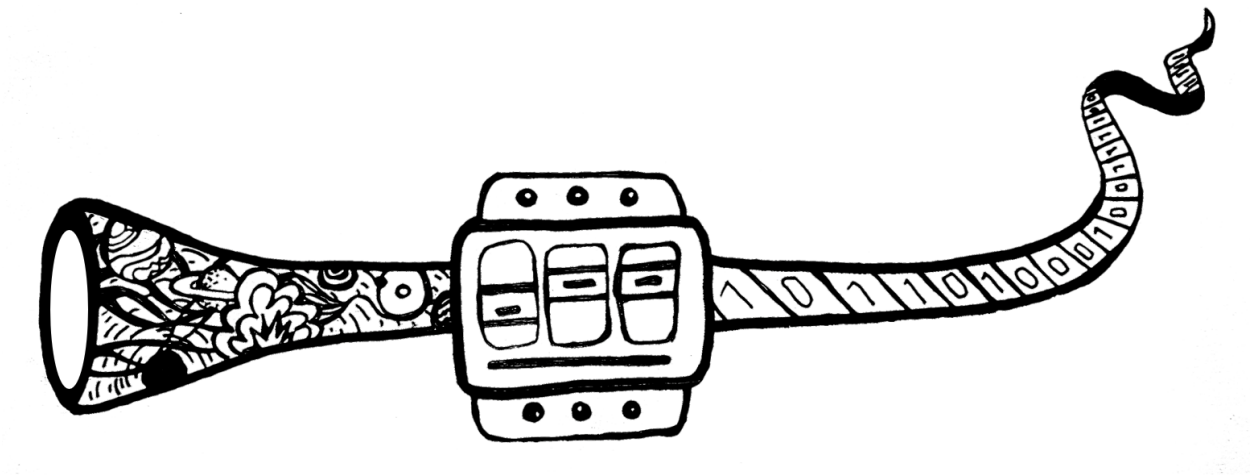
Algorithmic Physics



Samuel Epstein
samepst@jptheorygroup.org

May 10, 2024

Abstract

This article presents a survey of published and unpublished material of the intersection of algorithmic information theory with various areas of physics including quantum mechanics, thermodynamics, Newtonian physics, black holes, and constructor theory. The relationship between algorithmic information and quantum measurements is explored. There are no benefits to using quantum mechanics to compress classical information. One of the surprising results is that an overwhelming majority of non-pointer quantum (pure and mixed) states, when undertaking decoherence, will result in a classical probability with no algorithmic information. Thus most non-pointer quantum states decohere into white noise. Algorithmic information theory presents new complications for the many worlds theory, as it conflicts with the Independence Postulate. As for thermodynamics, new definitions of algorithmic coarse and fine grained entropy are introduced. The algorithmc fine grained entropy function oscillates during the course of dynamics. Small fluctuations are common and larger fluctuations are more rare. Coarse grained entropy is shown to be an excellent approximation to fine grained entropy. Marginal algorithmic thermodynamic entropies cannot be synchronized during the course of joint or independent dynamics. For Newtonian physics, a typicality measure is introduced that scores the level of algorithmic typicality of a position in Newtonian space. During the course of an orbit around a mass point, a point will oscillate in typicality. Furthermore two orbits that are not exotic cannot have synchronized typicality measures. The Kolmogorov complexity of Black holes is detailed and its relation to the Complexity/Volume Correspondence is described. The Independence Postulate is shown to be in conflict with the Many Worlds Theory and Constructor Theory.

Contents

Part I

Introduction

Chapter 1

Some Surprises About Algorithmic Physics

This manuscript deals with the application of algorithmic information theory to physics, namely quantum information theory, thermodynamics, newtonian gravity, and black holes. The reader is assumed to be familiar with these three areas, and the reader is refered to the books [LO97, Wil13, She15]. The main references to this manuscript are [Eps19c, Gác01, Gac94, HR09, SBC01, Vit01, NS19, Vai98, Eve57, Eps20, Vit00, BvL01, Mue07, Mul08, Mul09] and unpublished material from the author. In particular, the references of special import are [Gác01, Gac94] The reference [Gác01] introduces the central quantum matrix $\boldsymbol{\mu}$, which is used the quantum equivalent to the algorithmic semi-measure $\mathbf{m}$ and is used define quantum complexity, quantum mutual information, and to prove properties of quantum measurements. The reference [Gac94] introduces algorithmic (coarse and fine grained) thermodynamic entropy with the key insight that it is the negative logarithm of a universal lower computable test over the phase space. We now present some interesting facts about algorithmic physics that are detailed in this manuscript.

1.1 Müller's Theorem

A central topic of investigation in computer science is whether leveraging different physical models can change computability and complexity properties of constructs. In a remarkable result, Shor's factoring algorithm uses quantum mechanics to perform factoring in polynomial time. One question is whether quantum mechanics provides benefits to compressing classical information. In this manuscript, a negative answer is detailed and new shortened proofs are given. The plain Kolmogorov complexity of a string $x \in \{0,1\}^n$ is $\mathbf{C}(x)$, the size of the smallest program to a (plain) classical universal Turing machine that can produce x. The quantum Kolmogorov complexity of an n qubit state $|\psi\rangle$, which we call BvL complexity (named after its originators), is $\mathbf{Hbvl}(|\psi\rangle)$, the size of the smallest mixed quantum state input to a universal quantum Turing machine that produces $|\psi\rangle$ up to arbitrary fidelity. Due to [Mue07, Mul09],

$$\mathbf{C}(x) =^+ \mathbf{Hbvl}(|x\rangle \langle x|).$$

Thus there are no benefits to using quantum mechanics to compress classical information. The quantitative amount of information in bits is invariant to the physical model used.

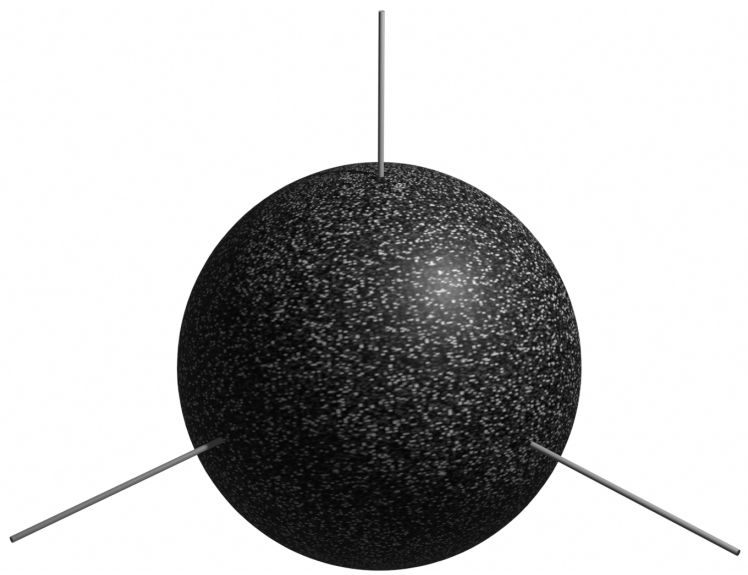


Figure 1.1: Over the uniform distribution of n qubit pure states $|\psi\rangle$, the self information function $\mathbf{I}(|\psi\rangle : |\psi\rangle)$ is almost always near 0, with sporadic spikes (such as at the basis state $|x\rangle$ for random string x.)

1.2 Quantum States Have No Self-Information

All strings of high Kolmogorov complexity have high self information, with $\mathbf{I}(x : x) =^+ \mathbf{K}(x)$. However the situation is much different in the quantum world, with respect to the definition of mutual information of quantum mixed states σ and ρ introduced in Chapter 8: $\mathbf{I}(\sigma : \rho)$. Almost all pure states $|\psi\rangle$ have low $\mathbf{I}(|\psi\rangle : |\psi\rangle)$, as shown in Figure 1.1. Indeed, let Λ be the uniform distribution over n qubit pure states:

$$\int 2^{\mathbf{I}(|\psi\rangle:|\psi\rangle)} d\Lambda = O(1).$$

This upper bound has several consequences, one being that given a (POVM) measurement, its application to overwhelming majority of quantum states produces white noise, as shown in Chapter 9. In addition, an overwhelming majority of non-pointer quantum states decohere into random noise. These results are a consequence of the vastness of Hilbert spaces opposed to the limited discretionary power of measurements. Conservation inequalities prevent any type of post-processing of the measured information. As discussed in Chapter 11, the only means to infuse quantum self information is with a projection operation caused by a quantum measurement.

1.3 Algorithmic Thermodynamic Entropy and Fluctuations

Thermodynamic entropy is subject to fluctuations. It will spend most of its time at its maximum value, will exhibit frequent small flucuations, and rarer large fluctuations. In this manuscript, we show that algorithmic fine grained entropy exihibits such oscillations, and even go one step further in proving the existence of synchronized oscillations for discrete dynamics.

The phase space Ω describes all possible states of the dynamic system, such as all the particles momentums and positions. The phase space is paired with (not necessarily probability) computable measure μ that represents the volume of the space. Like classical thermodynamic entropy, algorithmic fine grained entropy is defined with respect to a particular measure μ and phase space, denoted $\mathbf{H}_\mu(x)$ over $x \in \Omega$.

Due to the Liouville theorem, the dynamics of the system are volume invariant In this manuscript, it is proved that during the course of such dynamics, algorithmic fine grain thermodynamic have oscillations. Small dips in $\mathbf{H}_\nu$ are frequent, and larger dips are more rare. We get the following inequalities, where $\mathbf{K}$ is the prefix-free Kolmogorov complexity. This parallels the discrete ergodic transformation case, detailed in Chapter 16.

Let L be the Lebesgue measure over $\mathbb{R}$, and $(\mathcal{X}, \mu)$ be a computable measure space and $\alpha \in \mathcal{X}$ with finite mutual information with the halting sequence. For transformation group G^t acting on $\mathcal{X}$, there is a constant c with

$$2^{-n-\mathbf{K}(n)-c} < L\{t \in [0,1] : \mathbf{H}_\mu(G^t\alpha) < \log \mu(\mathcal{X}) - n\} < 2^{-n+c}.$$

The above result proves the existence of oscillations in thermodynamic entropy. However the situation becomes more complicated if one were to examine product spaces. If a product state is typical of the space then its marginal entropies cannot oscillate in synch. For every number there is a time where the marginal entropies differ by more than that amount. Thus, marginal algorithmic thermodynamic entropies cannot be synchronized. This result can be found in Chapter 17, where there exists an analogous theorem for discrete ergodic transforms.

Let $(\mathcal{X} \times \mathcal{Y}, \mu \times \nu)$ be a computable product measure space. Let G^t be a transformation group. Let $(\alpha, \beta) \in X \times Y$ with $(\alpha^t, \beta^t) = G^t(\alpha, \beta)$. If $\mathbf{H}_\mu(\alpha, \beta) < -\infty$ and (α, β) has finite mutual information with the halting sequence then $\sup_{t\in[0,1]} \left|\mathbf{H}_\mu(\alpha^t) - \mathbf{H}_\nu(\beta^t)\right| = \infty$.

1.4 Typicality and Newtonian Physics

This manuscript has the first, to the author's knowledge, application of algorithmic information theory to Newtonian Physics. Given a system representing a finite number of mass points, an infinite measure κ is defined equal to the magnitude of the gravitational vector field. In addition, a universal lower computable κ-test $\mathbf{T}_\kappa$ is defined that represents an atypicality score of points in this space. Points at the center of mass points have infinite atypicality, and somewhat surprisingly, $\mathbf{T}_\kappa$ functions like a typical universal test over compact spaces with suitable computability properties. This machinery can be applied to orbits, which are one dimensional rings around mass points. An orbit is specified by $(z, r, \hat{\mathbf{x}}, \hat{\mathbf{y}})$, where z is the center of the orbit, r is its radius, and $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ is its axis of rotation. As a object orbits a body of large mass, it will oscillate in algorithmic typicality, much like with oscillations of algorithmic fine grained entropy over dynamics. A graphical representation of this oscillation can be found in Figure 1.2

Let L be the one-dimensional Lebesgue measure. Let κ be a computable system with mass point $z \in \mathbb{R}^3$ and O be an orbit centered at z whose encoding has finite mutual information with the halting sequence. There is a $c \in \mathbb{N}$ where for all $n \in \mathbb{N}$,

$$2^{-n-c\mathbf{K}(n)} < L\{x : x \in O, \mathbf{T}_\kappa(x) > 2^n\}.$$

Figure 1.2: Orbits in Euclidean space will oscillate in typical measure $\mathbf{T}_\kappa$. Larger oscillations will be less frequent.

The above result shows that non-exotic orbits will oscillate in typicallity. The following result extends this result to the surfaces of spheres. Given a non-exotic sphere C, there is a lower bound on the measure of atypical points on the surface of C.

Let κ be a computable system. If sphere $C = (z, r)$ has finite mutual information with the halting sequence then there is a $c \in \mathbb{N}$ where for all $n \in \mathbb{N}$, $2^{-n-c\mathbf{K}(n)} < L_C\{\alpha : \|\alpha - z\| = r, \mathbf{T}_\kappa(\alpha) > 2^n\}$, where L_C is the uniform measure over C.

1.5 The Kolmogorov Complexity of Black Holes

This manuscript details and expands on the work of [BS18], which introduces the Kolmogorov complexity of black holes. Black holes are modelled as quantum circuits, or unitary matrices in $SU(2^K)$, where K is the number of qubits. The $SU(2^K)$ space is partitioned into cells by epsilon balls. This induces a graph where an edge is between two vertices (cells) if there is a k-local all-to-all gate between the two unitary operators. The k-local, all-to-all, requirement matches certain properties of black holes (namely that they are fast scramblers). Each vertex is assigned a label and the complexity of a unitary operator is the Kolmogorov complexity of the label of the cell. Dynamics of the black hole are modelled as a random walk along the graph, as seen in Figure 1.3.

Why is complexity important in the study of black holes? With this definition, [Sus20] conjectured the *Complexity/Volume Correspondence*, which states the growth of the volume of wormholes (Einstein-Rosen bridges) in black holes match their Kolmogorov complexity/time profile. The benefit to this definition of the Kolmogorov complexity of black holes can be summarized by the following statement.

The study of the Kolmogorov complexity of black holes can be reduced to the study of a fictious particle moving in the $SU(n)$ space.

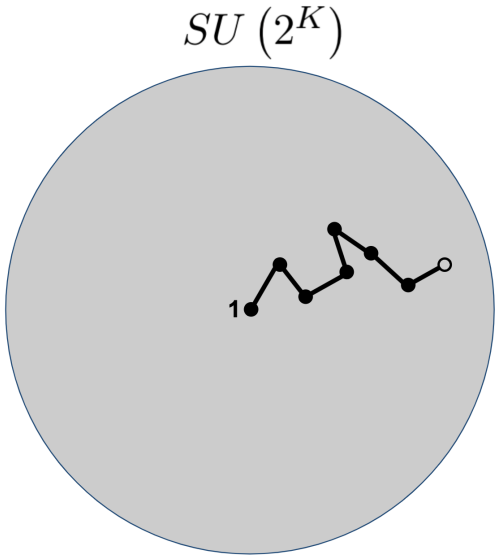


Figure 1.3: The dynamics of black holes can be absracted as a random walk in $SU(2^K)$. The starting position is the identity matrix $\mathbf{1}$.

1.6 The Independence Postulate

The Many Worlds Theory [Eve57] was formulated by Hugh Everett as a solution to the measurement problem of Quantum Mechanics. Branching (a.k.a. splitting of worlds) occurs during any process that magnifies microscopic superpositions to the macroscale. Thus occurs in event including human measurements such as the double slit experiment.

Conflicts emerge between the Many Worlds Theory and Algorithmic Information Theory. In particular, the Independence Postulate [Lev84, Lev13] is a finitary Church-Turing thesis, postulating that certain infinite and finite sequences cannot be found in nature i.e. have high addresses. One such class of forbidden sequences are large prefixes of the halting sequence. If a forbidden sequence is found in nature an "information leak" occurs. In this manuscript we detail the following assertion.

> *In the Many World Theory information leaks can occur.*

Another theory that comes into conflict with the Independence Postulate is Constructor Theory. Constructor Theory aims to unify many areas of science with counterfactuals. Counterfactuals describe whether a process can or cannot occur. The central tenet of Constructor Theory is as follows:

> *All other laws of physics are expressible entirely in terms of statements about which physical transformations are possible and which are impossible, and why.*

However the Independence Posultate poses the following unresolved question:

> *Is it possible or impossible to create or find large prefixes of the halting sequence?*

Chapter 2

Conventions

The following chapter details the conventions in algorithmic information theory, which will be used throughout the manuscript. We use $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\{0,1\}$, $\{0,1\}^*$, and $\{0,1\}^\infty$ to denote natural numbers, integers, rational numbers, reals, complex numbers, bits, finite strings, and infinite sequences. $\{0,1\}^{*\infty} \stackrel{\text{def}}{=} \{0,1\}^* \cup \{0,1\}^\infty$. $\|x\|$ denotes the length of the string. Let $X_{\geq 0}$ and $X_{>0}$ be the sets of non-negative and of positive elements of X. $[A] \stackrel{\text{def}}{=} 1$ if statement A holds, else $[A] \stackrel{\text{def}}{=} 0$. For set of strings $A \subseteq \{0,1\}^*$, $[\![A]\!] = \{x\alpha : x \in A, \alpha \in \{0,1\}^\infty\}$. When it is clear from the context, we will use natural numbers and other finite objects interchangeably with their binary representations. We let $[A] = 1$ if mathematical statement A is true, otherwise $[A] = 0$.

The ith bit of $\alpha \in \{0,1\}^{*\infty}$ is denoted α_i, and its n bit prefix is denoted $\alpha_{\leq n}$. $\langle x \rangle \in \{0,1\}^*$ for $x \in \{0,1\}^*$ is the self-delimiting code that doubles every bit of x and changes the last bit of the result. For positive real functions f, by $<^+ f$, $>^+ f$, $=^+ f$, and $<^{\log} f$, $>^{\log} f$, $\sim f$ we denote $\leq f + O(1)$, $\geq f - O(1)$, $= f \pm O(1)$ and $\leq f + O(\log(f+1))$, $\geq f - O(\log(f+1))$, $= f \pm O(\log(f+1))$. Furthermore, $\stackrel{*}{<} f$, $\stackrel{*}{>} f$ denotes $< O(1)f$ and $> f/O(1)$. The term and $\stackrel{*}{=} f$ is used to denote $\stackrel{*}{>} f$ and $\stackrel{*}{<} f$.

A probability measure Q over $\{0,1\}^*$ is elementary if it has finite support and range that is a subset of rationals. Elementary probability measures can be encoded into finite strings $\langle Q \rangle$ in the standard way.

2.1 Algorithmic Information Theory

$T_y(x)$ is the output of algorithm T (or $\perp$ if it does not halt) on input $x \in \{0,1\}^*$ and auxiliary input $y \in \{0,1\}^{*\infty}$. T is prefix-free if for all $x, s \in \{0,1\}^*$ with $s \neq \emptyset$, either $T_y(x) = \perp$ or $T_y(xs) = \perp$. The complexity of $x \in \{0,1\}^*$ with respect to T_y is $\mathbf{K}_T(x|y) \stackrel{\text{def}}{=} \inf\{\|p\| : T_y(p) = x\}$.

There exist optimal for $\mathbf{K}$ prefix-free algorithms U, meaning that for all prefix-free algorithms T, there exists $c_T \in \mathbb{N}$, where $\mathbf{K}_U(x|y) \leq \mathbf{K}_T(x|y) + c_T$ for all $x \in \{0,1\}^*$ and $y \in \{0,1\}^{*\infty}$. For example, one can take a universal prefix-free algorithm U, where for each prefix-free algorithm T, there exists $t \in \{0,1\}^*$, with $U_y(tx) = T_y(x)$ for all $x \in \{0,1\}^*$ and $y \in \{0,1\}^{*\infty}$. $\mathbf{K}(x|y) \stackrel{\text{def}}{=} \mathbf{K}_U(x|y)$ is the Kolmogorov complexity of $x \in \{0,1\}^*$ relative to $y \in \{0,1\}^{*\infty}$. Similarly, plain Kolmogorov complexity, $\mathbf{C}(x|y)$, is defined using algorithms that have a readable delimiter symbol # at the end of their inputs.

The chain rule is $\mathbf{K}(x,y) =^+ \mathbf{K}(x) + \mathbf{K}(y|x, \mathbf{K}(x))$. The algorithmic probability is $\mathbf{m}(x|y) = \sum\{2^{-\|p\|} : U_y(p) = x\}$. By the coding theorem $\mathbf{K}(x|y) =^+ -\log \mathbf{m}(x|y)$. The amount of mutual information between two strings x and y is $\mathbf{I}(x : y) = \mathbf{K}(x) + \mathbf{K}(y) - \mathbf{K}(x,y)$. By the chain

rule $\mathbf{K}(x,y) =^+ \mathbf{K}(x) + \mathbf{K}(y|x,\mathbf{K}(x))$. The halting sequence $\mathcal{H} \in \{0,1\}^\infty$ is the infinite string where $\mathcal{H}_i \stackrel{\text{def}}{=} [U(i) \text{ halts}]$ for all $i \in \mathbb{N}$. The amount of information that $\mathcal{H}$ has about $x \in \{0,1\}^*$ is $\mathbf{I}(x;\mathcal{H}) = \mathbf{K}(x) - \mathbf{K}(x|\mathcal{H})$. The randomness deficiency of $x \in \{0,1\}^*$ with respect to elementary probability P over $\{0,1\}^*$ is $\mathbf{d}(x|P) = \lfloor -\log P(x) - \mathbf{K}(x|\langle P\rangle)\rfloor$. we say $t : \{0,1\}^* \to \mathbb{R}_{\geq 0}$ is a P-test, for some probability P, if $\sum_x t(x)P(x) \leq 1$. Let $\mathbf{t}_P$ be a universal lower computable P-test, where for any other lower computable P-test t, $\mathbf{t}_P(x) \stackrel{*}{>} \mathbf{m}(t)t(x)$. Then by the universality of the deficiency of randomness, [GÓ1], $\mathbf{d}(x|P) =^+ \log \mathbf{t}_P(x)$. The transform of a probability Q by $f : \{0,1\}^* \to \{0,1\}^*$, is the probability fQ, where $fQ(x) = \sum_{f(y)=x} Q(y)$. Both randomness deficiency and information enjoy conservation inequalities.

