:279–294, 1966.

[50] J. Lutz and E. Mayordomo. Dimensions of points in self-similar fractals. *submitted*.

[51] J. H. Lutz. Almost everywhere high non-uniform complexity. *Journal of Computer and System Sciences*, 44:220, 1992.

[52] J. H. Lutz. The quantitative structure of exponential time. In L. A. Hemaspaandra and A. L. Selman, editors, *Complexity Theory Retrospective II*, pages 225–254. Springer-Verlag, 1997.

[53] J. H. Lutz. Resource-bounded measure. In *Proceedings of the 13th IEEE Conference on Computational Complexity*, page 236, New York, 1998. IEEE Computer Society Press.

[54] J. H. Lutz. Gales and the constructive dimension of individual sequences. In *Proceedings of the 27th International Colloquium on Automata, Languages, and Programming*, pages 902–913, 2000. Revised as [56].

[55] J. H. Lutz. Dimension in complexity classes. *SIAM Journal on Computing*, 32:1236–1259, 2003. Preliminary version appeared in *Proceedings of the Fifteenth Annual IEEE Conference on Computational Complexity*, pages 158–169, 2000.

[56] J. H. Lutz. The dimensions of individual strings and sequences. *Information and Computation*, 187:49–79, 2003. Preliminary version appeared as [54].

[57] K. Mahler. Arithmatische eigenschaften der Lösungen einer Klasse von Funktionalgrichungen. *Mathematische Annalen*, 101:342, 1929.

[58] K. Mahler. Arithmatische eigenschaften einer Klasse von Dezimalbrüchen. *Proc. Kon. Nederlandsche Akad. v. Wetenschappen*, 40:421, 1937.

[59] A. W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and its applications.* Academic Press, 1979.

[60] P. Martin-Löf. The definition of random sequences. *Information and Control*, 9:602–619, 1966.

[61] W. Merkle. The Kolmogorov-Loveland stochastic sequences are not closed under selecting subsequences. In *Proceedings of the 29th International Colloquium on Automata, Languages, and Programming*, pages 390–400, 2002.

[62] W. Merkle and J. Reimann. On selection functions that do not preserve normality. *Theory of computing systems*, 9:685–697, 2006.

[63] S. Nandakumar. An effective ergodic theorem and some applications. In *Proceedings of the 40th ACM Annual Symposium on the Theory of Computing*, pages 39–44. Association of Computing Machinery, 2008.

[64] S. Nandakumar. A charaterization of constructive dimension. *Mathematical Logic Quarterly*, 55:271–286, 2009.

[65] Ivan Niven. *Irrational Numbers.* Number 11 in The Carus Mathematical Monographs. 1956.

[66] J. Reimann. Random functions. *Unpublished Lecture, Computability and Randomness, Cordoba*, 2004.

[67] W. Schmidt. *Pacific Journal of Mathematics*, 10:661–672, 1960.

[68] C. P. Schnorr. Klassifikation der Zufallsgesetze nach Komplexität und Ordnung. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 16:1–21, 1970.

[69] C. P. Schnorr. A unified approach to the definition of random sequences. *Mathematical Systems Theory*, 5:246–258, 1971.

[70] C. P. Schnorr. *Zufälligkeit und Wahrscheinlichkeit.* Springer-Verlag, Berlin, 1971.

[71] C. P. Schnorr and H. Stimm. Endliche automaten und zufallsfolgen. *Acta Informatica*, 1:345–359, 1972.

[72] I. Schur. Über eine klasse von mittelbildungen mit anwendungen auf die determinantentheorie. *Sitzungsberichte der Berliner Mathematischen Gesellschaft*, 22:9–20, 1923.

[73] A. Kh. Shen′. The frequency approach to the definition of a random sequence. *Semiotika i Informatika*, 18:14–42, 1982. (In Russian.).

[74] R. J. Solomonoff. A formal theory of inductive inference. *Information and Control*, 7:1–22, 224–254, 1964.

[75] Ludwig Staiger. The kolmogorov complexity of real numbers. In *FCT*, pages 536–546, 1999.

[76] D. Sullivan. Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups. *Acta Mathematica*, 153:259–277, 1984.

[77] C. Tricot. Two definitions of fractional dimension. *Mathematical Proceedings of the Cambridge Philosophical Society*, 91:57–74, 1982.

[78] M. van Lambalgen. *Random Sequences*. PhD thesis, Department of Mathematics, University of Amsterdam, 1987.

[79] J. Ville. *Étude Critique de la Notion de Collectif*. Gauthier-Villars, Paris, 1939.

[80] R. von Mises. Grundlagen der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 5:52–99, 1919.

[81] V. Vovk. The law of the iterated logarithm for random kolmogorov, or chaotic, sequences. *Theory of Probability and Applications*, 32:413–425, 1987.

[82] V. G. Vovk and V. V. V'yugin. On the empirical validity of the bayesian method. *Journal of Royal Statistical Society B*, 55:253–266, 1993.

[83] V. V. V'yugin. Ergodic theorems for individual random sequences. *Theoretical Computer Science*, 207:343–361, 1998.

[84] D. D. Wall. *Normal Sequences*. PhD thesis, University of California, Berkeley, 1949.

[85] P. Walters. *An introduction to ergodic theory.* Springer Verlag, New York, 1982.

[86] K. Weihrauch. *Computable Analysis.* Springer Verlag, Berlin, 2000.

[87] B. Weiss. Single orbit dynamics. 2000.