Theorem 1 (See [GÓ1]) $\mathbf{d}(f(x)|fQ) <^+ \mathbf{d}(x|Q)$.

Theorem 2 ([Lev84]) $\mathbf{I}(f(x) : y) <^+ \mathbf{I}(x : y)$.

Proof. Due to the chain rule, $\mathbf{K}(x,y,z) <^+ \mathbf{K}(x,\mathbf{K}(x))+\mathbf{K}(y|x,\mathbf{K}(x))+\mathbf{K}(z|x,\mathbf{K}(x)) =^+ \mathbf{K}(x,y)+\mathbf{K}(x,z)-\mathbf{K}(x)$, since $\mathbf{K}(y,z|t) <^+ \mathbf{K}(y|t)+\mathbf{K}(z|t)$. So $\mathbf{I}((z,x) : y) >^+ \mathbf{I}(x : y)$. The statement follows from $\mathbf{I}(z : y) =^+ \mathbf{I}((z,x) : y)$ for $x = A(z)$ since z andd $(z, A(z))$ are computable from each other.

Lemma 1 ([Eps22a]) $\mathbf{I}(f(a);\mathcal{H}) <^+ \mathbf{I}(a;\mathcal{H}) + \mathbf{K}(f)$.

Proof.

$$\mathbf{I}(a;\mathcal{H}) = \mathbf{K}(a) - \mathbf{K}(a|\mathcal{H}) >^+ \mathbf{K}(a,f(a)) - \mathbf{K}(a,f(a)|\mathcal{H}) - \mathbf{K}(f).$$

The chain rule applied twice results in

$$\begin{aligned}
\mathbf{I}(a;\mathcal{H}) + \mathbf{K}(f) &>^+ \mathbf{K}(f(a)) + \mathbf{K}(a|f(a),\mathbf{K}(f(a))) - (\mathbf{K}(f(a)|\mathcal{H}) + \mathbf{K}(a|f(a),\mathbf{K}(f(a)|\mathcal{H}),\mathcal{H})\\
&=^+ \mathbf{I}(f(a);\mathcal{H}) + \mathbf{K}(a|f(a),\mathbf{K}(f(a))) - \mathbf{K}(a|f(a),\mathbf{K}(f(a)|\mathcal{H}),\mathcal{H})\\
&=^+ \mathbf{I}(f(a);\mathcal{H}) + \mathbf{K}(a|f(a),\mathbf{K}(f(a))) - \mathbf{K}(a|f(a),\mathbf{K}(f(a)),\mathbf{K}(f(a)|\mathcal{H}),\mathcal{H})\\
&>^+ \mathbf{I}(f(a);\mathcal{H}).
\end{aligned}$$

□

Lemma 2 *For program q that computes probability p over $\mathbb{N}$,* $\mathbf{E}_{a\sim p}\left[2^{\mathbf{I}(\langle q,a\rangle;\mathcal{H})}\right] \stackrel{*}{<} 2^{\mathbf{I}(q;\mathcal{H})}$.

Proof. The goal is to prove $\sum_a p(a)\mathbf{m}(a,q/\mathcal{H})/\mathbf{m}(a,q) \stackrel{*}{<} \mathbf{m}(q/\mathcal{H})/\mathbf{m}(q)$. Rewriting this inequality, it suffices to prove $\sum_a \big(\mathbf{m}(q)p(a)/\mathbf{m}(a,q)\big)\big(\mathbf{m}(a,q/\mathcal{H})/\mathbf{m}(q/\mathcal{H})\big) \stackrel{*}{<} 1$. The term $\mathbf{m}(q)p(a)/\mathbf{m}(a,q) \stackrel{*}{<} 1$ because $\mathbf{K}(q)-\log p(a) >^+ \mathbf{K}(a,q)$. Furthermore, it follows directly that $\sum_a \mathbf{m}(a,q/\mathcal{H})/\mathbf{m}(q/\mathcal{H}) \stackrel{*}{<} 1$. □

The stochasticity of a string $x \in \{0,1\}^*$ is $\mathbf{Ks}(x) = \min_{\text{Elementary } Q} \mathbf{K}(Q)+3\log\max\{\mathbf{d}(x|Q),1\}$. Strings with high stochasticity measures are exotic, in that they have high mutual information with the halting sequence. A proof to the following result can be found in Lemma 35 of Appendix A.

Lemma 3 ([Lev16, Eps21b]) $\mathbf{Ks}(x) < \mathbf{I}(x;\mathcal{H}) + O(\mathbf{K}(\mathbf{I}(x;\mathcal{H})))$.

The following definition is from [Lev74] .

Definition 1 (Information) *For infinite sequences $\alpha, \beta \in \{0,1\}^\infty$, their mutual information is defined to be* $\mathbf{I}(\alpha:\beta) = \log \sum_{x,y\in\{0,1\}^*} 2^{\mathbf{I}(x:y)-\mathbf{K}(x|\alpha)-\mathbf{K}(y|\beta)}$.

It is easy to see that $\mathbf{I}(f(\alpha):\beta) <^+ \mathbf{I}(\alpha:\beta) + \mathbf{K}(f)$.

Definition 2 (Randomness Deficiency of Infinite Sequences) *The deficiency of randomness of an infinite sequence $\alpha \in \{0,1\}^\infty$, relative to $x \in \{0,1\}^{*\infty}$ and (not necessarily probability) measure μ over $\{0,1\}^\infty$ is*

$$\mathbf{D}(\alpha|\mu, x) = \sup_n -\log \mu(\alpha[0\ldots n])) - \mathbf{K}(\alpha[0\ldots n]|\mu, x).$$

We say $\mathbf{D}(\alpha|\mu) = \mathbf{D}(\alpha|\mu, \emptyset)$.

2.2 Algorithmic Information Between Probabilities

We can generalize from information from strings to information between arbitrary probability measures over strings.

Definition 3 (Information, Probabilities)
For semi-measures p and q over $\{0,1\}^*$, $\mathbf{I}_{\text{Prob}}(p:q) = \log \sum_{x,y\in\{0,1\}^*} 2^{\mathbf{I}(x:y)} p(x)q(y)$.

Definition 4 (Channel) *A channel $f : \{0,1\}^* \times \{0,1\}^* \to \mathbb{R}_{\geq 0}$ has $f(\cdot|x)$ being a probability measure over $\{0,1\}^*$ for each $x \in \{0,1\}^*$. For probability p, channel f, $fp(x) = \sum_z f(x|z)p(z)$.*

Lemma 4 *Let ψ_a be an enumerable semi-measure, semi-computable relative to a.*
$\sum_c 2^{\mathbf{I}(\langle a,c\rangle:b)} \psi_a(c) \stackrel{*}{<} 2^{\mathbf{I}(a:b)}/\mathbf{m}(\psi)$.

Proof. This requires a slight modification of the proof of Proposition 2 in [Lev84], by requiring ψ to have a as auxilliary information. For completeness, we reproduce the proof. We need to show $\mathbf{m}(a,b)/(\mathbf{m}(a)\mathbf{m}(b)) \stackrel{*}{>} \sum_c (\mathbf{m}(a,b,c)/(\mathbf{m}(b)\mathbf{m}(a,c)))\mathbf{m}(\psi)\psi_a(c)$, or $\sum_c (\mathbf{m}(a,b,c)/\mathbf{m}(a,c))\mathbf{m}(c|a) \stackrel{*}{<} \mathbf{m}(a,b)/\mathbf{m}(a)$, since $\mathbf{m}(c|a) \stackrel{*}{>} \mathbf{m}(\psi)\psi_a(c)$. Rewrite it $\sum_c \mathbf{m}(c|a)\mathbf{m}(a,b,c)/\mathbf{m}(a,c) \stackrel{*}{<} \mathbf{m}(a,b)/\mathbf{m}(a)$ or $\sum_c \mathbf{m}(c|a)\mathbf{m}(a)\mathbf{m}(a,b,c)/\mathbf{m}(a,c) \stackrel{*}{<} \mathbf{m}(a,b)$. The latter is obvious since $\mathbf{m}(c|a)\mathbf{m}(a) \stackrel{*}{<} \mathbf{m}(a,c)$ and $\sum_c \mathbf{m}(a,b,c) \stackrel{*}{<} \mathbf{m}(a,b)$. □

Theorem 3 *For probabilities p and q over $\{0,1\}^*$, computable channel f,* $\mathbf{I}_{\text{Prob}}(fp:q) <^+ \mathbf{I}_{\text{Prob}}(p:q)$.

Proof. Using Lemma 1,

$$\mathbf{I}_{\text{Prob}}(fp:q) = \log \sum_{x,y} 2^{\mathbf{I}(x:y)} \sum_z f(x|z)p(z)q(y) <^+ \log \sum_{y,z} q(y)p(z) \sum_x 2^{\mathbf{I}((x,z):y)} f(x|z).$$

Using Lemma 4,

$$\mathbf{I}_{\text{Prob}}(fp:q) <^+ \log \sum_{z,y} q(y)p(z)2^{\mathbf{I}(z:y)} =^+ \mathbf{I}_{\text{Prob}}(p:q).$$

□

Theorem 4 *For enumerable semi-measures p, q,* $\mathbf{I}_{\text{Prob}}(p:q) <^+ \mathbf{I}(\langle p\rangle : \langle q\rangle)$.

Proof. Let T be a Turing machine, that when given an encoding of a lower semi-computable probabiliy p and an input x, lower enumerates $p(x)$. $\mathbf{I}_{\text{Prob}}(p : q) = \log \sum_{x,y} 2^{\mathbf{I}(x:y)} T_p(x) T_q(y)$. Using Theorems 3 and 4,

$$\begin{aligned}
&\mathbf{I}_{\text{Prob}}(p : q) \\
<^+ &\log \sum_{x,y} 2^{\mathbf{I}(\langle x,p\rangle:y)} T_p(x) T_q(y) \\
<^+ &\log \sum_{y} 2^{\mathbf{I}(\langle p\rangle:y)} q(y)/\mathbf{m}(T) \\
<^+ &\log \sum_{y} 2^{\mathbf{I}(\langle p\rangle:\langle y,q\rangle)} q(y)/\mathbf{m}(T) \\
<^+ &\log 2^{\mathbf{I}(\langle p\rangle:\langle q\rangle)}/\mathbf{m}(T)^2 \\
<^+ &\mathbf{I}(\langle p\rangle : \langle q\rangle).
\end{aligned}$$

Thus processing cannot increase information between two probabilities. If the the probability measure is concentrated at a single point, then it contains self-information equal to the complexity of that point. If the probability measure is spread out, then it is white noise, and contains no self-information. Some examples are as follows.

Example 1

- *In general, a probability p, will have low $\mathbf{I}_{\text{Prob}}(p : p)$ if it has large measure on simple strings, or low measure on a large number of complex strings, or some combination of the two.*
- *If probability p is concentrated on a single string x, then $\mathbf{I}_{\text{Prob}}(p : p) = \mathbf{K}(x)$.*
- *The uniform distribution over strings of length n has self information equal to (up to an additive constant) $\mathbf{K}(n)$.*
- *There are semi-measures that have infinite self information. Let α_n be the n bit prefix of a Martin Löf random sequence α and $n \in [2, \infty)$. Semi-measure $p(x) = [x = \alpha_n] n^{-2}$ has $\mathbf{I}_{\text{Prob}}(p : p) = \infty$.*
- *The universal semi-measure $\mathbf{m}$ has no self information.*

Example 2 (Uniform Spread) *An example channel f has $f(\cdot|x)$ be the uniform distribution over strings of length $\|x\|$. This is a cannonical spread function. Thus if p is a probability measure concentrated on a single string, then $\mathbf{I}_{\text{Prob}}(p : p) = \mathbf{K}(x)$, and $\mathbf{I}(fp : fp) =^+ \mathbf{K}(\|x\|)$. Thus f results in a decrease of self-information of p. This decrease of information occurs over all probabilities and computable channels.*

Part II

Quantum Mechanics

Chapter 3

Introduction

Classical information theory studies the communication of bits across a noisy channel. Quantum Information Theory (QIT) studies the kind of information ("quantum information") which is transmitted by microparticles from a preparation device (sender) to a measuring apparatus (receiver) in a quantum mechanical experiment—in other words, the distinction between carriers of classical and quantum information becomes essential. The notion of a qubit can be defined at an abstract level, without giving preference to any particular physical system such as a spin-12 particle or a photon. Qubits behave very differently than bits. To start, qubits can be in a linear superposition between 0 and 1. Qubits can have entanglement, where two objects at a distance become a single entity. The study of entanglement and in particular the question how it can be quantified is therefore a central topic within quantum information theory. However, due to the no-cloning theorem [WZ82], instant communication is not possible. Some other aspects of QIT are as follows.

1. **Quantum Computing**: includes hardware (quantum computers), software, algorithm such as Shor's factoring algorithm or Grover's algorithm, and applications.

2. **Quantum Communication**: quantum networking, quantum internet, quantum cryptography.

3. **Applications in Physics**: applications to convex optimizations, black holes, and exotic quantum phases of matter.

4. **Quantum Shannon Theory**: quantum channels, quantum protocols, quantum information and entropy.

One aspect of Quantum Shannon Theory (QST) that has had relatively little study is its relationship to Algorithmic Information Theory (AIT). AIT, in part, is the study of the information content of individual strings. A string is random if it cannot be compressed with respect to a universal Turing machine. This paper surveys the current state of research of QST and AIT and provides unpublished results from the author. Hopefully it will convince the reader that there is a fruitful area of research of QST and AIT. Some areas of this intersection include algorthmic content of quantum states, how typical a quantum state is with respect to a quantum source, and how to quantify the algorithmic content of a measurement. One can also gain further insight into quantum transformations, such as purification, decoherence, and approximations to quantum cloning.

As this manuscript will show, there are some aspects of AIT that directly transfer over to quantum mechanics. This includes comparable definitions of complexity, and conservation inequalities. In addition, there exist quantum versions of the EL Theorem, [Lev16, Eps19c] and the Outlier

Theorem, [Eps21b]. However there are some aspects of AIT that are different in the context of quantum mechanics. This includes the fact the self information of most quantum pure states is zero, with $\mathbf{I}(|\psi\rangle : |\psi\rangle) \approx 0$. This has implications on the algorithmic content of measurements and decoherence. The main quantum mechanical areas covered in this manuscript are

- **Chapter 4:** This chapter covers the background material on quantum mechanics needed for the article.
- **Chapter 5:** Three different algorithmic measures of quantum states are covered in the manuscript.. In this chapter, Gács complexity and Vitányi complexity are detailed. Their properties are described, including an addition inequality, a Quantum EL Theorem, and a generalized no-cloning theorem. Inequalities between the two complexities are proven.
- **Chapter 6**. The algorithmic entropy called BvL complexity is detailed. It is compared to Gács complexity and Kolmogorov complexity. A no cloning theorem is proved.
- **Chapter 7:** The notion of the algorithmic typicality of one quantum state with respect to another quantum state is introduced. Typicality is conserved with respect to quantum operations. A quantum outlier theorem is proven. This states that non-exotic projections must have atypical pure states in their images.
- **Chapter 8:** The definition of quantum algorithmic information is introduced. Quantum information differs from classical algorithmic information in that an overwhelming majority of pure states have negligible self-information. Information is conserved over quantum operations, with implications to quantum cloning, quantum decoherence, and purification.
- **Chapter 9:** Quantum algorithmic information upper bounds the amount of classical information produced by quantum measurements. Given a quantum measurement, for an overwhelming majority of pure states, the measurement will be random noise.
- **Chapter ??:** This chapter shows that with respect to a system-environment dynamics, an overwhelming majority of non-pointer states will decohere into algorithmic garbage.
- **Chapter 10:** A quantum equivalent to Martin Löf random sequence is introduced. Such quantum random states have incompressible initial segments with respect to a new measure quantum complexity called Quantum Operation Complexity. This complexity term measures the cost of approximating a state with a classical and quantum component.
- **Chapter 11:** This chapter concludes the quantum mechanical section of the manuscript with a discussion of the boundary between quantum information and classical information. We show that measurements are necessary to produce distributions over quantum states that have cloneable information.
- **Appendix A:** An extended coding theorem is proved with applications to proving inequalities of quantum complexities and the relation between dynamics and coarse grained entropy.

Chapter 4

Background

4.1 Quantum Mechanic Tools

We use the standard model of qubits used throughout quantum information theory. We deal with finite N dimensional Hilbert spaces $\mathcal{H}_N$, with bases $|\alpha_1\rangle, |\alpha_2\rangle, \ldots, |\alpha_n\rangle$. We assume $\mathcal{H}_{n+1} \supseteq \mathcal{H}_n$ and the bases for $\mathcal{H}_n$ are the beginning of that of $\mathcal{H}_{n+1}$. An n qubit space is denoted by $\mathcal{Q}_n = \bigotimes_{i=1}^n \mathcal{Q}_1$, where qubit space $\mathcal{Q}_1$ has bases $|0\rangle$ and $|1\rangle$. For $x \in \Sigma^n$ we use $|x\rangle \in \mathcal{Q}_n$ to denote $\bigotimes_{i=1}^n |x[i]\rangle$. The space $\mathcal{Q}_n$ has 2^n dimensions and we identify it with $\mathcal{H}_{2^n}$.

A pure quantum state $|\phi\rangle$ of length n is represented as a unit vector in $\mathcal{Q}_n$. Its corresponding element in the dual space is denoted by $\langle\phi|$. The tensor product of two vectors is denoted by $|\phi\rangle \otimes |\psi\rangle = |\phi\rangle |\psi\rangle = |\phi\psi\rangle$. The inner product of $|\psi\rangle$ and $\langle\phi|$ is denoted by $\langle\psi|\phi\rangle$.

The symbol Tr denotes the trace operation. The conjugate transpose of a matrix M is denoted by M^*. For Hermitian matrix with eigenvalue decomposition $A = \sum a_i |\psi_i\rangle \langle\psi_i|$, $|A| = \sum |a_i| |\psi_i\rangle \langle\psi_i|$. The tensor product of two matrices is denoted by $A \otimes B$. Projection matrices are Hermitian matrices with eigenvalues in $\{0, 1\}$. For tensor product space $\mathcal{G}_X \otimes \mathcal{G}_Y$, the partial trace is denoted by Tr_Y. For $B^X = \mathrm{Tr}_Y B$, $\mathrm{Tr}(A \cdot B^X) = \mathrm{Tr}((A \otimes I) \cdot B)$, which is used frequently throughout the manuscript. For positive semidefinite matrices, σ and ρ we say $\sigma \leq \rho$ if $\rho - \sigma$ is positive semidefinite. For positive semidefinite matrices A, B, C, if $A \leq B$ then $\mathrm{Tr}AC \leq \mathrm{Tr}BC$. Mixed states are represented by density matrices, which are, self adjoint, positive semidefinite, operators of trace 1. A semi-density matrix has non-negative trace less than or equal to 1. The von Neumann entropy of a density matrix σ with orthogonal decomposition $\sum p_i |\psi_i\rangle \langle\psi_i|$ is $S(\sigma) = -\sum p_i \log p_i$. Relative entropy is $S(\rho, \sigma) = -\mathrm{Tr}\rho \log \sigma - S(\rho)$. Holevo's Chi is defined as follows. For an ensemble $\mathcal{E} = \{(\rho_i, p_i)\}$, with $\rho = \sum_i p_i \rho_i$, $\chi(\mathcal{E}) = S(\rho) - \sum_i p_i S(\rho_i)$.

A number is algebraic if it is the root of a polynomial with rational coefficients. A pure quantum state $|\phi\rangle$ and (semi)density matrix σ are called *elementary* if their real and imaginary components have algebraic coefficients. The usual operations of linear algrebra such as orthogonalizaton, finding eigenvalues and vectors can stay within the realm of elementary constructs. Furthermore, elementary objects can be encoded into strings or integers and be the output of halting programs. Therefore one can use the terminology $\mathbf{K}(|\phi\rangle)$ and $\mathbf{K}(\sigma)$, and also $\mathbf{m}(|\phi\rangle)$ and $\mathbf{m}(\sigma)$.

We say program $q \in \{0, 1\}^*$ lower computes positive semidefinite matrix σ if, given as input to universal Turing machine U, the machine U reads $\leq \|q\|$ bits and outputs, with or without halting, a sequence of elementary semi-density matrices $\{\sigma_i\}$ such that $\sigma_i \leq \sigma_{i+1}$ and $\lim_{i\to\infty} \sigma_i = \sigma$. A matrix is lower computable if there is a program that lower computes it.

4.2 Quantum Operations

A quantum operation is the most general type of operation than can be applied to a quantum state. In Chapters 7 and 8, conservation inequalities will be proven with respect to quantum operations. A map transforming a quantum state σ to $\varepsilon(\sigma)$ is a quantum operation if it satisfies the following three requirements

1. The map of ε is positive and trace preserving, with $\text{Tr}(\sigma) = \text{Tr}(\varepsilon(\sigma))$.
2. The map is linear with $\varepsilon(\sum_i p_i \sigma_i) = \sum_i p_i \varepsilon(\sigma_i)$.
3. The map is completely positive, were any map of the form $\varepsilon \otimes \mathbf{M}$ acting on the extended Hilbert space is also positive.

Another means to describe quantum operations is through a series of operators. A quantum operation ε on mixed state σ_A can be seen as the appending of an ancillia state σ_b, applying a unitary transform U, then tracing out the ancillia system with

$$\varepsilon(\sigma_A) = \text{Tr}_B \left(U(\sigma_A \otimes \sigma_B) U^* \right). \tag{4.1}$$

A third way to characterize a quantum operation is through Kraus operators, which can be derived using an algebraic reworking of Equation 4.1. Map ε is a quantum operation iff it can be represented (not necessarily uniquely) using a set of matrices $\{M_i\}$ where $\varepsilon(\sigma) = \sum_i M_i \varepsilon M_i^*$ and $\sum_i M_i^* M_i = I$, where I is the identity matrix.

A quantum operation ε is elementary iff it admits a represented of the form in Equation 4.1 where B, U, and σ_B are each elementary, in that they each can be encoded with a finite string. The encoding of an elementary quantum operation is denoted by $\langle \varepsilon \rangle = \langle B \rangle \langle U \rangle \langle \sigma_B \rangle$. Each elementary quantum operation admits an elementary Kraus operator representation $\{M_i\}$, in that each M_i is an elementary matrix. This elementary Kraus operator is computable from $\langle \varepsilon \rangle$.

Chapter 5

Gács and Vitányi Complexity

In traditional Quantum Information Theory, the entropy of a pure or mixed state ρ is measured by the von Neumann entropy, $S(\rho)$ defined in Chapter 4. This measures the amount of mixing that a state has, or put another way, the amount of departure of the the state from a pure state. Thus the von Neumann entropy of a pure state is 0 and that of a totally mixed n qubit state is n. The min-entropy of a quantum state is

$$\mathcal{H}_{\min}(\rho) = \max_{\Pi} \frac{1}{\max_i \operatorname{Tr}(\Pi_i \rho)},$$

where Π is over all projective measurements (as described in chapter 9). The min-entropy can be interpreted as the distance of a state from a maximally entangled state. This concept is useful in quantum cryptography, in the context of privacy amplification.

However these scores fail to measure how complicated the state is, or rather the amount of computational resources it takes to create the state. For example take states $|0^n\rangle$ and $|x\rangle$ where x is a random n-bit string. For both strings, their von Neumann entropy is 0. Algorithmic Information Theory provides a suitable area of research into the individual complexity of sequences of 1s and 0s. This chapter details the application of AIT to quantum mechanics toward defining the algorithmic content of quantum states.

The formal study of Algorithmic Information Theory and Quantum Mechanics began with the introduction of three independent measures of the algorithmic content of a mixed or pure quantum state, detailed in the papers [BvL01, GÓ1, Vit01]. In [BvL01], the complexity of a pure or mixed quantum state $|\psi\rangle$ is measured by the length of the smallest input to a universal quantum Turing machine that outputs a good approximation of σ. Vitányi complexity [Vit01] measures the entropy of a pure state $|\psi\rangle$ as the amount of classical information needed to reproduce a good approximation of $|\psi\rangle$. Gács complexity measures the entropy of a pure or mixed quantum state by using a quantum analogue of the universal semi-measure $\mathbf{m}$. In this chapter, Gács complexity and Vitányi complexity are studied. In Chapter 6, BvL complexity is detailed.

5.1 Vitányi Complexity

One approach to measuring the algorithmic complexity of a state is to use a normal Turing machine to measure the information content of a quantum state. Vitányi complexity, [Vit01], of a pure state $|\psi\rangle$ is equal to the minimum size of a program to a universal Turing machine that outputs an approximation that is an elementary pure state $|\theta\rangle$ of the target state plus a score of their closeness. The cost term is $-\log\langle\theta|\psi\rangle$. We use a slightly different definition than the original [Vit01], in that

we use a classical universal Turing machine and not a quantum universal Turing machine . Let n be the number of qubits. The Vitányi complexity, $\mathbf{Hv}$, is defined by

$$\mathbf{Hv}(|\psi\rangle) = \min_{\text{Elementary } |\theta\rangle \in \mathcal{Q}_n} \mathbf{K}(|\theta\rangle\,|n) - \log |\,\langle\psi|\theta\rangle\,|^2.$$

Theorem 5 ([Vit01]) *For all n qubit pure states $|\psi\rangle$, $\mathbf{Hv}(|\psi\rangle) <^+ 2n$.*

Proof. Let $|\psi\rangle \in \mathcal{Q}_n$ with basis vectors $\{e_i\}_{i=1}^{2^n}$ and $\sum_{i=1}^{2^n} |e_i|\psi\rangle^2 = 1$. So there is an i with $\langle e_i|\psi\rangle^2 \geq 2^{-n}$. Let p be an $n + O(1)$ program that constructs $|e_i\rangle$. So $\mathbf{Hv}(|psi\rangle) \leq \|p\| - \log 2^{-n} <^+ 2n$. □

Exercise 1 ([Vit01]) *Show that for the uniform probability Λ over n qubit states, $\Lambda\{|\psi\rangle : \mathbf{Hv}(|\psi\rangle \geq n - c)\} > 1 - 1/2^c$.*

Exercise 2 ((Hard)[GÓ1]) *Prove that for large enough n, there are states $|\psi\rangle \in \mathcal{Q}_n$ with $\mathbf{Hv}(|\psi\rangle) > 2n - 2\log n$.*

Theorem 6 ([Vit01]) *Let $|\psi\rangle$ be a basis vector for an elementary orthonormal basis $\mathcal{B}$, $\mathbf{K}(|\psi\rangle\,|n) <^+ \mathbf{Hv}(|\psi\rangle) + \mathbf{K}(\mathcal{B}|n)$.*

Proof. Let elementary $|\phi\rangle \in \mathcal{Q}_n$ be defined such that $\mathbf{Hv}(|\psi\rangle) = \mathbf{K}(|\phi\rangle\,|n) - \log \langle\phi|\psi\rangle^2$. Let $\{|e_i\rangle\}$ be the basis vectors of $\mathcal{B}$. So $\sum_{|e_i\rangle} \langle\phi|e_i\rangle^2 = 1$. Let $n_i = \langle\phi|e_i\rangle^2$. So if $|\psi\rangle = e_i$, we have $\mathbf{Hv}(|\psi\rangle) = \mathbf{K}(|\phi\rangle\,|n) + n_i$. Since $\sum_i n_i = 1$, there is a prefix free code that can identify $|e_i\rangle$. So $\mathbf{K}(|\psi\rangle\,|n) <^+ \mathbf{K}(|\phi\rangle\,|n) - \log n_i + \mathbf{K}(\mathcal{B}|n) <^+ \mathbf{Hv}(|\psi\rangle) + \mathbf{K}(\mathcal{B}|n)$. □ The following corollary shows that over classical strings, Vitányi complexity is equal to Kolmogorov complexity. This is also true for Gács complexity and BvL complexity.

Corollary 1 ([Vit01]) *For $x \in \{0,1\}^n$, $\mathbf{K}(x|n) =^+ \mathbf{Hv}(|x\rangle)$.*

Proof. This follows from the basis $\mathcal{B}$ corresponding to n bit classical strings, with $\mathbf{K}(\mathcal{B}|n) = 0$. □

5.2 Gács Complexity

Gács complexity, [GÓ1], takes a different approach. The Kolmogorov complexity of a string x is equal to, up to an additive factor, $-\log \mathbf{m}(x)$. Similarly Gács complexity is defined using the following universal lower computable semi-density matrix, parametered by $x \in \{0,1\}^*$, with

$$\boldsymbol{\mu}_x = \sum_{\text{Eementary } |\phi\rangle \in \mathcal{Q}_n} \mathbf{m}(|\phi\rangle\,|x, n)\,|\phi\rangle\langle\phi|.$$

The parameter n represents number of qubits used. We use $\boldsymbol{\mu}_X$ to denote the matrix $\boldsymbol{\mu}$ over the Hilbert space denoted by symbol X. The matrix $\boldsymbol{\mu}$ will be very useful in the subsequent chapters. The Gács entropy of a mixed state σ, conditioned on $x \in \{0,1\}^*$ is defined by

$$\mathbf{Hv}(\sigma|x) = \lceil -\log \mathrm{Tr}\boldsymbol{\mu}_x\sigma \rceil.$$

We use the following notation for pure states, with $\mathbf{Hg}(|\phi\rangle\,|x) = \mathbf{Hg}(\,|\phi\rangle\langle\phi|\,|x)$. For empty x we use the notaion $\mathbf{Hg}(\sigma)$. Whereas BvL complexity and Vitányi complexity are defined solely on pure states, this definition also applies to mixed states. This generalizes the definition $\underline{H}$ in [GÓ1], which was solely over pure states.

Note than in [GÓ1], there is another measure of quantum algorithmic entropy, $\overline{H}$, which we will not cover in this manuscript. An infinite version of algorithmic entropy can be found at [BOD14].

5.3 Properties of the Universal Matrix and Gács Complexity

The matrix $\boldsymbol{\mu}$ is important in Algorithmic Information Theory and Quantum Mechanics, as it is the foundation for the information term defined in Chapter 8. The following theorem shows that the lower computable semi-density matrix $\boldsymbol{\mu}$ is universal. It is greater than any other lower computable matrix, weighted by their complexity. This parallels the classical case, where universal measure $\mathbf{m}$ majorizes lower computable semi measure p, with $\mathbf{m}(x) \stackrel{*}{>} \mathbf{m}(p)p(x)$. This theorem is used throughout the paper, and will not be explicitly cited.

Theorem 7 ([GÓ1]) *Let $q \in \{0,1\}^*$ lower compute semi-density matrix A, relativized to the number of qubits n,. Then $\boldsymbol{\mu} \stackrel{*}{>} \mathbf{m}(q|n)A$.*

Proof. A can be composed into a sum $\sum_i p(i) |\psi_i\rangle \langle\psi_i|$, where each $|\psi_i\rangle$ is elementary, p is a semi-measure, with $\sum_i p(i) \leq 1$, and p is lower computable from q and n. Thus,

$$A = \sum_i p(i) |\psi_i\rangle \langle\psi_i| \stackrel{*}{<} \mathbf{m}(p|n)^{-1} \sum_i \mathbf{m}(i|n) |\psi_i\rangle \langle\psi_i| \stackrel{*}{<} \mathbf{m}(q|n)^{-1} \sum_i \mathbf{m}(i|n) |\psi_i\rangle \langle\psi_i| \stackrel{*}{<} \boldsymbol{\mu}/\mathbf{m}(q|n).$$

□

Exercise 3 ([GÓ1]) *Let ν vary over elementary matrices and P vary over elementary projections. Prove $\mu \stackrel{*}{=} \sum_\nu \mathbf{m}(\nu|n)\nu \stackrel{*}{=} \sum_P \mathbf{m}(P|n)P$.*

Theorem 8 ([GÓ1]) $\boldsymbol{\mu}_{ii} \stackrel{*}{=} \mathbf{m}(i|n)$.

Proof. The matrix $\rho = \sum_i \mathbf{m}(i|n) |i\rangle \langle i|$ is lower computable, so $\rho \stackrel{*}{<} \boldsymbol{\mu}$ so $\boldsymbol{\mu}_{ii} \stackrel{*}{>} \mathbf{m}(i|n)$. Furthermore, $f(i) = \langle i| \boldsymbol{\mu} |i\rangle$ is a lower computable semi-measure, so $\mathbf{m}(i|n) \stackrel{*}{>} \boldsymbol{\mu}_{ii}$. □

Theorem 9 ([GÓ1]) $\mathrm{Tr}_Y \boldsymbol{\mu}_{XY} \stackrel{*}{=} \boldsymbol{\mu}_X$.

Proof. Let $\rho = \mathrm{Tr}_Y \boldsymbol{\mu}_{XY}$, which is a lower computable semi-density matrix because one can enumerate elementary pure states $|\psi\rangle \langle\psi|$ in the space XY, take their partial trace, $\mathrm{Tr}_Y |\psi\rangle \langle\psi|$, and then add the resulting pure or mixed state to the sum ρ. Thus $\rho \stackrel{*}{<} \boldsymbol{\mu}_X$. Let $\sigma = \boldsymbol{\mu}_X \otimes |\psi\rangle \langle\psi|$, where $|\psi\rangle$ is a reference elementary state. Thus $\sigma \stackrel{*}{<} \boldsymbol{\mu}_{XY}$ so

$$\boldsymbol{\mu}_X = \mathrm{Tr}_Y \sigma \stackrel{*}{<} \mathrm{Tr}_Y \boldsymbol{\mu}_{XY}.$$

□

Theorem 10 ([GÓ1]) $\mathbf{Hg}(\sigma) <^+ \mathbf{Hg}(\sigma \otimes \rho)$.

Proof. Note that this theorem is not less general than that of Theorem 12, because both σ and ρ can be non-elementary. Using Theorem 9 and the properties of partial trace,

$$2^{-\mathbf{Hg}(\sigma)} \stackrel{*}{>} \mathrm{Tr}\sigma\boldsymbol{\mu}_X \stackrel{*}{>} \mathrm{Tr}\sigma\mathrm{Tr}_Y\boldsymbol{\mu}_{XY} \stackrel{*}{>} \mathrm{Tr}(\sigma \otimes I)\boldsymbol{\mu}_{XY} \stackrel{*}{>} \mathrm{Tr}(\sigma \otimes \rho)\boldsymbol{\mu}_{XY} \stackrel{*}{=} 2^{-\mathbf{Hg}(\sigma\otimes\rho)}.$$

□

Theorem 11 ([GÓ1]) *For mixed state ρ, unitary U, $\mathbf{Hg}(\rho) =^+ \mathbf{H}(U\rho U^*) \pm \mathbf{K}(U)$.*

Proof. This follows from $\boldsymbol{\mu} \overset{*}{>} U\boldsymbol{\mu}U^*$ and $U\boldsymbol{\mu}U^* \overset{*}{>} U(U^*\boldsymbol{\mu}U)U^* \overset{*}{=} \boldsymbol{\mu}$. □

Exercise 4 ([GÓ1]) *Let P be a lower semicomputable projection with $d = \mathrm{Tr}P$. Prove* $\mathbf{Hg}(\sigma) <^+ \mathbf{K}(P|n) + \log d - \log \mathrm{Tr}\sigma P$.

Example 3
Prove the folllowing properties about Gács complexity.

1. $0 <^+ \mathbf{Hg}(\rho) <^+ n$.
2. $\mathbf{Hg}(|0^n\rangle) = O(1)$.
3. $\mathrm{Tr}\boldsymbol{\mu} \overset{*}{=} O(1)$.
4. *For the maximally mixed state* $U = 2^{-n}\mathbf{1}$, $\mathbf{Hg}(U) =^+ n$.
5. *For string* $x \in \{0,1\}^n$, $\mathbf{Hg}(|x\rangle) =^+ \mathbf{K}(x|n)$.
6. *Given mixed states* σ *and* ρ, $\mathbf{Hg}(a\sigma + b\rho) \leq a\mathbf{Hg}(\sigma) + b\mathbf{Hg}(\rho)$.

5.4 Addition Inequality

The addition theorem for classical entropy asserts that the joint entropy for a pair of random variables is equal to the entropy of one plus the conditional entropy of the other, with $\mathcal{H}(\mathcal{X}) + \mathcal{H}(\mathcal{Y}/\mathcal{X}) = \mathcal{H}(\mathcal{X}, \mathcal{Y})$. For algorithmic entropy, the chain rule is slightly more nuanced, with $\mathbf{K}(x) + \mathbf{K}(y|x, \mathbf{K}(x)) =^+ \mathbf{K}(x, y)$. An analogous relationship cannot be true for Gács entropy, **Hg**, since as shown in Theorem 18, there exists elementary $|\phi\rangle$ where $\mathbf{Hg}(|\phi\rangle\,|\phi\rangle) - \mathbf{Hg}(|\phi\rangle)$ can be arbitrarily large, and $\mathbf{Hg}(|\phi\rangle\,/\,|\phi\rangle) =^+ 0$. However, the following theorem shows that a chain rule inequality does hold for **Hg**.

For $n^2 \times n^2$ matrix A, let $A[i,j]$ be the $n \times n$ submatrix of A starting at position $(n(i-1)+1, n(j-1)+1)$. For example for $n = 2$ the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix}$$

has $A[1,1] = \begin{bmatrix} 1 & 2 \\ 5 & 6 \end{bmatrix}$, $A[1,2] = \begin{bmatrix} 3 & 4 \\ 7 & 8 \end{bmatrix}$, $A[2,1] = \begin{bmatrix} 9 & 10 \\ 13 & 14 \end{bmatrix}$, $A[2,2] = \begin{bmatrix} 11 & 12 \\ 15 & 16 \end{bmatrix}$.

For $n^2 \times n^2$ matrix A and $n \times n$ matrix B, let M_{AB} be the $n \times n$ matrix whose (i,j) entry is equal to $\mathrm{Tr}A[i,j]B$. For any $n \times n$ matrix C, in can be seen that $\mathrm{Tr}A(C \otimes B) = \mathrm{Tr}M_{AB}C$. Furthermore if A is lower computable and B is elementary, then M_{AB} is lower computable.

For elementary semi density matrices ρ, we use $\langle \rho, \mathbf{Hg}(\rho)\rangle$ to denote the encoding of the pair of an encoded ρ and an encoded natural number $\mathbf{Hg}(\rho)$.

Theorem 12 ([Eps19a]) *For semi-density matrices σ, ρ, elementary ρ,* $\mathbf{Hg}(\rho) + \mathbf{Hg}(\sigma|\langle \rho, \mathbf{Hg}(\rho)\rangle) <^+ \mathbf{Hg}(\sigma \otimes \rho)$.

Proof. Let $\boldsymbol{\mu}_{2n}$ be the universal lower computable semi density matrix over the space of 2n qubits, $\mathcal{Q}_{2n} = \mathcal{Q}_n \otimes \mathcal{Q}_n = \mathcal{Q}_A \otimes \mathcal{Q}_B$. Let $\boldsymbol{\mu}_n$ be the universal matrix of the space over n qubits. We define the following bilinear function over complex matrixes of size $n \times n$, with $T(\nu, \delta) = \mathrm{Tr}\boldsymbol{\mu}_{2n}(\nu \otimes \delta)$. For fixed ρ, $T(\nu, \rho)$ is of the form $T(\nu, \rho) = \mathrm{Tr}M_{\boldsymbol{\mu}_{2n}\rho}\nu$. The matrix $M_{\boldsymbol{\mu}_{2n}\rho}$ has trace equal to

$$\begin{aligned}
\mathrm{Tr}M_{\boldsymbol{\mu}_{2n}\rho} &= T(\rho, I) \\
&= \mathrm{Tr}\boldsymbol{\mu}_{2n}(\rho \otimes I) \\
&= \mathrm{Tr}\left((\mathrm{Tr}_{\mathcal{Q}_B}\boldsymbol{\mu}_{2n})\rho\right) \\
&\stackrel{*}{=} \mathrm{Tr}\boldsymbol{\mu}_n\rho \\
&\stackrel{*}{=} 2^{-\mathbf{Hg}(\rho)},
\end{aligned}$$

using Theorem 9, which states $\mathrm{Tr}_Y\boldsymbol{\mu}_{XY} \stackrel{*}{=} \boldsymbol{\mu}_X$. By the definition of M, since $\boldsymbol{\mu}_{2n}$ and ρ are positive semi-definite, it must be that $M_{\boldsymbol{\mu}_{2n}\rho}$ is positive semi-definite. Since the trace of $M_{\boldsymbol{\mu}_{2n}\rho}$ is $\stackrel{*}{=} 2^{-\mathbf{Hg}(\rho)}$, it must be that up to a multiplicative constant, $2^{\mathbf{Hg}(\rho)}M_{\boldsymbol{\mu}_{2n}\rho}$ is a semi-density matrix.

Since $\boldsymbol{\mu}$ is lower computable and ρ is elementary, by the definition of M, $2^{\mathbf{Hg}(\rho)}M_{\boldsymbol{\mu}_{2n}\rho}$ is lower computable relative to the string $\langle \rho, \mathbf{Hg}(\rho)\rangle$. Therefore we have that $2^{\mathbf{Hg}(\rho)}M_{\boldsymbol{\mu}_{2n}\rho} \stackrel{*}{<} \boldsymbol{\mu}_{\langle \rho, \mathbf{Hg}(\rho)\rangle}$. So we have that $-\log \mathrm{Tr}2^{\mathbf{Hg}(\rho)}M_{\boldsymbol{\mu}_{2n}\rho}\sigma = -\mathbf{Hg}(\rho) - \log T(\sigma, \rho) =^+ \mathbf{Hg}(\sigma \otimes \rho) - \mathbf{Hg}(\rho) >^+ -\log \boldsymbol{\mu}_{(\rho, \mathbf{Hg}(\rho))}\sigma =^+ \mathbf{Hg}(\sigma|\langle \rho, \mathbf{Hg}(\rho)\rangle)$. □

Exercise 5 *Prove that relativized to a basis* $|1\rangle, |2\rangle, \ldots$ *over elementary orthogonal states,* $\mathbf{Hg}(\sigma \otimes |i\rangle\langle i|) =^+ \mathbf{Hg}(|i\rangle) + \mathbf{Hg}(\sigma||i\rangle, \mathbf{Hg}(|i\rangle))$.

5.5 Subadditivity, Strong Subadditivity, Strong Superadditivity

Theorem 13 ([GÓ1]) $\mathbf{Hg}(\sigma)$ *is subadditive, with* $\mathbf{Hg}(\sigma \otimes \rho) <^+ \mathbf{Hg}(\sigma) + \mathbf{Hg}(\rho)$.

Proof.

$$\begin{aligned}
&2^{-\mathbf{Hg}(\sigma)-\mathbf{Hg}(\rho)} \\
=&\left(\mathrm{Tr}\boldsymbol{\mu}_X\sigma\right)\left(\mathrm{Tr}\boldsymbol{\mu}_Y\rho\right) \\
=&\mathrm{Tr}(\sigma \otimes \rho)(\boldsymbol{\mu}_X \otimes \boldsymbol{\mu}_Y) \\
\stackrel{*}{>}&\mathrm{Tr}(\sigma \otimes \rho)(\boldsymbol{\mu}_{XY}) \\
\stackrel{*}{=}&2^{-\mathbf{Hg}(\sigma\otimes\rho)}.
\end{aligned}$$

□

A function $\mathbf{L}$ from quantum mixed states to whole numbers is strongly subadditive if there exists a constant $c \in \mathbb{N}$ such that for all mixed states ρ_{123}, $\mathbf{L}(\rho_{123}) + \mathbf{L}(\rho_2) < \mathbf{L}(\rho_{12}) + \mathbf{L}(\rho_{23}) + c$. Similarly $\mathbf{L}$ is strongly superadditive if there exists a constant $c \in \mathbb{N}$ such that for all mixed states ρ_{123}, $\mathbf{L}(\rho_{12}) + \mathbf{L}(\rho_{23}) < \mathbf{L}(\rho_{123}) + \mathbf{L}(\rho_2) + c$. In [GÓ1], it was asked if $\mathbf{Hg}$ was strongly superadditive. In this section we provide a negative answer, and it is also shown that $\mathbf{Hg}$ is not strongly superadditive or subadditive.

Theorem 14 $\mathbf{Hg}$ *is not strongly subadditive.*

Proof. We fix the number of qubits n, and for $i \in [1..2^n]$, $|i\rangle$ is the ith basis state of the n qubit space. Let $|\psi\rangle = \sum_{i=1}^{2^n} 2^{-n/2} |i\rangle |i\rangle$. The pure state $|\psi\rangle$ is elementary, with $\mathbf{K}(|\psi\rangle |2^{2n}) =^+ 0$. We define the the $3n$ qubit mixed state $\rho_{123} = .5 |\psi\rangle \langle\psi| \otimes |1\rangle \langle 1| + .5 |1\rangle \langle 1| \otimes |\psi\rangle \langle\psi|$. $\rho_{12} = .5 |\psi\rangle \langle\psi| + .5 |1\rangle \langle 1| \otimes 2^{-n} I$. $\rho_{23} = .5 * 2^{-n} I \otimes |1\rangle \langle 1| + .5 |\psi\rangle \langle\psi|$. $\rho_2 = 2^{-n} I$. $\mathbf{Hg}(\rho_{12}) =^+ -\log \mathrm{Tr} \boldsymbol{\mu}^{2n} \rho_{12} <^+ -\log \mathrm{Tr} \boldsymbol{\mu}^{2n} |\psi\rangle \langle\psi| <^+ -\log \mathbf{m}(|\psi\rangle |2^{2n})| \langle\psi|\psi\rangle |^2 <^+ 0$. Similarly, $\mathbf{Hg}(\rho_{23}) =^+ 0$. $\mathbf{Hg}(\rho_2) =^+ n$. So $\mathbf{Hg}(\rho_{123}) + \mathbf{Hg}(\rho_2) >^+ n$ and $\mathbf{Hg}(\rho_{12}) + \mathbf{Hg}(\rho_{23}) =^+ 0$, proving that $\mathbf{Hg}$ is not strongly subadditive. □

Theorem 15 $\mathbf{Hg}$ *is not strongly superadditive.*

Proof. We fix the number of qubits n, and for $i \in [1..2^n]$, $|i\rangle$ is the ith basis state of the n qubit space. Let $|\phi\rangle = \sum_{i=1}^{2^n} 2^{-n/2} |i\rangle |i\rangle |i\rangle$, with $\mathbf{K}(|\phi\rangle |2^{3n}) = 0$. Let $\sigma_{123} = |\phi\rangle \langle\phi|$. $\sigma_{12} = \sigma_{23} = \sum_{i=1}^{2^n} 2^{-n} |i\rangle \langle i| \otimes |i\rangle \langle i|$. $\mathbf{Hg}(\sigma_{123}) =^+ -\log \mathrm{Tr} \sigma_{123} \boldsymbol{\mu}^{3n} <^+ -\log \mathrm{Tr}\, \mathbf{m}(|\phi\rangle |2^{3n})| \langle\phi|\phi\rangle |^2 <^+ 0$. Let D be a unitary transform where $D |i\rangle |i\rangle = |i\rangle |1\rangle$ and $\mathbf{K}(D|2^{2n}) =^+ 0$. So $\mathbf{Hg}(\sigma_{12}) =^+ \mathbf{Hg}(D\sigma_{12}D^*) =^+ \mathbf{Hg}(2^{-n} I \otimes |1\rangle \langle 1|) =^+ n - \log \mathrm{Tr}(I \otimes |1\rangle \langle 1|) \boldsymbol{\mu}^{2n}$. By Theorem 8 and properties of partial trace, $\mathbf{Hg}(2^{-n} I \otimes |1\rangle \langle 1|) =^+ n - \log \mathrm{Tr} |1\rangle \langle 1| \boldsymbol{\mu}^n =^+ n$. So $\mathbf{Hg}(\sigma_{12}) = \mathbf{Hg}(\sigma_{23}) =^+ n$. So $\mathbf{Hg}(\sigma_{123}) + \mathbf{Hg}(\sigma_2) <^+ n$, and $\mathbf{Hg}(\sigma_{12}) + \mathbf{Hg}(\sigma_{23}) >^+ 2n$, proving that $\mathbf{Hg}$ is not strongly superadditive. □

5.6 Vitányi Complexity and Gács Complexity

A natural question to ask is the relationship between $\mathbf{Hg}$ and $\mathbf{Hv}$. As this section will show, Vitányi complexity and Gács complexity are virtually identical, except for a small subset of exotic states. By definition $\mathbf{Hg}(|\psi\rangle) <^+ \mathbf{Hv}(|\psi\rangle)$. In fact, as shown in the following theorem, Vitányi complexity is bounded with respect to Gács complexity.

Theorem 16 ([GÓ1]) $\mathbf{Hg}(|\psi\rangle) <^+ \mathbf{Hv}(|\psi\rangle) <^{\log} 4\mathbf{Hg}(|\psi\rangle)$.

Proof. For semi-density matrix A with eigenvectors $\{|a_i\rangle\}$ and decreasing eigenvectors $\{a_i\}$ with $\langle\psi| A |\psi\rangle \geq 2^{-k}$ and $|\psi\rangle = \sum c_i |a_i\rangle$, let A_m be a projector onto the m largest eigenvectors. Let m be the first i where $a_i \leq 2^{-k-1}$. Since $\sum a_i \leq 1$, we have $m \leq 2^{k+1}$. Since

$$\sum_{i \geq m} a_i |c_i|^2 < 2^{-k-1} \sum_i |c_i|^2 = 2^{-k-1},$$

we have

$$\langle\psi| A_m |\psi\rangle \geq \sum_{i<m} |c_i|^2 \geq \sum_{i<m} a_i |c_i|^2 \geq 2^{-k} - \sum_{i \geq m} a_i |c_i|^2 > 2^{-k-1}.$$

Thus there is some $i \leq m$ such that $|\langle\psi|a_i\rangle|^2 \geq 2^{-2k-2}$. Let $\nu = \mathrm{Tr}\boldsymbol{\mu}$ and $\nu_k \in \mathbb{Q}$ be a rational created from the first k digits of ν. Let $\hat{\boldsymbol{\mu}}$ be a lower approximation of $\boldsymbol{\mu}$, with trace greater than ν_k. So $\mathbf{K}(\hat{\boldsymbol{\mu}}) <^{\log} k$. Thus if $\langle\psi| \boldsymbol{\mu} |\psi\rangle \geq 2^{-k}$, then $\langle\psi| \hat{\boldsymbol{\mu}} |\psi\rangle \geq 2^{-k-1}$. Thus there is an eigenvector $|u\rangle$ of $\hat{\boldsymbol{\mu}}$ of complexity $\mathbf{K}(|u\rangle |n) <^{\log} 2k$ and $|\langle\psi|u\rangle|^2 \stackrel{*}{>} 2^{-2k}$, so

$$\mathbf{Hv}(|\psi\rangle) \leq \mathbf{K}(|u\rangle |n) - \log |\langle\psi|u\rangle|^2 <^{\log} 4k <^{\log} 4\mathbf{Hg}(|\psi\rangle).$$

□

Definition 5 *We now describe an infinite encoding scheme for an arbitrary (not necessarily elementary) quantum pure state $|\psi\rangle$. This scheme is defined as an injection between the set of pure states and $\{0,1\}^\infty$. We define $\langle\langle|\psi\rangle\rangle$ to be an ordered list of the encoded tuples $\langle\langle|\theta\rangle\rangle, q, [|\langle\psi|\theta\rangle|^2 \geq q]\rangle$, over all elementary states $|\theta\rangle$ and rational distances $q \in \mathbb{Q}_{>0}$.*

The following theorem states that only exotic pure states will have a Vitányi complexity much greater than Gács complexity. States are exotic if they have high mutual information, $\mathbf{I}$ (Definition 1), with the halting sequence $\mathcal{H} \in \{0,1\}^\infty$.

Lemma 5 *For pure quantum state $|\psi\rangle$,*
$\min_{|\phi\rangle} \mathbf{K}(|\phi\rangle) - \log|\langle\psi|\phi\rangle|^2 <^{\log} -\log\sum_{|\phi\rangle}\mathbf{m}(|\phi\rangle)|\langle\psi|\phi\rangle|^2 + \mathbf{I}(\langle|\psi\rangle\rangle : \mathcal{H})$.

Proof. Let $\mathcal{D}$ be a finite set of elementary pure states, computable from $\langle|\psi\rangle\rangle$ and the value $g = \lceil -\log\sum_{|\phi\rangle}\mathbf{m}(|\phi\rangle)|\langle\psi|\phi\rangle|^2\rceil$ such that

$$-\log\sum_{|\theta\rangle\in\mathcal{D}}\mathbf{m}(|\theta\rangle)|\langle\psi|\theta\rangle|^2 \leq g+1.$$

It is computable because there exists an algorithm that can find $\mathcal{D}$ by the following method. The algorithm enumerates all elementary states $|\theta\rangle$. This algorithm approximates the algorithmic probabilities $\mathbf{m}(|\theta\rangle)$ (from below) with $\widehat{\mathbf{m}}(|\theta\rangle)$. This algorithm uses $\langle|\psi\rangle\rangle$ to approximate $|\langle\theta|\psi\rangle|^2$ from below with $\widehat{|\langle\theta|\psi\rangle|}^2$. This algorithm stops when it finds a finite set $\mathcal{D}$ such that

$$-\log\sum_{|\theta\rangle\in\mathcal{D}}\widehat{\mathbf{m}}(|\theta\rangle)\widehat{|\langle\theta|\psi\rangle|}^2 \leq g+1.$$

Thus we have that $\mathbf{K}(\mathcal{D}|g,\langle|\psi\rangle\rangle) = O(1)$. Let $f:\mathcal{D}\to\mathbb{W}$ be a elementary function such that $|-\log|\langle\psi|\theta\rangle|^2 - f(|\theta\rangle)| \leq 1$. One such f is computable relative to $\langle|\psi\rangle\rangle$, and g. Firstly this is because D is computable from $\langle|\psi\rangle\rangle$ and g. The individual values of f are computable from $\langle|\psi\rangle\rangle$, since $|\langle\psi|\theta\rangle|^2$ can be computed to any degree of accuracy. So $\mathbf{K}(f|g,\langle|\psi\rangle\rangle) = O(1)$ and $-\log\sum_{|\theta\rangle\in\mathcal{D}}\mathbf{m}(|\theta\rangle)2^{-f(|\theta\rangle)} \leq g+2$. One then has that

$$\begin{aligned}
\min_{|\phi\rangle}\mathbf{K}(|\phi\rangle) - \log|\langle\psi|\phi\rangle|^2 &<^+ \min_{\theta\in\mathcal{D}}\mathbf{K}(|\theta\rangle) + f(|\theta\rangle) \\
&<^{\log} -\log\sum_{|\theta\rangle\in\mathcal{D}}\mathbf{m}(|\theta\rangle)2^{-f(|\theta\rangle)} + \mathbf{I}(\langle f\rangle;\mathcal{H}). && (5.1)\\
&<^{\log} g + \mathbf{I}(\langle f\rangle;\mathcal{H}) && (5.2)\\
&<^{\log} g + \mathbf{I}(\langle|\psi\rangle\rangle:\mathcal{H}) + \mathbf{K}(\langle f\rangle|\langle|\psi\rangle\rangle) && (5.3)\\
&<^{\log} g + \mathbf{I}(\langle|\psi\rangle\rangle:\mathcal{H}) + \mathbf{K}(g) \\
&<^{\log} -\log\sum_{|\phi\rangle}\mathbf{m}(|\phi\rangle)|\langle\psi|\phi\rangle|^2 + \mathbf{I}(\langle|\psi\rangle\rangle:\mathcal{H}).
\end{aligned}$$

Inequality 6.2 is due to Theorem 114. Inequality 6.3 is due to the definition of f and $\mathcal{D}$. Inequality 6.4 is due to the definition of $\mathbf{I}$, where $\mathbf{I}(x;\mathcal{H}) <^+ \mathbf{I}(\alpha:\mathcal{H}) + \mathbf{K}(x|\alpha)$. □

Theorem 17 $\mathbf{Hg}(|\psi\rangle) <^+ \mathbf{Hv}(|\psi\rangle) <^{\log} \mathbf{Hg}(|\psi\rangle) + \mathbf{I}(|\psi\rangle:\mathcal{H}|n)$.

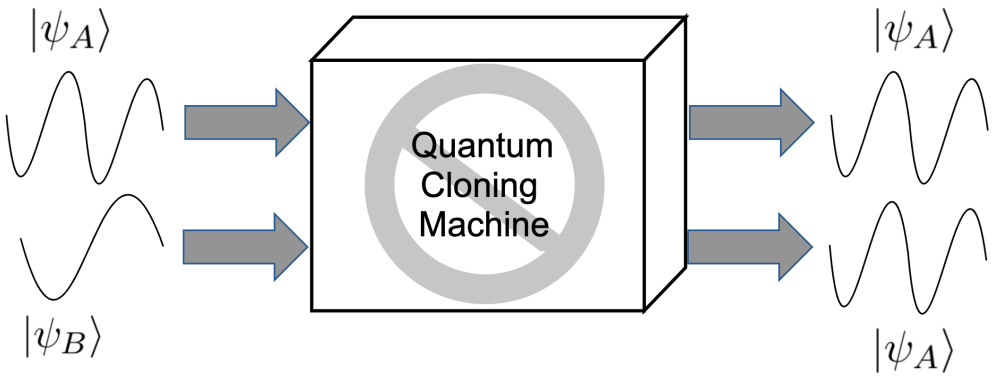


Figure 5.1: The no cloning theorem states that there is no method for cloning an arbitrary quantum state. However, approximate quantum cloning is possible.

Proof. This follows directly from Lemma 5, relativized to n, and the fact that $\mathbf{Hg}(|\psi\rangle) =^+ -\log \mathrm{Tr}\boldsymbol{\mu}\,|\psi\rangle\langle\psi| =^+ -\log \sum_{|\phi\rangle} \mathbf{m}(|\phi\rangle\,|n\rangle)|\,\langle\phi|\psi\rangle\,|^2$. □

5.7 No Cloning Theorem

In classical algorithmic information theory, one can easily reproduce a string x, with

$$\mathbf{K}(x) =^+ \mathbf{K}(x,x).$$

However the situation is much different in the quantum case. The no cloning theorem [WZ82]. is as follows. Say Alice has arbitrary state $|\psi_A\rangle$ in Hilbert space $\mathcal{A}$ and Bob has base state $|\psi_B\rangle$ in Hilbert space $\mathcal{B}$, where spaces $\mathcal{A}$ and $\mathcal{B}$ are identical. Alice wants to perform to perform the two operations to clone her state to produce $|\psi_A\rangle \otimes |\psi_B\rangle \to |\psi_A\rangle \otimes |\psi_A\rangle$:

1. An observation that will cause a collapse to an eigenstate.
2. A time independent Hamiltonian of the combined system .

The no-cloning theorem (Figure 5.1) says Alice will not be successful, she cannot clone all states. In addition, there exists several generalizations to the no-cloning theorem, showing that imperfect clones can be made. In [BH96], a universal cloning machine was introduced that can clone an arbitrary state with the fidelity of 5/6.

The following theorem generalizes this no-go result, by showing there exist tensor products $|\psi\rangle^m$ that has significantly more algorithmic quantum complexity measure than $|\psi\rangle\,|0\rangle^{m-1}$. This result can be proved for all three complexities: **Hg**, **Hv** and **Hbvl**. The following theorem has a new proof to the **Hg** result.

Theorem 18 ([GÓ1]) $\log \binom{m+2^n-1}{m} <^+ \max_{|\psi\rangle} \mathbf{Hg}(|\psi\rangle^{\otimes m}) <^+ \mathbf{K}(m) + \log \binom{m+2^n-1}{m}$.

Proof. Let $N = 2^n$. Let $\mathcal{H}_N$ be an N dimensional Hilbert space and let $\mathcal{H}_{N^m}$ be an m-fold tensor space of $\mathcal{H}_N$. Let $\mathrm{Sym}(\mathcal{H}_N^m)$ be the subspace of $\mathcal{H}_N^m$ consisting of all pure states of the form $|\psi\rangle^{\otimes m}$. The subspace $\mathrm{Sym}(\mathcal{H}_N^m)$ is spanned by M basis vectors, where M is the number of multisets of size m from the set $\{1,\ldots,N\}$. This is because for each such multiset $S = \{i_1,\ldots,i_m\}$, one can construct a basis vector $|\psi_S\rangle$ that is the normalized superposition of all basis vectors of $\mathrm{Sym}(\mathcal{H}_N^m)$

that are permutations of S. If $S' \neq S$, then $|\psi_S\rangle$ is orthogonal to $|\psi_{S'}\rangle$. Thus the dimension of $\mathrm{Sym}(\mathcal{H}_N^m)$ M, is $\binom{m+N-1}{m}$ because choosing a multiset is the same as splitting an interval of size m into N intervals. For the upper bounds, let P_S be the projector onto $\mathrm{Sym}(\mathcal{H}_N^m)$. If $|\psi\rangle \in \mathrm{Sym}(\mathcal{H}_N^m)$, then $\langle\psi| P_S |\psi\rangle = 1$ so

$$\mathbf{Hg}(|\psi\rangle) <^+ \mathbf{K}(P_S/M|n \times m) - \log \langle\psi| \frac{1}{M} P_S |\psi\rangle <^+ \mathbf{K}(m) + \log \binom{m+N-1}{m}.$$

For the lower bound, let $c = \max_{|\psi\rangle \in \mathcal{Q}_n} \mathbf{Hg}(|\psi\rangle^{\otimes m})$. We have for all $|\psi\rangle \in \mathcal{Q}_n$,

$$\mathrm{Tr}\boldsymbol{\mu} |\psi\rangle^m \langle\psi|^m \stackrel{*}{>} 2^{-c}. \tag{5.4}$$

Let Λ be the uniform distribution on the unit sphere of $\mathcal{Q}_n$. And let

$$\rho = \int |\psi\rangle^m \langle\psi|^m \, d\Lambda.$$

$\mathrm{Tr}\rho = \int \mathrm{Tr} |\psi\rangle^m \langle\psi|^m \, d\Lambda = \int d\Lambda = 1$. Furthermore for $|\phi\rangle^m, |\nu\rangle^m \in \mathrm{Sym}(\mathcal{H}_N^m)$, with unitary transform U such that $U^m |\phi\rangle^m = |\nu\rangle^m$, we have

$$\langle\nu|^n \rho |\nu\rangle^n = \int \langle\phi|^m (U^{*m} |\psi\rangle^m \langle\psi|^m U^m) |\phi\rangle^m \, d\Lambda = \int \langle\phi^m|\psi^m\rangle \langle\psi^m|\phi^m\rangle^m \, d\Lambda = \langle\phi|^n \rho |\phi\rangle^n .$$

For any pure state $|\psi\rangle \in \mathcal{H}_N^m$, such that $\langle\psi| P_S |\psi\rangle = 0$, then $\langle\psi| \rho |\psi\rangle = 0$. Thus $\rho = P_S/M$. Integrating Equation 5.4, by $d\Lambda$ results in

$$2^{-c} \stackrel{*}{<} \mathrm{Tr}\boldsymbol{\mu}\rho \stackrel{*}{=} \mathrm{Tr}\boldsymbol{\mu} P_S/M \stackrel{*}{=} \binom{m+N-1}{m}^{-1}$$
$$c >^+ \log \binom{m+N-1}{m}.$$

□

Corollary 2 ([Vit01]) $\log \binom{m+2^n-1}{m} <^+ \max_{|\psi\rangle} \mathbf{Hv}(|\psi\rangle^{\otimes m}) <^{\log} 4 \left(\mathbf{K}(m) + \log \binom{m+2^n-1}{m} \right)$.

Proof. This follows from Theorems 18 and 16. □

5.8 Quantum EL Theorem

In this paper we prove a Quantum EL Theorem. In AIT, the EL Theorem [Lev16, Eps19d] states that sets of strings that contain no simple member will have high mutual information with the halting sequence. It is also stated in Corollary 47.

$$\textit{For finite set } D \subset \{0,1\}^*, \min_{x \in D} \mathbf{K}(x) <^{\log} -\log \sum_{x \in D} \mathbf{m}(x) + \mathbf{I}(D; \mathcal{H}).$$

It has many applications, including that all sampling methods produce outliers [Eps21b]. The Quantim EL Theorem states that elementary projections P of large rank must have simple quantum pure states in their images, otherwise they have high $\mathbf{I}(P : \mathcal{H})$. By the Independence Postulate (see Chapter 26), constructs with non-negligible information with the halting sequence cannot be found

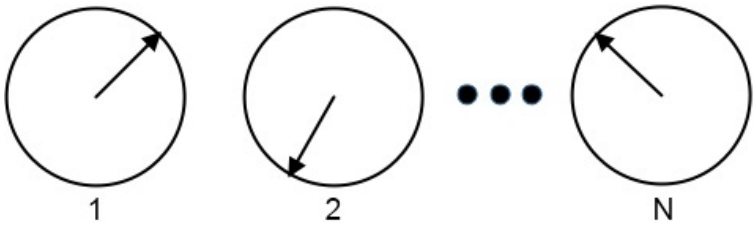


Figure 5.2: The proof of the Quantum EL Theorem relies on sampling a large number (N) of pure n qubit states according to the uniform distribution. The proof has a collection of elementary projections R under consideration. For most R, there will be a randomly selected state $|\psi\rangle$ that will have a high $\langle\psi| R |\psi\rangle$.

in the physical world and are thus exotic. The Quantum EL Theorem has the following consequence.

Claim. *As the von Neumann entropy associated with the quantum source increases, the lossless quantum coding projectors have larger rank and thus must have simpler (in the algorithmic quantum complexity sense) pure states in their images.*

Theorem 19 (Quantum EL Theorem [Eps23d]) *Fix an n qubit Hilbert space. Let P be a elementary projection of rank $> 2^m$. Then, relativized to (n, m), $\min_{|\phi\rangle \in \text{Image}(P)} \mathbf{Hv}(|\phi\rangle) <^{\log} 3(n - m) + \mathbf{I}(\langle P\rangle; \mathcal{H})$.*

Proof. We assume P has rank 2^m. Let Q be the elementary probability measure that realized the stochasticity, $\mathbf{Ks}(P)$, of an encoding of P. We can assume that every string in the support of Q encodes an elementary projection of rank 2^m. We sample N independent pure states according to the uniform distribution Λ on the n qubit space, as shown in Figure 5.2. N is to be defined later. For each pure state $|\psi_i\rangle$ and projection R in the support of Q, the expected value of $\langle\psi_i| R |\psi_i\rangle$ is

$$\int \langle\psi_i| R |\psi_i\rangle \, d\Lambda = \text{Tr}R \int |\psi_i\rangle \langle\psi_i| \, d\Lambda = 2^{-n}\text{Tr}RI = 2^{m-n}.$$

Let random variable $X_R = \frac{1}{N}\sum_{i=1}^{N} \langle\psi_i| R |\psi_i\rangle$ be the average projection size of the random pure states onto the projection R. Since $\langle\psi_i| R |\psi_i\rangle \in [0, 1]$ with expectation 2^{m-n}, by Hoeffding's inequality,

$$\Pr(X_R \leq 2^{m-n-1}) < \exp\left[-N2^{-2(m-n)-1}\right]$$

Let $d = \mathbf{d}(P|Q)$. Thus if we set $N = d2^{2(m-n)+1}$, we can find N elementary n qubit states such that $Q(\{R : X_R \leq 2^{m-n-1}\}) \leq \exp(-d)$, where X_R is now a fixed value and not a random variable. Thus $X_P > 2^{m-n-1}$ otherwise one can create a Q-expectation test, t, such that $t(R) = \exp d$. This is a contradiction because

$$1.44d <^{+} \log(P) <^{+} \mathbf{d}(P|Q, d) <^{+} d + \mathbf{K}(d),$$

for large enough d which we can assume without loss of generality. Thus there exists i such that $\langle\psi_i| P |\psi_i\rangle \geq 2^{m-n-1}$. Thus $|\phi\rangle = P |\psi_i\rangle / \sqrt{\langle\psi_i| P |\psi_i\rangle}$ is in the image of P and $|\langle\psi_i|\phi\rangle|^2 = \langle\psi_i| P |\psi_i\rangle \geq 2^{m-n-1}$. The elementary state $|\psi_i\rangle$ has classical Kolmogorov complexity $\mathbf{K}(|\psi_i\rangle) <^{\log} \log N +$

$\mathbf{K}(Q,d) <^{\log} 2(m-n) + \mathbf{Ks}(P)$. Thus by Lemma 3,

$$\begin{aligned}
&\min\{\mathbf{Hv}(|\psi\rangle) : |\psi\rangle \in \text{Image}(P)\} \\
&\leq \mathbf{Hv}(|\phi\rangle) \\
&<^{\log} \mathbf{K}(|\psi_i\rangle) + |\langle\psi_i|\phi\rangle|^2 \\
&<^{\log} 3(n-m) + \mathbf{Ks}(P) \\
&<^{\log} 3(n-m) + \mathbf{I}(P;\mathcal{H}).
\end{aligned}$$

□

5.8.1 Computable Projections

Theorem 19 is in terms of elementary described projecctions and can be generalized to arbitrarily computable projections. For a matrix M, let $\|M\| = \max_{i,j} |M_{i,j}|$ be the max norm. A program $p \in \{0,1\}^*$ computes a projection P of rank ℓ if it outputs a series of rank ℓ projections $\{P_i\}_{i=1}^{\infty}$ such that $\|P - P_i\| \leq 2^{-i}$. For computable projection operator P, $\mathbf{I}(P;\mathcal{H}) = \min\{\mathbf{K}(p) - \mathbf{K}(p|\mathcal{H}) : p \text{ is a program that computes } P\}$.

Corollary 3 ([Eps23d]) *Fix an n qubit Hilbert space. Let P be a computable projection of rank $> 2^m$. Then, relativized to (n,m), $\min_{|\phi\rangle \in \text{Image}(P)} \mathbf{Hv}(|\phi\rangle) <^{\log} 3(n-m) + \mathbf{I}(P;\mathcal{H})$.*

Proof. Let p be a program that computes P. There is a simply defined algorithm A, that when given p, outputs P_n such that $\min_{|\psi\rangle \in \text{Image}(P)} \mathbf{Hv}(|\psi\rangle) =^{+} \min_{|\psi\rangle \in \text{Image}(P_n)} \mathbf{Hv}(|\psi\rangle)$. Thus by Lemma 1, one gets that $\mathbf{I}(P_n;\mathcal{H}) <^{+} \mathbf{I}(P;\mathcal{H})$. The corollary follows from Theorem 19. □

5.8.2 Quantum Data Compression

The Quantum EL Theorem can be used to address open issues in Quantum Information Theory. In [Gó1] the following remark was made.

Remark 1 ([Gó1]) *Maybe the study of the problem for quantum description complexity helps with the understanding of the problem for von Neumann entropy, and its relation to coding tasks of quantum information theory.*

A quantum source consists of a set of pure quantum states $\{|\psi_i\rangle\}$ and their corresponding probabilities $\{p_i\}$, where $\sum_i p_i = 1$. The pure states are not necessarily orthogonal. The sender, Alice wants to send the pure states to the receiver, Bob. Let $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ be the density matrix associated with the quantum source. Let $S(\rho)$ be the von Neumann entropy of ρ. By Schumacher compression, [Sch95], in the limit of $n \to \infty$, Alice can compress n qubits into $S(\rho)n$ qubits and send these qubits to Bob with fidelity approaching 1. For example, if the message consists of n photon polarization states, we can compress the inital qubits to $nS(\rho)$ photons. Alice cannot compress the initial qubits to $n(S(\rho) - \delta)$ qubits, as the fidelity will approach 0. The qubits are compressed by projecting the message onto a typical subspace of rank $nS(\rho)$ using a projector P. The projection occurs by using a quantum measurement consisting of P and a second projector $(I - P)$, which projects onto a garbage state.

> *The results of this paper says that as $S(\rho)$ increases, there must be simple states in the range of P. There is no way to communicate a quantum source with large enough $S(\rho)$ without using simple quantum states.*

Chapter 6

BvL Complexity

Kolmogorov complexity measures the smallest program to a universal Turing machine that produces a string. Thus it is natural to adapt this notion to defining the complexity of a pure or mixed quantum state ρ to be the shortest program to a universal quantum Turing machine that approximates or produces ρ. This definition was introduced in [BvL01] and we call it BvL complexity. Whereas Gács complexity and Vitányi can be thought of as scores of the algorithmic entropy of a state, BvL complexity enjoys a direct interpretation of the amount of resources in quantum mechanics needed to approximate or produce a state.

All quantum Turing machines used in this manuscript are the well formed QTMs defined in [BV93]. Well formed QTM preserve length and their time evolution is unitary. In this manuscript, BvL complexity is defined with respect to a universal quantum Turing machine introduced in [Mul08]. This is different than the work in [BvL01], which uses the universal quantum machine from [BV93]. Another notion of algorithmic quantum complexity can be defined with respect to *deterministic control* quantum Turing machines, as seen in [Lem]. This seems like a natural definiton, as virtually all quantum algorithms known (such as Shor's factoring algorithm) have deterministic control. This notion enjoys the chain rule relation, but the complexity is not covered in this manuscript.

The input, output and auxiliary tapes of M consists of symbols of the type $\Sigma = \{0, 1, \#\}$. The input is an ensemble $\{p_i\}$ of pure states $|\psi_i\rangle$ of the same length n, where $p_i \geq 0$, $\sum_i p_i = 1$, and $p_i \in Q_{\geq 0}$. Each pure state $|\psi_i\rangle$ is a complex linear superposition over all inputs of length n. Thus the input can be seen as an ensemble of states $|\psi_i \# \# \# \ldots\rangle$. This ensemble can be represented as a mixed state ρ of n qubits. The auxiliary tape can contain quantum or classical information. The quantum transition function is

$$\delta : Q \times \Sigma^3 \to \mathbb{C}^{Q \times \Sigma^3 \times \{L,R\}^3}.$$

Note that each complex number must be computable. Q is the set of states, Σ is the alphabets on the auxiliary, output, and input tapes, and $\{L, R\}^3$ is the action taken by the three heads. The evolution of M is a computable unitary matrix u_M.

Definition 6 (Indeterminate Length Quantum States) *The separable Hilbert space $\mathcal{Q} = \bigoplus_{n \in \mathbb{W}} \mathcal{Q}_n$ is the space of indeterminate length quantum states. An example indeterminate length quantum state is*

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11011\rangle).$$

There is a start state $|s_C\rangle$ and a final state $|f_C\rangle$. If there exists a $t \in \mathbb{N}$, where during the operation of M input ρ, the control state $M_C^{t'}(\rho)$ is orthogonal to the final state $|f_C\rangle$ for all $t' < t$,

with $\langle f_C|\, M_C^{t'}(\rho)\, |f_C\rangle = 0$, and $\langle f_C|\, M_C^{t}(\rho)\, |f_C\rangle = 1$, then $M(\rho)$ is defined to be the qubit mixed state σ corresponding to the ensemble of pure states determined by ensemble of pure states over the contents of the output tapes at halting time. If one such pure state of the output tape is $|\psi\rangle = \sum_{i=1}^{N} \alpha_i\, |s_i \#\#\#\#\# \ldots\rangle$, where each $\|s_i\|$ can be different, then the resultant output pure state is $|\widetilde{\psi}\rangle = \sum_{i=1}^{N} \alpha_i\, |s_i\rangle$. In this case, $M(\sigma)$ is defined. If $M(\sigma)$ is defined and runs in time t, then for times $t' > t$, the output tape is not changed. Otherwise, if the control state evolution is not defined as above, $M(\sigma)$ is undefined. Thus the output can be a superposition of pure states of different lengths, indeterminate length quantum states. Thus QTMs M can be thought of as partial functions of the following form.

$$M : \bigcup_n \mathcal{Q}_n \to \mathcal{Q}.$$

Thus we only consider *fixed-length* inputs to QTMs M. This consists of elements of $\mathcal{Q}$ that are superpositions of basis quantum states $|e_i\rangle$ of the same length.

One might argue that this definition with regard to the halting state is too restrictive, but as shown [Mue07], for every input σ to a QTM that almost halts within a certain computable level of precision, there is another state σ' such that $\|\sigma'\| <^{+} \|\sigma\|$ that makes the universal QTM $\mathfrak{U}$ halt perfectly.

Quantum machines are not expected to produce the target states exactly, only an approximation is required. To measure the closeness of states, the *trace distance* function is used.

Definition 7 (Trace Distance and Fidelity of Quantum States) $D(\sigma, \rho) = \frac{1}{2}\|\sigma - \rho\|_1$, *where* $\|A\|_1 = \mathrm{Tr}\sqrt{A^* A}$. *The trace distance obeys the triangle inequality. Fidelity is* $F(\sigma, \rho) = \left(\mathrm{Tr}\sqrt{\sqrt{\sigma}\rho\sqrt{\rho}}\right)^2$, *with* $F(|\psi\rangle, \sigma) = \langle\psi|\, \sigma\, |\psi\rangle$ *and* $1 - D(\rho, |\psi\rangle) < F(\rho, |\psi\rangle)$.

Theorem 20 ([Mul08]) *There is quantum Turing machine* $\mathfrak{U}$ *such that for every QTM* M *and mixed state* σ *for which* $M(\sigma)$ *is defined, there is mixed state* σ' *such that*

$$D\left(\mathfrak{U}(\sigma'), M(\sigma)\right) < \delta,$$

for every $\delta \in \mathbb{Q}_{>0}$ *where* $\|\sigma'\| <^{+} \|\sigma\| + \mathbf{K}(M, \delta$.

One can define the complexity of a state σ with respect to an arbitrary quantum Turing machine.

Definition 8 *The BvL Complexity of mixed state* ρ *with respect to QTM* M *and trace distance* ϵ *is*

$$\mathbf{Hbvl}_M^{\epsilon}(\rho) = \min_{\sigma}\{\|\sigma\| : D(M(\sigma), \rho) < \epsilon\}.$$

The BvL Complexity of mixed state ρ *with respect to QTM* M *is*

$$\mathbf{Hbvl}_M(\rho) = \min_{\sigma}\left\{\|\sigma\| : \forall_k, D\left(M\left(\sigma, k\right), \rho\right) < \frac{1}{k}\right\}.$$

Due to Theorem 20 and the fact that the trace distance D follows the triangle inequality, using the universal quantum Turing machine $\mathfrak{U}$, one can define the BvL complexity of a quantum state. This differs from the original definition in [BvL01] where the program must achieve any degree of precision.

Theorem 21 ([Mul08]) *For* $\delta < \epsilon \in \mathbb{Q}_{>0}$, *universal QTM* $\mathfrak{U}$, *for every QTM* M,

- $\mathbf{Hbvl}^{\epsilon}_{\mathfrak{U}}(\sigma) < \mathbf{Hbvl}^{\delta}_{M}(\sigma) + \mathbf{K}(\epsilon - \delta, M)$.
- $\mathbf{Hbvl}_{\mathfrak{U}}(\sigma) < \mathbf{Hbvl}_{M}(\sigma) + \mathbf{K}(M)$.

Definition 9 (BvL Complexity)

- $\mathbf{Hbvl}^{\epsilon}(\sigma) = \mathbf{Hbvl}^{\epsilon}_{\mathfrak{U}}(\sigma)$.
- $\mathbf{Hbvl}(\sigma) = \mathbf{Hbvl}_{\mathfrak{U}}(\sigma)$.

6.1 An Elementary Approximation of $\mathfrak{U}$

Remark 2 *Let $\mathcal{H}^t_k$ be the linear subspace of $\mathcal{Q}_k$ that spans pure states $|\psi\rangle \in \mathcal{Q}_k$ such that $\mathfrak{U}(|\psi\rangle)$ is defined and halts in t steps. Due to [Mue07, Mul08], if $t \neq t'$ then $\mathcal{H}^t_k \perp \mathcal{H}^{t'}_k$.*

Theorem 22 ([Mue07, Mul08]) *Given k, t, there is an algorithm that can enumerate $\mathcal{H}^t_k$ in the form of elementary projections $\{P_i\}$, such that $P_iP_j = \delta_{ij}P_i$ and $\sum_i P_i$ projects onto $\mathcal{H}^t_k$. Furthermore, all valid inputs σ to $\mathfrak{U}$ have $\sigma \leq P_i$ for some P_i.*

Lemma 6 *Given t, k, δ one can compute an elementary quantum operation $\Psi^{t,\delta}_k : \mathcal{Q}_k \to \mathcal{Q}$ such that if $\sigma \in \mathcal{H}^t_k$ then $D(\Psi^{t,\delta}_k(\sigma), \mathfrak{U}(\sigma)) \leq \delta$.*

Proof. Let $\Psi = \Psi^{t,\delta}_k$. The quantum operation Ψ starts by first applying quantum operation $\mathcal{E}_1$, which appends $2t$ spaces to the auxiliary, input, and output tape, and then treating the tapes as loops. The start state is appended as well as the header pointer at origin. Then it applies the approximating elementary unitary matrix $\tilde{u}$ corresponding to the unitary matrix u of $\mathfrak{U}$ (with shortened tapes) t times. Then it applies quantum operation $\mathcal{E}_2$, which projects all configuratons in the halting state $|q_f\rangle$ of the form $|s_i\#\#\ldots\rangle$ to $|s_i\rangle$ and projects configurations with states other than $|q_f\rangle$ to $\lambda \in \mathcal{Q}_0$. So $\Psi(\sigma) = \mathcal{E}_2(\tilde{u}^t\mathcal{E}_1(\sigma)\tilde{u}^{t*})$. It remains to determine the approximation matrix $\tilde{u}$.

Let $\mathcal{C}$ be the finite configuration space. Let γ be a parameter to be determined later. First cover $\mathcal{C}$ by elementary mixed states $\rho \in Q$, such that $\max_{\sigma\in\mathcal{C}} \min_{\rho\in Q} D(\sigma, \rho) < \gamma/3$. Next run the algorithm to compute the transition function of $\mathfrak{U}$ long enough to produce unitary matrix $\tilde{u}$ such that for all $\rho \in Q$, $D(u\rho u^*, \tilde{u}\rho\tilde{u}^*) < \gamma/3$. This is possible because the amplitudes of the transition function of $\mathfrak{U}$ can be computed to any accuracy. Thus for any $\sigma \in \mathcal{C}$, for proper choice of $\rho \in Q$, by the triangle inequality of trace distance,

$$\begin{aligned} D(u\sigma u^t, \tilde{u}\sigma\tilde{u}^*) &< D(u\sigma u^*, u\rho u^*) + D(u\rho u^*, \tilde{u}\rho\tilde{u}^*) + D(\tilde{u}\rho\tilde{u}^*, \tilde{u}\sigma\tilde{u}^*) \\ &< D(\sigma, \rho) + \gamma/3 + D(\rho, \sigma) \\ &< \gamma. \end{aligned}$$

If $\tilde{u}$ is run twice with any input $\sigma \in \mathcal{C}_n$, the error is bounded by

$$\begin{aligned} D(\tilde{u}^2\sigma\tilde{u}^{2*}, u^2\sigma u^{2*}) &< D(\tilde{u}^2\sigma\tilde{u}^{2*}, \tilde{u}u\sigma u\tilde{u}) + D(\tilde{u}u\sigma u\tilde{u}, u^2\sigma u^{2*}) \\ &< D(u\sigma u^*, \tilde{u}\sigma\tilde{u}^*) + \gamma \\ &< 2\gamma. \end{aligned}$$

With similar reasoning, one can see that running $\tilde{u}$ a total of ℓ times will produce a maximum error of $\gamma\ell$. So γ is set to equal δ/t. So for all $\sigma \in \mathcal{Q}_k$,

$$D(u^t\mathcal{E}_1(\sigma)u^{t*}, \tilde{u}^t\mathcal{E}_1(\sigma)\tilde{u}^{t*}) < \delta. \tag{6.1}$$

If $\sigma \in \mathcal{H}^t_{k,n}$, then $\mathcal{E}_2(u^t\mathcal{E}_1(\sigma)u^{t*}) = \mathfrak{U}(\sigma)$, so

$$\begin{aligned}\delta &\geq D(u^t\mathcal{E}_1(\sigma)u^{t*}, \tilde{u}^t\mathcal{E}_1(\sigma)\tilde{u}^{t*}) \\ &\geq D(\mathcal{E}_2(\tilde{u}^t\mathcal{E}_1(\sigma)\tilde{u}^{t*}), \mathcal{E}_2(u^t\mathcal{E}_1(\sigma)u^{t*})) \\ &= D(\Psi(\sigma), \mathfrak{U}(\sigma)).\end{aligned}$$

□

6.2 BvL Complexity and Gács Complexity

We lower and upper bound BvL complexity by Gács complexity.

6.2.1 Lower Bound

Theorem 23 *For $|\psi\rangle \in \mathcal{Q}_n$, $\mathbf{Hg}(|\psi\rangle) <^+ \mathbf{Hbvl}^\epsilon(|\psi\rangle\,|n) + \mathbf{K}(\mathbf{Hbvl}^\epsilon(|\psi\rangle\,|n), \epsilon|n) - \log(1-1.01\epsilon)$.*

Proof. Let $k = \mathbf{Hbvl}^\epsilon(|\psi\rangle\,|n)$. We construct a lower computable semi-density matrix ν by using the algorithm (relativized to n) in Theorem 22, with fixed k and all t, which results in the enumeration $\{P_i\}$, P_i is a projection operator for $\mathcal{H}^{t(i)}_k$. We define ν to be equal to $2^{-k}Q\sum_i \Psi^{t(i),0.01\epsilon}_k(P_i)Q$, where Ψ is defined by Lemma 6, and Q projects out everything but n qubit outputs.

Let σ realize $\mathbf{Hbvl}^\epsilon(|\psi\rangle)$, where $\rho = \mathfrak{U}(\sigma)$ in s steps, and $D(\rho, |\psi\rangle) < \epsilon$, and due to Theorem 22, $\sigma \leq P_i$ for some i. So due to Lemma 6, if $\xi = \Psi^{s,0.01\epsilon}_k(\sigma)$, then $D(\xi, \rho) \leq 0.01\epsilon$. So $D(\xi, |\psi\rangle) < 1.01\epsilon$. Since $|\psi\rangle \in \mathcal{Q}_n$, $D(Q\xi Q, |\psi\rangle) < 1.01\epsilon$. So, due to the definition of trace distances and fidelty of quantum states, $F(|\psi\rangle, Q\xi Q) = \langle\psi|\, Q\xi Q\,|\psi\rangle > 1-1.01\epsilon$. So, using reasoning analogous to Theorem 9 in [Gó1],

$$\begin{aligned}\mathbf{m}(k,\epsilon|n)\nu \overset{*}{<} \mathbf{m}(k,\epsilon|n)2^{-k}Q\sum_j \Psi^{t(j),0.01\epsilon}_k(P_j)Q &\overset{*}{<} \boldsymbol{\mu} \\ \mathbf{m}(k,\epsilon|n)2^{-k}Q\Psi^{t(i),0.01\epsilon}_k(P_i)Q &\overset{*}{<} \boldsymbol{\mu} \\ \mathbf{m}(k,\epsilon|n)2^{-k}Q\Psi^{s,0.01\epsilon}_k(\sigma)Q &\overset{*}{<} \boldsymbol{\mu} \\ \mathbf{m}(k,\epsilon|n)2^{-k}\,\langle\psi|\,Q\xi Q\,|\psi\rangle &\overset{*}{<} \langle\psi|\,\boldsymbol{\mu}\,|\psi\rangle \\ \mathbf{m}(k,\epsilon|n)2^{-k}(1-1.01\epsilon) &\overset{*}{<} \langle\psi|\,\boldsymbol{\mu}\,|\psi\rangle \\ k + \mathbf{K}(k,\epsilon|n) - \log(1-1.01\epsilon) &>^+ \mathbf{Hg}(|\psi\rangle).\end{aligned}$$

□

Proposition 1 *For $k \in \mathbb{N}$, $\mathbf{Hbvl}^{\frac{1}{k}}(\sigma|k) \leq \mathbf{Hbvl}(\sigma)$.*

Proof. Let $\mathcal{M}$ be the set of inputs to $\mathfrak{U}$ that realize $\mathbf{Hbvl}^{\frac{1}{k}}(\sigma|k)$. Let $\mathcal{N}$ be the set of inputs to $\mathfrak{U}$ that realize $\mathbf{Hbvl}(\sigma)$. Clearly $\mathcal{N} \subseteq \mathcal{M}$. □

Corollary 4 *For $|\psi\rangle \in \backslash$, $\mathbf{Hg}(|\psi\rangle) <^+ \mathbf{Hbvl}(|\psi\rangle\,|n) + \mathbf{K}(\mathbf{Hbvl}(|\psi\rangle\,|n)|n)$.*

Proof. From Theorem 23,

$$\mathbf{Hg}(|\psi\rangle\,|2) <^+ \mathbf{Hbvl}^{\frac{1}{2}}(|\psi\rangle) + \mathbf{K}(\mathbf{Hbvl}^{\frac{1}{2}}(|\psi\rangle), 1/2|n, 2).$$

From Propositions 1 and 2,

$$\mathbf{Hg}(|\psi\rangle) <^+ \mathbf{Hbvl}(|\psi\rangle) + \mathbf{K}(\mathbf{Hbvl}(|\psi\rangle)|n).$$

□

Proposition 2 *For every c, there is a c' such that if $a < b + c$ then $a + \mathbf{K}(a) < b + \mathbf{K}(b) + c'$.*

Proof. So $\mathbf{K}(a-b) < 2\log c + O(1)$. So $\mathbf{K}(a) < \mathbf{K}(b) + 2\log c + O(1)$. Assume not, then $b - a + c' < \mathbf{K}(a) - \mathbf{K}(b) + O(1) < 2\log c + O(1)$, which is a contradiction for $c' > 2\log c + O(1)$.□

6.2.2 Upper Bound

Theorem 24 ([Eps20]) *for $|\psi\rangle \in \mathcal{Q}_n$, $\mathbf{Hbvl}^{\epsilon}(|\psi\rangle\,|n) <^{\log} \mathbf{Hg}(|\psi\rangle)$, where $\epsilon = \sqrt{1 - 2^{-\mathbf{Hg}(|\psi\rangle) - O(\log \mathbf{Hg}(|\psi\rangle))}}$.*

Proof. We use reasoning from Theorem 7 in [GÓ1]. From Theorem 115 there exists a ρ such that $\mathbf{K}(\rho|n) - \log \langle\psi|\,\rho\,|\psi\rangle <^{\log} k$. Let $\lceil -\log \langle\psi|\,\rho\,|\psi\rangle \rceil = m$. Let $|u_1\rangle, |u_2\rangle, |u_3\rangle, \ldots$ be the eigenvectors of ρ with eigenvalues $u_1 \geq u_2 \geq u_3 \ldots$ For $y \in \mathbb{N}$, let $\rho_y = \sum_{i=1}^{y} u_i\,|u_i\rangle\,\langle u_i|$. We expand $|\psi\rangle$ in the basis of $\{|u_i\rangle\}$ with $|\psi\rangle = \sum_i c_i\,|u_i\rangle$. So we have that $\sum_i u_i |c_i|^2 \geq 2^{-m}$. Let $s \in \mathbb{N}$ be the first index i with $u_i < 2^{-m-1}$. Since $\sum_i u_i \leq 1$, it must be that $s \leq 2^{m+2}$. So

$$\sum_{i \geq s} u_i |c_i|^2 < 2^{-m-1} \sum_i |c_i|^2 \leq 2^{-m-1},$$

$$\langle\psi|\,\rho_{2^{m+2}}\,|\psi\rangle \geq \mathrm{Tr}\,\langle\psi|\,\rho_s\,|\psi\rangle > \sum_{i<s} u_i |c_i|^2 \geq 2^{-m} - \sum_{i \geq s} u_i |c_i|^2 > 2^{-m-1}.$$

We now describe a quantum Turing machine M that will construct $\rho_{2^{m+2}}$. The input σ is an ensemble $\{u_i\}_{i=1}^{2^{m+2}}$ of vectors $\{|cB(i)\rangle\}$, where $B(i)$ is the binary encoding of index $i \in \mathbb{N}$ which is of length $m+2$. Helper code c of size $=^+ \mathbf{K}(p|n)$ transforms each $|cB(i)\rangle$ into $|u_i\rangle$. Thus the size of the input is $<^+ \mathbf{K}(p|n) + m <^{\log} k$. The fidelity, F, of the approximation is

$$\langle\psi|\,\rho_{2^{m+2}}\,|\psi\rangle > 2^{-m-1} \geq 2^{-k - c\log k},$$

for proper choice of $c \in \mathbb{N}$. Thus, given the properties of trace distance and fidelity, $D(M(\sigma), |\psi\rangle) < \sqrt{1 - 2^{-k-c\log k}}$. Thus there is a progam σ' to the universal QTM $\mathfrak{U}$ with $\|\sigma'\| <^+ \|\sigma\|$ and $c' > c$ such that $D(\mathfrak{U}(\sigma'; k, n), |\psi\rangle) < \sqrt{1 - 2^{-k-c'\log k}}$. The theorem follows from this inequality. □

6.3 BvL Complexity and Kolmogorov Complexity

One question is whether quantum mechanics provides benefits to compressing classical information. This section answers this question in the negative: there are no such benefits. The quantitative amount of information in classical sources is invariant to the physical model used. These consequences make this section arguably the most important part of this manuscript. This section details the work of [Mue07, Mul09], which proves that plain Kolmogorov complexity equals BvL complexity. The original theorems of [Mue07, Mul09] in this section require in depth properties of unversal QTM $\mathfrak{U}$ in their proofs. In this section, a simpler proof is presented, only relying on Theorem 22 for characterization of $\mathfrak{U}$.

6.3.1 Prefix Free Complexity with Error

In this section, prefix free Kolmogorov complexity is related to $\mathbf{Hbvl}^\epsilon$ complexity. The first corollary has n as a conditional term. In the rest of the section, the result of the corollary without this term is proved.

Corollary 5 *For $x \in \{0,1\}^n$,*

$$\mathbf{K}(x|n) <^+ \mathbf{Hbvl}^\epsilon(|x\rangle \langle x| \, |n) + \mathbf{K}(\mathbf{Hbvl}^\epsilon(|x\rangle \langle x| \, |n), \epsilon|n) - \log(1 - 1.01\epsilon).$$

Proof. The corollary comes from Theorems 8 and 23. □

We the notion of the universal lower computable semi density operator $\boldsymbol{\nu}$ over the space $\mathcal{Q}$ of indeterminate quantum states introduced in Section 6.7.

Theorem 25 $\mathbf{K}(x) <^+ \mathbf{Hbvl}^\epsilon(|x\rangle \langle x|) + \mathbf{K}(\mathbf{Hbvl}^\epsilon(|x\rangle \langle x|), \epsilon) - \log(1 - 1.01\epsilon)$.

Proof. Let $k = \mathbf{Hbvl}^\epsilon(|y\rangle \langle y|)$. We use the algorithm in Theorem 22 to enumerate projections P_i for $\mathcal{H}_k^t$, for fixed k and all t. We construct the semi-density operator $\nu = 2^{-k} \sum_i \Psi_k^{t(i),0.01\epsilon}(P_i)$ over the space of indeterminate quantum states $\mathcal{Q}$.

Let σ realize $\mathbf{Hbvl}^\epsilon(|x\rangle)$, where $\rho = \mathfrak{U}(\sigma)$ in s steps, and $D(\rho, |x\rangle) < \epsilon$, and due to Theorem 22, $\sigma \leq P_i$ for some i. So due to Lemma 6, if $\xi = \Psi_k^{s,0.01\epsilon}(\sigma)$, then $D(\xi, \rho) \leq 0.01\epsilon$. So $D(\xi, |\psi\rangle) < 1.01\epsilon$. So, due to the definition of trace distances and fidelty of quantum states, $F(|\psi\rangle, \xi) = \langle \psi| \, \xi \, |\psi\rangle > 1 - 1.01\epsilon$. So, using reasoning analogous to Theorem 9 in [Gó1], and due to Theorem 35,

$$\begin{aligned}
\mathbf{m}(k,\epsilon)\nu \stackrel{*}{<} \mathbf{m}(k,\epsilon)2^{-k} \sum_j \Psi_k^{t(j),0.01\epsilon}(P_j) &\stackrel{*}{<} \boldsymbol{\nu} \\
\mathbf{m}(k,\epsilon)2^{-k} \Psi_k^{t(i),0.01\epsilon}(P_i) &\stackrel{*}{<} \boldsymbol{\nu} \\
\mathbf{m}(k,\epsilon)2^{-k} \Psi_k^{s,0.01\epsilon}(\sigma) &\stackrel{*}{<} \boldsymbol{\nu} \\
\mathbf{m}(k,\epsilon)2^{-k} \langle x| \, \xi \, |x\rangle &\stackrel{*}{<} \langle x| \, \boldsymbol{\nu} \, |x\rangle \\
\mathbf{m}(k,\epsilon)2^{-k}(1 - 1.01\epsilon) &\stackrel{*}{<} \mathbf{m}(x) \\
k + \mathbf{K}(k,\epsilon) - \log(1 - 1.01\epsilon) &>^+ \mathbf{K}(x).
\end{aligned}$$

□

Theorem 25 can be used to prove a weaker version of Müller's Theorem, as shown in the following corollary.

Corollary 6 $\mathbf{K}(x) <^+ \mathbf{Hbvl}(|x\rangle \langle x|) + \mathbf{K}(\mathbf{Hbvl}(|x\rangle \langle x|))$.

Proof. By Theorem 25,

$$\mathbf{K}(x) <^+ \mathbf{Hbvl}^{\frac{1}{2}}(|x\rangle \langle x|) + \mathbf{K}(\mathbf{Hbvl}^{\frac{1}{2}}(|x\rangle \langle x|), 1/2).$$

By Propositions 1 and 2 ,

$$\mathbf{K}(x) <^+ \mathbf{Hbvl}(|x\rangle \langle x|) + \mathbf{K}(\mathbf{Hbvl}(|x\rangle \langle x|)).$$

□

6.3.2 Müller's Theorem

The following new proof of Müller's Theorem is self contained, in that the only characterization of the universal QTM $\mathfrak{U}$ needed is Theorem 22.

Theorem 26 ([Mue07, Mul09])

$$\mathbf{C}(x) =^+ \mathbf{Hbvl}(|x\rangle \langle x|).$$

Proof. $\mathbf{Hbvl}(|x\rangle \langle x|) <^+ \mathbf{C}(x)$ because a universal QTM can simulate a classical Turing machine. Let $k = \mathbf{Hbvl}(|x\rangle \langle x|)$. Let $j = 2^{k+5}$ be the precision parameter. Let $\Psi_k^{t,\delta}(\cdot|j)$ be equal to $\Psi^{t,\delta}(\cdot)$ with the universal QTM $\mathfrak{U}$ (and the QTMs it simulates) with j on the auxilliary tape. Using Theorem 22, enumerate all projection operators P_i of $\mathcal{H}_k^t$ (relativized to j) for fixed k over all t. So $\mathrm{Tr}\sum_i P_i \leq 2^k$. For each P_i enumerated, compute $O_i = \Psi_k^{t(i),1/j}(P_i)$, where each O_i is a positive operator over $\mathcal{Q}$ with $\mathrm{Tr}\sum_i O_i \leq 2^k$.

Assume there is a k qubit input $\sigma \leq P_i$ and a pure state $|\psi\rangle \in \mathcal{Q}_\ell$ such that $D(\mathfrak{U}(\sigma, j), |\psi\rangle) < 1/j$. If $\xi = \Psi_k^{t(i),1/j}(\sigma|j) \leq O_i$ then $D(\xi, \mathfrak{U}(\sigma, j)) < 1/j$ and by the triangle inequality of trace distances, $D(\xi, |\psi\rangle) < 2/j$ and so $1 - 2/j < F(\xi, |\psi\rangle) = \langle\psi|\,\xi\,|\psi\rangle \leq \langle\psi|\,O_i\,|\psi\rangle = \langle\psi|\,O_i^\ell\,|\psi\rangle$, where $O_i^\ell = Q_\ell O_i Q_\ell$, where Q_ℓ is the projector onto $\mathcal{Q}_\ell$.

Let N_i^ℓ be a projection over $\mathcal{Q}_\ell$ defined from O_i^ℓ in the following way. Since $O_i^\ell = \sum_i v_i |e_i\rangle \langle e_i|$ for some orthonormal basis $\{|e_i\rangle\}$, of $\mathcal{Q}_\ell$, we define N_i^ℓ to be equal to $\sum_i [1/2 \leq v_i] |e_i\rangle \langle e_i|$. So $\mathrm{Tr}N_i^\ell \leq 2\mathrm{Tr}O_i^\ell \leq 2^{k+1}$. Some simple math shows that if $\langle\psi|\,O_i^\ell\,|\psi\rangle \geq 1 - 2/j$, then $\langle\psi|\,N_i^\ell\,|\psi\rangle \geq 1 - 4/j = 1 - 2^{-k-3}$. By Lemma 7, there can be only at most $2\mathrm{Tr}N_i^\ell$ classical states $|y\rangle$, $y \in \{0,1\}^\ell$, with $\langle y|\,N_i^\ell\,|y\rangle \geq 1 - 2^{-k-3}$. Since $\mathrm{Tr}\sum_{i,j} N_i^j \leq 2^{k+1}$, there only at most 2^{k+1} classical strings $|y\rangle$ such that there is a k qubit state ρ such that $D(\mathfrak{U}(\rho, j), |y\rangle) < j^{-1}$.

So we define an algorithm that takes in a $k+1$ bit number b. For all i, j, it enumerates P_i, O_i, and then each O_i^j and N_i^j. Then it determines the set $\{|y\rangle\}$ for classical strings $y \in \{0,1\}^\ell$ such that $\langle y|\,N_i^\ell\,|y\rangle > 1 - 2^{-k-3}$ for some $i \in \mathbb{N}$. If $|y\rangle$ is the bth state discovered with this condition, then return y. By the definition of k, there is a k qubit input ρ and $P_i \geq \rho$ such that $D(\mathfrak{U}(\rho, j), |x\rangle) < 1/j$, so x will be returned for proper choice of b. So $\mathbf{C}(x) <^+ \mathbf{Hbvl}(|x\rangle)$. □

Lemma 7 *For a rank m projection matrix P in $\mathbb{C}^n$, assume there is a orthonormal set $\{|e_i\rangle\}_{i=1}^N$ such that $\langle e_i|\,P\,|e_i\rangle > 1 - 1/4m$ for all i. Then $N < 2m$.*

Proof. Let $Q = I_n - P$. So $\langle e_i|\,Q\,|e_i\rangle \leq 1/4m$. By the Cauchy Schwarz inequality $|\,\langle e_i|Q|e_j\rangle\,|^2 \leq \langle e_i|\,Q\,|e_i\rangle \langle e_j|\,Q\,|e_j\rangle \leq (1/4m)^2$. So $|\,\langle e_i|Q|e_j\rangle\,| \leq 1/4m$.

$$\begin{aligned} 0 = \langle e_i|e_j\rangle &= \langle e_i|P + Q|e_j\rangle \\ 0 &= \langle e_i|P|e_j\rangle + \langle e_i|Q|e_j\rangle \\ |\,\langle e_i|P|e_j\rangle\,| &\leq |\,\langle e_i|Q|e_j\rangle\,| \leq 1/4m. \end{aligned}$$

Let $c_i = (\langle e_i|P|e_i\rangle)^{1/2}$, where $c_i^2 \geq 1 - 1/4m$. Let $|f_i\rangle = c_i^{-1} P\,|e_i\rangle$ be a normalized vector. So for $i \neq j$,

$$|\,\langle f_i|f_j\rangle\,| \leq |\,\langle e_i|P|e_j\rangle\,|/(c_i c_j) \leq (1/4m)/(1 - 1/4m) \leq m^{-1/2}/2.$$

The following reasoning is due to [Tao]. Suppose for contradiction $N \geq 2m$. We consider the $2m \times 2m$ Gram matrix $(\langle f_i|f_j\rangle)$, $1 \leq i, j \leq 2m$. This matrix is positive semi-definite with rank at

most m. Thus if one subtracts off the identity matrix, it has an eigenvalue of -1 with multiplicity at least m. Taking Hilbert-Schmidt norm, we conclude

$$\sum_{1\leq i,j\leq 2n; i\neq j} |\langle f_i, f_j\rangle|^2 \geq m.$$

But the left-hand side is at most $2m(2m-1)\frac{1}{4m} = m - \frac{1}{2}$, giving the desired contradiction. □

The following theorem compliments Corollary 5(1). The full proof is in [Mue07] and a proof sketch is in [Mul09].

Theorem 27 ([Mue07, Mul09]) *Given $\epsilon \in (0, 1/2e) \cap \mathbb{Q}$,*

$$\mathbf{C}(x) <^+ \frac{1}{1-4\epsilon}\mathbf{Hbvl}^\epsilon(|x\rangle) + \mathbf{K}(\epsilon).$$

6.4 BvL Complexity and Vitányi Complexity

Theorem 28 *For $|\psi\rangle \in \mathcal{Q}_n$, $\mathbf{Hbvl}^\epsilon(|\psi\rangle|n) <^+ \mathbf{Hv}(|\psi\rangle) + \mathbf{K}(\mathbf{Hv}(|\psi\rangle))$ where $\epsilon = \sqrt{1-2^{-\mathbf{Hv}(|\psi\rangle)-1}}$.*

Proof. Let $k = \lceil \mathbf{Hv}(|\psi\rangle)\rceil$. Let $|\phi\rangle$ be the elementary state that realizes $\mathbf{Hv}(|\psi\rangle)$. Thus $\langle\phi|\psi\rangle^2 \geq 2^{-k}$, which implies $D(|\psi\rangle, |\phi\rangle) \leq \sqrt{1-2^{-k}}$. Let M be a quantum Turing machine that takes the input $|0\rangle$ to $|\phi\rangle$. Thus $\mathbf{K}(M) <^+ \mathbf{K}(|\phi\rangle) \leq k$.

Thus there is an input to the universal quantum Turing machine $\mathfrak{U}$ consisting of a description (M, δ), where parameter δ such that $\sqrt{1-2^{-k}} + \delta = \sqrt{1-2^{-k-1}}$, as well as the input $|0\rangle$ to M. This parameter δ is simple relative to k. Thus thus the input is size $k + \mathbf{K}(k) + O(1)$ and will output a state $|\phi'\rangle$ such that $D(|\psi\rangle, |\phi'\rangle) \leq \sqrt{1-2^{-k-1}}$. □

The following inequality uses the Definition 5, $\langle|\psi\rangle\rangle$ and Definition 1, $\mathbf{I}$.

Theorem 29 $\mathbf{Hv}(|\psi\rangle) <^{\log} \mathbf{Hbvl}^\epsilon(|\psi\rangle) + \mathbf{K}(\epsilon|n) + \mathbf{I}(\langle|\psi\rangle\rangle : \mathcal{H}|n) - \log(1-1.01\epsilon)$.

Proof. This follows from Theorems 17 and 23. □

6.5 Incompressibility of BvL Complexity

In this section, lower bounds on the BvL complexity of ensembles of quantum states is proved. This is a presentation of Section 6.2 in [BvL01] with some clarifications added to the main proof. We note that χ is Holevo's Chi.

Theorem 30 (Lindblad-Uhlmann monotonicity) *For ensemble $\mathcal{E} = \{(\rho_i, p_i)\}$ and completely positive, trace preserving mapping map Ψ, $\chi(\Psi(\mathcal{E})) \leq \chi(\mathcal{E})$, where $\Psi(\mathcal{E}) = \{\Psi(\rho_i, p_i)\}$.*

Lemma 8 (Von Neumann Entropy Limit) *If sequence $\rho_1, \rho_2, \ldots$ has $\lim_{k\to\infty}\rho_k = \rho$ then $\lim_{k\to\infty} S(\rho_k) = S(\rho)$.*

Theorem 31 ([BvL01]) *For any set of states $\rho_1, \ldots, \rho_M$ such that $\forall_i \mathbf{Hbvl}(\rho_i) \leq l$, then*

$$l \geq S(\rho) - \frac{1}{M}\sum_i S(\rho_i),$$

where $\rho = \frac{1}{M}\sum_i \rho_i$.

Proof. Let $\sigma_1, \ldots, \sigma_M$ be the minimal $\mathfrak{U}$ programs for $\rho_1, \ldots, \rho_M$. Let $\ell = \max_i \|\sigma_i\|$. Let $\$^k$ be the quantum operation associated with the universal QTM $\mathfrak{U}$, with parameter k.

- $\mathcal{E} = \{(\rho_i, \frac{1}{M})\}$,
- $\mathcal{E}_\sigma = \{(\sigma_i, \frac{1}{M})\}$,
- $\mathcal{E}_k = \{(\rho_{k,i}, \frac{1}{M})\}$, $\rho_i = \$^k(\sigma_i)$.

By the monotonicity of Theorem 30, $\chi(\mathcal{E}_k) \leq \chi(\mathcal{E}_\sigma)$. The chi factor of the ensemble $\mathcal{E}_\sigma$ is upper bounded by the maximum length of its inputs, with $\chi(\mathcal{E}_\sigma) \leq \max_i \|\sigma_i\| \leq \ell$. The only thing to prove is that $\chi(\mathcal{E}_k)$ for sufficiently big k is close to $\chi(\mathcal{E})$.

By definition , for all i, $\lim_{k\to\infty} D(\rho_i, \rho_{k,i}) = 0$, so $\lim_{k\to\infty} \rho_{k,i} = \rho_i$. Because $\mathcal{E}$ and $\mathcal{E}_k$ have only a finite number (M) of states, by Lemma 8, one has that $\lim_{k\to\infty} \chi(\mathcal{E}_k) = \chi(\mathcal{E})$. This means for any δ there exists a k such that $\chi(\mathcal{E}) - \delta \leq \chi(\mathcal{E}_k)$. With the above inequalities we can conclude that $\chi(\mathcal{E}) - \delta \leq l$ holds for arbitrarily small $\delta > 0$ and so $l \geq \chi(\mathcal{E})$. □

Corollary 7 *For any set of orthogonal pure states $|\phi_1\rangle, \ldots, |\phi_M\rangle$ of length n, there is an i such that* **Hbvl**$(\phi_i) \geq \log M$.

Proof. All pure states have zero entropy $S(\phi_i) = 0$, hence by Theorem 31, $l \geq S(\rho)$. Because all ϕ_is are mutually orthogonal, this Von Neumann entropy $S(\rho)$ of the average state $\rho = \frac{1}{M}\sum_i |\phi_i\rangle\langle\phi_i|$ equals $\log M$.

Corollary 8 *For every length n, at least $2^n - 2^{n-c} + 1$ qubit strings of length n have complexity at least $n - c$.*

Corollary 9 *For any set of pure states $|\phi_1\rangle, \ldots, |\phi_M\rangle$ of length n, there is an i such that* **Hbvl**$(\phi_i) \geq S(\rho)$, *where $\rho = \frac{1}{M}\sum_i |\phi_i\rangle\langle\phi_i|$.*

6.5.1 A Quantum Counting Argument

Lemma 9 (Fannes Inequality) *For density matrices ρ and σ with $D(\rho, \sigma) \leq 1/e$,*

$$|S(\rho) - S(\sigma)| \leq 2D(\rho, \sigma) \log d + \eta(2D(\rho, \sigma)),$$

where d is the dimension of the Hilbert space and $\eta(x) = -x \log x \geq 0$.

Lemma 10 ([Mue07, Mul09]) *Let $\mathcal{H}$ and $\mathcal{H}'$ be separable Hilbert spaces with $0 < d = \mathrm{Dim}(\mathcal{H}) < \infty$ and $0 \leq \delta < \frac{1}{2e}$. If $\mathcal{E}$ is a quantum operation from $\mathcal{H}$ to $\mathcal{H}'$ then the maximal number N of mutually orthonormal vectors on $\mathcal{H}'$ which are produced by $\mathcal{H}$ within trace distance δ is not more than*

$$\log N \leq \frac{\log d + 4\delta \log \frac{1}{\delta}}{1 - 4\delta}.$$

Proof. For $\delta = 0$ the theorem is trivial so we can assume that $\delta > 0$. Let N_δ be a set of orthonormal vectors $|\psi\rangle \in \mathcal{H}'$ such that for each $|\psi\rangle$ there is a density operator σ on $\mathcal{H}$ with

$$D(\mathcal{E}(\sigma), |\psi\rangle\langle\psi|) \leq \delta.$$

Let $N_\delta = \{|\varphi_i\rangle\}_{i=1}^N$. Thus there exists σ_i on $\mathcal{H}$ with $D(\mathcal{E}(\sigma_i), |\varphi_i\rangle\langle\varphi_i|) \leq \delta$. For $1 \leq i \leq N$, define the projectos $P_i = |\varphi_i\rangle\langle\varphi_i|$ and $P_{N+1} = I - \sum_{i=1}^N |\varphi_i\rangle\langle\varphi_i|$. Let $\{|k\rangle\}_{k=1}^{\mathrm{Dim}(\mathcal{H}')}$ be an orthonormal basis of $\mathcal{H}'$. We define a quantum operation $Q : \mathcal{H}' \to \mathbb{C}^{N+1}$ with

$$Q(\rho) = \sum_{i=1}^{N+1} \sum_{k=1}^{\mathrm{Dim}(\mathcal{H}')} |e_i\rangle\langle k| P_i \rho P_i |k\rangle\langle e_i| .$$

The set $\{|e_i\rangle\}_{i=1}^{N+1}$ is the canonical orthonormal basis of $\mathbb{C}^{N+1}$. The operation $\mathcal{Q}$ is completely positive and trace-preserving. For $j \in [1, \ldots, N]$,

$$Q(P_j) = |e_j\rangle\langle e_j| .$$

Let $\mathbb{E}_\sigma$ be equal to the ensemble $\left\{\frac{1}{N}, \sigma_i\right\}_{i=1}^N$ and let $\sigma = \frac{1}{N}\sum_{i=1}^N \sigma_i$. We have, with χ equal to Holevo's Chi, and $S(\cdot,\cdot)$ being relative entropy.

$$\begin{aligned}\chi(\mathcal{Q}\circ\mathcal{E}(\mathbb{E}_\sigma)) &= \frac{1}{N}\sum_{i=1}^N S(\mathcal{Q}\circ\mathcal{E}(\sigma_i), \mathcal{Q}\circ\mathcal{E}(\sigma)) \leq \frac{1}{N}\sum_{i=1}^N S(\sigma_i, \sigma) \\ &= \chi(\mathbb{E}_\sigma) \leq \log d.\end{aligned}$$

Since trace distance is monotone, for every $1 \leq i \leq N$.

$$D(\mathcal{Q}\circ\mathcal{E}(\sigma_i), \mathcal{Q}(P_i)) \leq D(\mathcal{E}(\sigma_i), P_i) = D(\mathcal{E}(\sigma_i), |\varphi_i\rangle\langle\varphi_i|) \leq \delta.$$

Let $\Delta = \frac{1}{N}\sum_{i=1}^N \mathcal{Q}(P_i) = \frac{1}{N}\sum_{i=1}^N |e_i\rangle\langle e_i|$, with $S(\delta) = \log N$, and

$$D(\mathcal{Q}\circ\mathcal{E}(\sigma)), \Delta)\frac{1}{N}\sum_{i=1}^N D(\mathcal{Q}\circ\mathcal{E}(\sigma_i), \mathcal{Q}(P_i)) \leq \delta.$$

Using Fannes Inequality (Lemma 9) gives us

$$\begin{aligned}S(\mathcal{Q}\circ\mathcal{E}(\sigma_i)) = |S(\mathcal{Q}\circ\mathcal{E}(\sigma_i)) - S(Q(P_i))| &\leq 2\delta\log(N+1) + \eta(2\delta), \\ |S(\mathcal{Q}\circ\mathcal{E}(\sigma_i)) - S(Q(\Delta))| &\leq 2\delta\log(N+1) + \eta(2\delta).\end{aligned}$$

So we get the final inequality, with

$$\begin{aligned}\log d \geq \chi(\mathcal{Q}\circ\mathcal{E}(\mathbb{E}_\sigma)) &= S(\mathcal{Q}\circ\mathcal{E}(\sigma)) - \frac{1}{N}\sum_{i=1}^N S(\mathcal{Q}\circ\mathcal{E}(\sigma_i)) \\ &\geq S(\Delta) - 2\delta\log(N+1) - \eta(2\delta) - \frac{1}{N}\sum_{i=1}^N (2\delta\log(N+1) + \eta(2\delta)) \\ &= \log N - 4\delta\log(N+1) - 2\eta(2\delta) \\ &\geq (1-4\delta)\log N - 4\delta\log 2 + 4\delta\log(2\delta).\end{aligned}$$

□

6.5.2 Incompressibility for Approximate BvL Complexity

Theorem 32 ([Mue07]) *Let $\delta \in \left(0, \frac{1}{2e}\right)$ and let $|\psi_1\rangle, \ldots, |\psi_n\rangle \in \mathcal{Q}$ be e a set of mutually orthonormal pure qubit strings. Then there is an $i \in \{1, \ldots, n\}$ where*

$$\mathbf{Hbvl}^\delta(|\psi_i\rangle) > (1-4\delta)\log n - 1 - 4\delta\log\frac{1}{\delta}.$$

Proof. Let $l = \max\{\mathbf{Hbvl}^{\delta}(|\psi_i\rangle : i \in \{1, \dots n\}\}$. Then there exists σ_i, $\|\sigma_i\| \leq l$ where $D(\mathcal{U}(\sigma_i), |\psi_i\rangle \langle\psi|_i) < \delta$, where $\mathcal{U}$ is the quantum operation that corresponds to the universal QTM $\mathfrak{U}$. Thus, Theorem 10 gives

$$\log n \leq \frac{l + 1 + 4\delta \log \frac{1}{\delta}}{1 - 4\delta}.$$

So $l > (1 - 4\delta) \log n - 1 - 4\delta \log \frac{1}{\delta}$.

Theorem 33 ([Mue07]) *Let $|\psi_1\rangle, \dots, |\psi_n\rangle \in \mathcal{Q}$ be a set of pure states. There exists an $i \in \{1, \dots, n\}$ such that*

$$\mathbf{Hbvl}^{\delta}(|\psi_i\rangle) > S\left(\frac{1}{n}\sum_{j=1}^{n} |\psi_j\rangle \langle\psi_j|\right) - 4\delta \log \frac{n+1}{2\delta} - 1,$$

where S is von Neumann entropy.

Proof. Let $l = \max\{\mathbf{Hbvl}^{\delta}(|\psi_i\rangle) : i \in \{1, \dots, n\}\}\}$. Then there exists operators $\{\sigma_i\}$ of length not more than l with $D(\mathcal{U}(\sigma_i), |\psi_i\rangle \langle\psi_i|) < \delta$, where $\mathcal{U}$ is the quantum operation that corresponds to the universal QTM $\mathfrak{U}$.

Let $\mathcal{H} = \mathrm{Span}(\{|\psi_i\rangle_{i=1}^{n}\})$ and $N = \mathrm{Dim}(\mathcal{H})$ and $V : \mathcal{H} \to \mathbf{C}^{N+1}$ be an arbitray isometry (a unitary map from $\mathcal{H}$ to some N-dimensional subspace of $\mathcal{C}^{N+1}$). Let $|e\rangle \in \mathbf{C}^{N+1}$ be a normalized vector from $\mathrm{Range}(V)^{\perp}$. Define a quantum operation $\mathcal{O}$ from $\mathcal{Q}$ to $\mathbf{C}^{N+1}$ with

$$\mathcal{O}(\rho) = V P_{\mathcal{H}} \rho P_{\mathcal{H}} V^* + \sum_{k=1}^{\infty} |e\rangle \langle k| (I - P_{\mathcal{H}}) \rho (I - P_{\mathcal{H}}) |k\rangle \langle e|,$$

where $\{|k\rangle\}_{k=1}^{\infty}$ is an orthonormal basis of $\mathcal{H}^{\perp}$ in $\mathcal{Q}$, and $P_{\mathcal{H}}$ is the orthogonal projector onto $\mathcal{H}$. It easy to see that $\mathcal{O}$ is a trace preserving quantum operation. Also, for every $i \in \{1, \dots, n\}$,

$$\mathcal{O}(|\psi_i\rangle \langle\psi_i|) = V |\psi_i\rangle \langle\psi_i| V^*$$

Since trace distance is monotone with respect to quantum operations,

$$D(\mathcal{O} \circ \mathcal{U}(\sigma_i), \mathcal{O}(|\psi_i\rangle \langle\psi_i|)) \leq D(\mathcal{U}(\sigma_i), |\psi_i\rangle \langle\psi_i|) \leq \delta.$$

Let $\Delta = \frac{1}{n}\sum_{i=1}^{n} \mathcal{O}(|\psi_i\rangle \langle psi_i|)$. Since the trace distance is jointly convex,

$$D\left(\mathcal{O} \circ \mathcal{U}\left(\frac{1}{n}\sum_{i=1}^{n} \sigma_i\right), \Delta\right) \leq \frac{1}{n}\sum_{i=1}^{n} D(\mathcal{O} \circ \mathcal{U}(\sigma_i), \mathcal{O}(|\psi_i\rangle \langle\psi_i|)) \leq \delta.$$

For $i \in \{1, \dots, n\}$, the Fannes inequality (Lemma 9) gives

$$\left| S(\Delta) - S\left(\frac{1}{n}\sum_{i=1}^{n} \mathcal{O} \circ \mathcal{U}(\sigma_i)\right)\right| \leq 2\delta \log(N+1) + \eta(2\delta),$$
$$|S(\mathcal{Q} \circ \mathcal{U}(\sigma)) - S(V |\psi_i\rangle \langle\psi_i| V^*)| \leq 2\delta \log(N+1) + \eta(2\delta),$$
$$S(\mathcal{Q} \circ \mathcal{U}(\sigma)) \leq 2\delta \log(N+1) + \eta(2\delta),$$

where $\eta(x) = -x \log x > 0$. Consider the ensemble $\mathcal{E}_\sigma = \left\{\frac{1}{n}, \sigma_i\right\}_{i=1}^n$. The monotonicity property of Holevo's χ results in

$$\begin{aligned} l+1 > \chi(\mathcal{E}_\sigma) &\geq \chi(\mathcal{U}(\mathcal{E}_\sigma)) \geq \chi(\mathcal{O} \circ \mathcal{U}(\mathcal{E}_\sigma)) \\ &= S\left(\frac{1}{n}\sum_{i=1}^n \mathcal{O} \circ \mathcal{U}(\sigma_i)\right) - \frac{1}{n}\sum_{i=1}^n S(\mathcal{O} \circ \mathcal{U}(\sigma_i)) \\ &\geq S(\Delta) - 2\delta \log(N+1) - \eta(2\delta) - \frac{1}{n}\sum_{i=1}^n (2\delta \log(N+1) + \eta(2\delta)) \\ &= S\left(\frac{1}{n}\sum_{i=1}^n |\psi\rangle\langle\psi|\right) - 4\delta \log(N+1) - 4\delta \log \frac{1}{\delta}. \end{aligned}$$

The theorem follows from the fact that $N \leq n$.

6.6 No Cloning Theorem

Theorem 34 ([BvL01]) $\log \binom{m+2^n-1}{m-1} \leq \max_{|\psi\rangle} \mathbf{Hbvl}(|\psi\rangle^m) <^+ \log \binom{m+2^n-1}{m-1} + \mathbf{K}(m,n)$.

Proof. Let $N = 2^n$. We proof the upper bound first. As shown in the proof of Theorem 18, the dimension of $\mathrm{Sym}(\mathcal{H}_N^m)$ is $d = \binom{n+N-1}{N-1}$. So there is a unitary transform u from $\mathcal{H}_d$ to $\mathrm{Sym}(\mathcal{H}_N^m)$. Thus there is a quantum Turing machine M such that for every $|\psi\rangle \in \mathrm{Sym}(\mathcal{H}_N^m)$, there is a pure state $|\phi\rangle \in \mathcal{Q}_d$, such that $M(|\phi\rangle) = |\psi\rangle$. Thus the universal QTM $\mathfrak{U}$ on input $(M, |\phi\rangle, k)$ can produce an output $|\tilde{\psi}\rangle$ such that $D(|\psi\rangle, |\tilde{\psi}\rangle) < 1/k$. This implies

$$\mathbf{Hbvl}(|\psi\rangle^m) <^+ \log d + \mathbf{K}(m,n).$$

Now we prove the lower bound. Consider an orthogonal basis of pure states $\{|\psi\rangle_i\}$ over $\mathrm{Sym}(\mathcal{H}_N^m)$ (defined in the proof of Theorem 18) with $i \in \{1, \ldots, d\}$, where $d = \binom{m+N-1}{m-1}$. By Corollary 7, there is an i where $\mathbf{Hbvl}(|\psi\rangle_i) \geq \log d$. □

6.7 Indeterminate Gács Complexity

In this section, we introduce indeterminate Gács complexity, **Hn**, which is a version of Gács entropy, except over the space $\mathcal{Q}$ of indeterminate length quantum states. Indeterminate Gács complexity has deep connections to the universal QTM $\mathfrak{U}$. **Hn** is related to the expected fidelity of a state with the output of $\mathfrak{U}$ when given a random input.

6.7.1 Properties of Hn

A semi-density operator σ is a positive semidefinite operator over $\mathcal{Q}$ of non negative trace no more than 1. An elementary pure state $|\psi\rangle \in \mathcal{Q}$ is a normalized vector with elementary coefficients residing in a finite number of subspaces $\mathcal{Q}_n$. An elementary semi-density operator can be decomposed into $\sum_{i=1}^N v_i |\psi_i\rangle\langle\psi_i|$, where $|\psi_i\rangle$ is an elementary pure state and each $v_i \in \mathbb{R}_{\geq 0}$ is algebraic. A semi-density operator σ is lower computable if there is an algorithm that outputs a sequence $\{v_i, |\psi_i\rangle\}_{i=1}^\infty$, where each $v_i \in \mathbb{R}_{\geq 0}$ is algebraic and $|\psi_i\rangle$ is elementary and $\sigma = \sum_{i=1}^\infty v_i |\psi_i\rangle\langle\psi_i|$. The lower complexity of such σ is $\underline{\mathbf{m}}(\sigma) = \sum\{\mathbf{m}(p) : p \text{ lower computes } \sigma\}$. There exists a universal lower computable semi-density operator $\boldsymbol{\nu}$, such that for all lower computable semi-density operators σ, $\boldsymbol{\nu} \stackrel{*}{>} \underline{\mathbf{m}}(\sigma)\sigma$. This is constructed in the standard way in algorithmic information theory.

Definition 10 (Indeterminate Gács Complexity) *For operator σ over $\mathcal{Q}$,* $\mathbf{Hn}(\sigma) = \lceil -\log \mathrm{Tr}\boldsymbol{\nu}\sigma \rceil$.

Exercise 6 *For density operator σ over $\mathcal{Q}_n$, prove* $\mathbf{Hg}(\sigma) <^+ \mathbf{Hn}(\sigma) <^+ \mathbf{Hg}(\sigma) + \mathbf{K}(n)$.

The following theorem is analagous to Theorem 8, except the conditioning on the number of qubits, n, is removed. Thus, this theorem displays the advantage of using $\boldsymbol{\nu}$ instead of $\boldsymbol{\mu}$.

Theorem 35 *For $x \in \{0,1\}^*$,* $\mathbf{Hn}(|x\rangle) =^+ \mathbf{K}(x)$.

Proof. Since $\boldsymbol{\nu}$ is a lower computable semi-density operator, its trace is not more than 1, so $p(x) = \langle x| \boldsymbol{\nu} |x\rangle$ is a lower computable semi-measure. So $p(x) \stackrel{*}{<} \mathbf{m}(x)$. Let σ be the lower computable semi density operator $\sigma = \sum_{x\in\{0,1\}^*} \mathbf{m}(x) |x\rangle \langle x|$. So $\boldsymbol{\nu} \stackrel{*}{>} \sigma$ which implies $\langle x| \boldsymbol{\nu} |x\rangle \stackrel{*}{>} \langle x| \sigma |x\rangle \stackrel{*}{>} \mathbf{m}(x)$.□

Similarly to $\boldsymbol{\mu}$, the universal matric $\boldsymbol{\nu}$ can be decomposed into a weighted sum of elementary indeterminate states.

Theorem 36 $\boldsymbol{\nu} \stackrel{*}{=} \sum\{\mathbf{m}(|\psi\rangle) |\psi\rangle \langle\psi| : |\psi\rangle$ *is elementary*$\}$.

Proof. The operator $\rho = \sum\{\mathbf{m}(|\psi\rangle) |\psi\rangle \langle\psi| : |\psi\rangle$ is elementary$\}$ is a lower computable semi-density operator, so $\boldsymbol{\nu} \stackrel{*}{>} \rho$. Furthermore, any lower computable semi-density operator σ can be decomposed into $\sum_{i\in\mathbb{N}} p(i) |\psi_i\rangle \langle\psi_i|$, where each $|\psi_i\rangle$ is elementary and p is a lower computable semi-measure. Thus we have that $\mathbf{m}(p)p \stackrel{*}{<} \mathbf{m}$ and so $\sigma \stackrel{*}{<} \sum_i \mathbf{m}(i) |\psi_i\rangle \langle\psi_i| / \mathbf{m}(p)$. Thus the bounds come from setting σ to $\boldsymbol{\nu}$. □

Indeterminate Gács complexity is non increasing on elementary quantum operations that preserve rank.

Theorem 37 *Let $\mathcal{E} : \mathcal{Q} \to \mathcal{Q}$ be an elementary quantum operation and let $\mathcal{E}(|\psi\rangle)$ be a pure state.* $\mathbf{Hn}(\mathcal{E}(|\psi\rangle)) <^+ \mathbf{Hn}(|\psi\rangle) + \mathbf{K}(\mathcal{E})$.

Proof. Let $|\xi\rangle = \mathcal{E}(|\psi\rangle)$. The semi-density operator $\boldsymbol{\nu}_{\mathcal{E}} = \sum_{\text{Elementary } |\phi\rangle\in\mathcal{Q}} \mathbf{m}(\mathcal{E}(|\phi\rangle \langle\phi|))\mathcal{E}(|\phi\rangle \langle\phi|)$ is lower computable, with $\mathbf{m}(\mathcal{E})\boldsymbol{\nu}_{\mathcal{E}} \stackrel{*}{<} \boldsymbol{\nu}$. So, due to monotonicity of fidelity with respect to quantum operations and Theorem 36,

$$\begin{aligned}
\langle\psi| \boldsymbol{\nu} |\psi\rangle &= \sum_{\text{Elementary } |\phi\rangle\in\mathcal{Q}} \mathbf{m}(|\phi\rangle) F(|\psi\rangle, |\phi\rangle) \\
&\leq \sum_{\text{Elementary } |\phi\rangle\in\mathcal{Q}} \mathbf{m}(|\phi\rangle) F(|\xi\rangle, \mathcal{E}(|\phi\rangle)) \\
&\stackrel{*}{<} \sum_{\text{Elementary } |\phi\rangle\in\mathcal{Q}} \mathbf{m}(\mathcal{E}(|\phi\rangle)) F(|\xi\rangle, \mathcal{E}(|\phi\rangle)) / \mathbf{m}(\mathcal{E}) \\
&\stackrel{*}{=} \langle\xi| \boldsymbol{\nu}_{\mathcal{E}} |\xi\rangle / \mathbf{m}(\mathcal{E}) \\
&\stackrel{*}{<} \langle\xi| \boldsymbol{\nu} |\xi\rangle / \mathbf{m}(\mathcal{E}).
\end{aligned}$$

□

6.7.2 Expected Fidelity

In this section, the indeterminate Gács complexity, **Hn** of an indeterminate length state $|\psi\rangle \in \mathcal{Q}$ is related the expected fidelity of $|\psi\rangle$ with the output of the universal QTM $\mathfrak{U}$ when given a random input. This result requires using the mutual information with the halting sequence $\mathbf{I}(\cdot;\mathcal{H})$ term. We start with a straightforward reworking Lemma 5 for indeterminate length states.

Definition 11 *We now describe an infinite encoding scheme for an arbitrary (not necessarily elementary) quantum indeterminate pure state $|\psi\rangle \in \mathcal{Q}$. This scheme is defined as an injection between the set of indeterminate length pure states and $\{0,1\}^\infty$. We define $\langle\langle|\psi\rangle\rangle\rangle$ to be an ordered list of the encoded tuples $\langle\langle|\theta\rangle\rangle, q, [|\langle\psi|\theta\rangle|^2 \geq q]\rangle$, over all elementary states $|\theta\rangle \in \mathcal{Q}$ and rational distances $q \in \mathbb{Q}_{>0}$.*

Note that the information term $\mathbf{I}$ used in Lemma 11 is from Definition 1 and $\mathcal{H} \in \{0,1\}^\infty$ is the infinite halting sequence.

Lemma 11 *For indeterminate length pure quantum state $|\psi\rangle \in \mathcal{Q}$, and $|\phi\rangle$ varying over elementary states of $\mathcal{Q}$,*

$$\min_{|\phi\rangle} \mathbf{K}(|\phi\rangle) - \log|\langle\psi|\phi\rangle|^2 <^{\log} -\log\sum_{|\phi\rangle}\mathbf{m}(|\phi\rangle)|\langle\psi|\phi\rangle|^2 + \mathbf{I}(\langle|\psi\rangle\rangle : \mathcal{H}).$$

Proof. Let $\mathcal{D}$ be a finite set of elementary pure states of $\mathcal{Q}$, computable from $\langle|\psi\rangle\rangle$ and the value $g = \lceil -\log\sum_{|\phi\rangle}\mathbf{m}(|\phi\rangle)|\langle\psi|\phi\rangle|^2\rceil$ such that

$$-\log\sum_{|\theta\rangle\in\mathcal{D}}\mathbf{m}(|\theta\rangle)|\langle\psi|\theta\rangle|^2 \leq g+1.$$

It is computable because there exists an algorithm that can find $\mathcal{D}$ by the following method. The algorithm enumerates all indeterminate length elementary states $|\theta\rangle$ of $\mathcal{Q}$. This algorithm approximates the algorithmic probabilities $\mathbf{m}(|\theta\rangle)$ (from below) with $\widehat{\mathbf{m}}(|\theta\rangle)$. This algorithm uses $\langle|\psi\rangle\rangle$ to approximate $|\langle\theta|\psi\rangle|^2$ from below with $\widehat{|\langle\theta|\psi\rangle|}^2$. This algorithm stops when it finds a finite set $\mathcal{D}$ such that

$$-\log\sum_{|\theta\rangle\in\mathcal{D}}\widehat{\mathbf{m}}(|\theta\rangle)\widehat{|\langle\theta|\psi\rangle|}^2 \leq g+1.$$

Thus we have that $\mathbf{K}(\mathcal{D}|g,\langle|\psi\rangle\rangle) = O(1)$. Let $f:\mathcal{D}\to\mathbb{W}$ be a elementary function such that $|-\log|\langle\psi|\theta\rangle|^2 - f(|\theta\rangle)| \leq 1$. One such f is computable relative to $\langle|\psi\rangle\rangle$, and g. Firstly this is because D is computable from $\langle|\psi\rangle\rangle$ and g. The individual values of f are computable from $\langle|\psi\rangle\rangle$, since $|\langle\psi|\theta\rangle|^2$ can be computed to any degree of accuracy. So $\mathbf{K}(f|g,\langle|\psi\rangle\rangle) = O(1)$ and

$-\log \sum_{|\theta\rangle \in \mathcal{D}} \mathbf{m}(|\theta\rangle) 2^{-f(|\theta\rangle)} \leq g+2$. One then has that

$$\begin{aligned}
\min_{|\phi\rangle} \mathbf{K}(|\phi\rangle) - \log |\langle \psi|\phi\rangle|^2 &<^{+} \min_{\theta \in \mathcal{D}} \mathbf{K}(|\theta\rangle) + f(|\theta\rangle) \\
&<^{\log} -\log \sum_{|\theta\rangle \in \mathcal{D}} \mathbf{m}(|\theta\rangle) 2^{-f(|\theta\rangle)} + \mathbf{I}(\langle f \rangle; \mathcal{H}). && (6.2) \\
&<^{\log} g + \mathbf{I}(\langle f \rangle; \mathcal{H}) && (6.3) \\
&<^{\log} g + \mathbf{I}(\langle |\psi\rangle \rangle : \mathcal{H}) + \mathbf{K}(\langle f \rangle | \langle |\psi\rangle \rangle) && (6.4) \\
&<^{\log} g + \mathbf{I}(\langle |\psi\rangle \rangle : \mathcal{H}) + \mathbf{K}(g) \\
&<^{\log} -\log \sum_{|\phi\rangle} \mathbf{m}(|\phi\rangle) |\langle \psi|\phi\rangle|^2 + \mathbf{I}(\langle |\psi\rangle \rangle : \mathcal{H}).
\end{aligned}$$

Inequality 6.2 is due to Theorem 114. Inequality 6.3 is due to the definition of f and $\mathcal{D}$. Inequality 6.4 is due to the definition of $\mathbf{I}$, where $\mathbf{I}(x;\mathcal{H}) <^{+} \mathbf{I}(\alpha : \mathcal{H}) + \mathbf{K}(x|\alpha)$. □

Theorem 38 *Let $|\phi\rangle$ vary over elementary pure states in $\mathcal{Q}$.*

$$\mathbf{Hn}(|\psi\rangle <^{+} \min_{|\phi\rangle \in \mathcal{Q}} \mathbf{K}(|\phi\rangle) - \log |\langle \phi|\psi\rangle|^2 <^{\log} \mathbf{Hn}(|\psi\rangle) + \mathbf{I}(\langle |\psi\rangle \rangle : \mathcal{H}).$$

Proof. The upper bound comes from Lemma 11. □

We now define the distribution over $\mathcal{Q}$ which will be the input to the universal QTM, $\mathfrak{U}$. The number of qubits $n$is first chosen by the algorithmic probability $\mathbf{m}$. Then an n qubit pure state is chosen uniformly over $\mathcal{Q}_n$.

Definition 12 ($\mathcal{P}_{\mathfrak{U}}$) *We assume a distribution $\mathcal{P}_{\mathfrak{U}}$ over pure states in $\bigcup_n \mathcal{Q}_n$. First a length $n \in \mathbb{N}$ is chosen with probability $\mathbf{m}(n)$, and $\emptyset$ is chosen for $1 - \mathbf{m}(\mathbb{N})$. Then a random pure state $|\psi\rangle$ is chosen uniformly over $\mathcal{Q}_n$.*

We set the fidelity between an undefined output of $\mathfrak{U}$ and any other pure state to equal 0.

Remark 3 *Let $|\phi\rangle$ be a pure state such that $\mathfrak{U}(|\phi\rangle)$ is undefined. This means there is no projection P_i from the algorithm in Theorem 22 such that $|\phi\rangle\langle\phi| \leq P_i$. We set $F(\mathfrak{U}(|\phi\rangle), |\psi\rangle) = 0$ for all pure states $|\psi\rangle$.*

Definition 13 *Let $\mathbf{bb}^{-1}(n) = \min\{\ell : \exists$ program $p \in \{0,1\}^{\ell}, \|U(p)\| > n\}$. $a <^{\mathbf{bb}} b$ is equal to $a < b + \mathbf{bb}^{-1}(b) + O(1)$.*

Lemma 12 *There is an algorithm that on input of elementary matrix P, $t \in \mathbb{N}$ and $\delta \in \mathbb{Q}_{>0}$, where P is a projection for a subspace of $\mathcal{H}_k^t$, outputs an elementary quantum operation $\Psi_k^{t,\delta}$ such that such that $\mathrm{Tr}\Psi_k^{t,\delta}(P) = \mathrm{Tr}P$, $\mathfrak{U}(P) < \Psi_k^{t,\delta}(P) + I_\delta$, where I_δ is an operator over $\mathcal{Q}$ with the property $I_\delta = \sum_{x \in \{0,1\}^*} c_x |x\rangle\langle x|$, and $c_x \in \mathbb{Q}_{\geq 0}$, $\sum_{x \in \{0,1\}^*} c_x \leq \delta$.*

Proof. We describe the quantum operation $\Psi_k^{t,\delta}$. First a quantum operation $\mathcal{E}_1$ is applied, which appends $2t$ blank cells to the auxiliary input, input and output tape. The start state is appended as well as the header pointer at origin. Then an approximation $\tilde{u}$ of the unitary matrix is run t times. Then it applies quantum operation $\mathcal{E}_2$, which projects all configurations in the halting state $|q_f\rangle$ of the form $|s_i \# \# \ldots\rangle$ to $|s_i\rangle$ and projects configurations with states other than $|q_f\rangle$ to

$\lambda \in \mathcal{Q}_0$. So $\Psi_k^{t,\delta}(\sigma) = \mathcal{E}_2(\tilde{u}^t \mathcal{E}_1(\sigma)\tilde{u}^{t*})$. Since the transition function of $\mathfrak{U}$ is computable, $\tilde{u}$ can be computed to any degree of accuracy. Compute $\tilde{u}$ such that $\|\Psi_k^{t,\delta}(P) - \mathfrak{U}(P)\|_2 \leq \delta 2^{-2t-1}$ and construct $I_\delta = \sum_{n=1}^{2t} \sum_{x\in\{0,1\}^n} \delta 2^{-2t-1} |x\rangle \langle x|$. Thus, $\mathfrak{U}(P) < \Psi^{t,\delta}(P) + I_\delta$. □

Proposition 3 *For mixed states ρ, σ,*

$$\|\rho - \sigma\|_2 \leq D(\rho, \sigma).$$

Proof. Let $\Delta = \rho - \sigma$, where Δ is Hermitian. We can assume that on of its eigenvalues which has largest absolute value is positive. So

$$\|\Delta\|_2 = \max_{\||v\rangle\|=1} \langle v| \Delta |v\rangle = \max_{\text{proj } P, \mathrm{Tr}P=1} \mathrm{Tr}P\Delta \leq \max_{\text{proj } P} \mathrm{Tr}P\Delta = D(\rho, \sigma).$$

□

Remark 4 *For elementary indeterminate state $|\phi\rangle$, let $M_{|\phi\rangle}$ be the QTM that takes no input and then performs a unitary transform, rotating $|\#\ldots\#\rangle$ from the input tape to $|\phi\rangle$ on the output tape. Because $M_{|\phi\rangle}$ takes no input, $|x_{|\phi\rangle}\rangle \langle x_{|\phi\rangle}|$ is a valid input to $\mathfrak{U}$, where $x_{|\phi\rangle} \in \{0,1\}^*$ is a classical program to compute $(M_{|\phi\rangle}, j)$, with j being the precision parameter expected by $\mathfrak{U}$.*

Theorem 39

$$\mathbf{Hn}(|\psi\rangle) <^{\mathbf{bb}} -\log \mathbf{E}_{|\phi\rangle\sim\mathcal{P}_\mathfrak{U}} [F(\mathfrak{U}(|\phi\rangle), |\psi\rangle)] <^{\log} \mathbf{Hn}(|\psi\rangle) + \mathbf{I}(\langle|\psi\rangle\rangle : \mathcal{H}).$$

Proof. Let $b = \mathbf{bb}^{-1}(\mathbf{Hn}(|\psi\rangle))$. We first prove the lower bound. Let $\ell = 2^{-\mathbf{bb}(b)}$. By Lemma 12,

$$\begin{aligned}
&\mathbf{E}_{|\phi\rangle\sim\mathcal{P}} [F(\mathfrak{U}(\phi), |\psi\rangle)] \\
&= \sum_{k=1}^{\infty} \mathbf{m}(k) 2^{-k} \left(\sum_t \sum_{P_i \in \mathcal{H}_k^t} \langle\psi| \mathfrak{U}(P_i) |\psi\rangle \right) \\
&= \langle\psi| \sum_{k=1}^{\infty} \mathbf{m}(k) 2^{-k} \left(\sum_t \sum_{P_i \in \mathcal{H}_k^t} \mathfrak{U}(P_i) \right) |\psi\rangle \\
&\leq \langle\psi| \sum_{k=1}^{\infty} \mathbf{m}(k) 2^{-k} \left(\sum_t \sum_{P_i \in \mathcal{H}_k^t} \Psi_k^{t,\ell}(P_i) + I_\ell \right) |\psi\rangle .
\end{aligned}$$

Note that $\xi = \sum_{k=1}^{\infty} \mathbf{m}(k) 2^{-k} \left(\sum_t \sum_{P_i \in \mathcal{H}_k^t} \Psi_k^{t,\ell}(P_i) \right)$ is a lower computable semi-density operator, with $\mathbf{m}(\ell)\xi \stackrel{*}{<} \boldsymbol{\nu}$, so

$$\begin{aligned}
\mathbf{E}_{|\phi\rangle\sim\mathcal{P}} [F(\mathfrak{U}(\phi), |\psi\rangle)] &\stackrel{*}{<} \langle\psi| \boldsymbol{\nu} |\psi\rangle / \mathbf{m}(\ell) + \langle\psi| I_\ell |\psi\rangle \stackrel{*}{<} \langle\psi| \boldsymbol{\nu} |\psi\rangle / \mathbf{m}(\ell) \\
\mathbf{Hn}(|\psi\rangle) &<^{\mathbf{bb}} -\log \mathbf{E}_{|\phi\rangle\sim\mathcal{P}} [F(\mathfrak{U}(\phi), |\psi\rangle)] .
\end{aligned}$$

We now prove the upper bound. Let

$$|\phi\rangle = \arg \min_{\text{Elementary } |\xi\rangle\in\mathcal{Q}} \mathbf{K}(|\xi\rangle) - \log |\langle\xi|\psi\rangle|^2.$$

Let $x_{|\phi\rangle} \in \mathcal{X}$ be such that $U(x_{|\phi\rangle}) = (M_{|\phi\rangle}, j)$, where U is a universal classical Turing machine $M_{|\phi\rangle}$ is a QTM that outputs $|\phi\rangle \in \mathcal{Q}$, and j is the precision parameter for $\mathfrak{U}$ to be determined later. Furthermore $\|x_{|\psi\rangle}\| = \mathbf{K}(M_{|\phi\rangle}, j)$. The QTM $M_{|\phi\rangle}$ is described in Remark 4. Let $P(\mathcal{H}_k^t)$ be the projections enumerated by the algorithm in Theorem 22. Thus there exists one elementary projection operator P_i over $\mathcal{H}^t_{\|x_{|\psi\rangle}\|}$ such that $|x_{|\phi\rangle}\rangle \langle x_{|\phi\rangle}| \leq P_i$. So,

$$\begin{aligned}
\mathbf{E}_{|\phi\rangle\in\mathcal{P}}\left[F(\mathfrak{U}(|\phi\rangle), |\psi\rangle)\right] &= \langle\psi| \sum_{k,t\in\mathbb{N}} \sum_{P\in P(\mathcal{H}_k^t)} \mathbf{m}(k) 2^{-k} \mathfrak{U}(P) |\psi\rangle \\
&\geq \mathbf{m}(\|x_{|\phi\rangle}\|) 2^{-\|x_{|\phi\rangle}\|} \langle\psi| \mathfrak{U}(|x_{|\phi\rangle}\rangle \langle x_{|\phi\rangle}|) |\psi\rangle \\
&\stackrel{*}{>} \mathbf{m}(\|x_{|\phi\rangle}\|) \mathbf{m}\left(M_{|\phi\rangle}, j\right) \langle\psi| \mathfrak{U}(|x_{|\phi\rangle}\rangle \langle x_{|\phi\rangle}|) |\psi\rangle .
\end{aligned}$$

Let $c = 2^{-\mathbf{Hn}(|\psi\rangle) - \mathbf{I}(\langle|\psi\rangle\rangle : \mathcal{H})}$, and $d \in O(\log(\mathbf{Hn}(|\psi\rangle) + \mathbf{I}(\langle|\psi\rangle\rangle : \mathcal{H})))$ be the error term of Theorem 38. Taking into account Proposition 3, choose j so that $\|\mathfrak{U}(|x_{|\phi\rangle}\rangle \langle x_{|\phi\rangle}|) - |\phi\rangle \langle\phi|\|_2 \leq D(|\phi'\rangle \langle\phi'|, |\phi\rangle \langle\phi|) \leq c/2d$. The state $|\phi'\rangle$ is the result of simulating $M_{|\phi\rangle}$ (with no input) with $\mathfrak{U}$ with precision parameter j. So $\mathbf{K}(j|c, |\phi\rangle) = O(1)$. If $p_{|\phi\rangle}$ is a shortest program to compute $|\phi\rangle$, we have,

$$\begin{aligned}
\mathbf{E}_{|\phi\rangle\in\mathcal{P}}\left[F(\mathfrak{U}(|\phi\rangle), |\psi\rangle)\right] \geq & \mathbf{m}(\mathbf{K}(|\phi\rangle, j)) \mathbf{m}(|\phi\rangle, j) \left(|\langle\psi|\phi\rangle|^2 - c/2d\right) \\
\stackrel{*}{>} & \mathbf{m}(\mathbf{K}(|\phi\rangle, c)) \mathbf{m}(|\phi\rangle, c) \left(|\langle\psi|\phi\rangle|^2 - c/2d\right) && (6.5) \\
\stackrel{*}{>} & \mathbf{m}(c) \mathbf{m}(\mathbf{K}(c|p_{|\phi\rangle})) \mathbf{m}(\mathbf{K}(|\phi\rangle)) \mathbf{m}(|\phi\rangle) \left(|\langle\psi|\phi\rangle|^2 - c/2d\right) && (6.6) \\
\stackrel{*}{>} & \mathbf{m}(c) \mathbf{m}(\mathbf{K}(c|p_{|\phi\rangle})) \mathbf{m}(\mathbf{K}(|\phi\rangle)) (c/d - c/2d) \\
\stackrel{*}{>} & \mathbf{m}(c) \mathbf{m}(\mathbf{K}(c|p_{|\phi\rangle})) c/d && (6.7)
\end{aligned}$$

$$\begin{aligned}
-\log \mathbf{E}_{|\phi\rangle\in\mathcal{P}}\left[F(\mathfrak{U}(|\phi\rangle), |\psi\rangle)\right] &<^{\log} \mathbf{Hn}(|\psi\rangle + \mathbf{I}(\langle|\psi\rangle\rangle : \mathcal{H}) + \mathbf{K}(c) + \mathbf{K}(\mathbf{K}(c|p_{|\psi\rangle})) \\
&<^{\log} \mathbf{Hn}(|\psi\rangle + \mathbf{I}(\langle|\psi\rangle\rangle : \mathcal{H}).
\end{aligned}$$

Equation 6.5 uses Proposition 2. Equation 6.6 is due to Theorem 38. Equation 6.7 is due to a redefinition of d. □

Corollary 10 *Let $\mathcal{P}_k$ be the uniform distribution of pure states over $\mathcal{Q}_k$. For $|\psi\rangle \in \mathcal{Q}$,*

$$\mathbf{Hn}(|\psi\rangle) <^{\mathbf{bb}} -\log \mathbf{E}_{|\phi\rangle\sim\mathcal{P}_k}\left[F(\mathfrak{U}(|\phi\rangle), |\psi\rangle)\right] + \mathbf{K}(k).$$

Chapter 7

Quantum Typicality

There is no standard definition for what constitutes an outlier. Some reasonable definitions are:

- *An observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism.*
- *A datapoint generated by contaminating models different from the one generating the rest of the data.*
- *An observation that lies outside the overall pattern of a distribution. Usually, the presence of an outlier indicates some sort of problem.*

In machine learning, an outlier is a point residing a distance away from the general population, usually in a low (compared to that used in complexity theory) dimensional space. An example of an outlier in the task of regression is shown in Figure 7.1. If one were to define a score of how much

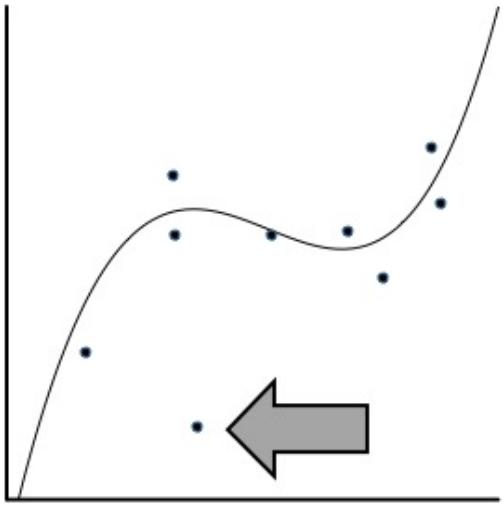


Figure 7.1: In the task of regression, a (usually parameterized) smooth function is fitted to a population of datapoints. The datapoints which are far away from the curve can be considered to outliers, i.e. the product of some sort of error or white noise effect.

a datapoint is an outlier, it should take into account perfectly randomly generated observations have an absence of regularity. Thus the theory of algorithms and recursive functions should be used to quantify this regularity.

To this end, in Algorithmic Information Theory, outliers are modeledd with the randomness deficiency function, which is defined to be $\mathbf{d}(x|p) = -\log p(x) - \mathbf{K}(x|p)$, where x is a string and p is a computable probability. Thus it is the difference between the length of x's Shannon-Fano code with respect to p and its Kolmogorov complexity. It is a score in refutation of the statement thst x was generated by model p. If the sequence is atypical, say $x = 0^n$, for the uniform distribution p over strings of length n, then $-\log p(a) = n$ which is much larger than $\mathbf{K}(a|p) = \mathbf{K}(0^n|p) = O(1)$, causing a high deficiency of randomness, with $\mathbf{d}(a|P) = n - O(1)$. Thus $\mathbf{d}$ is a score of how sequences have regularities that makes them compressible.

The exponent of the randomness deficiency is a p- test, with $\sum_x p(x)2^{\mathbf{d}(a|p)} \leq 1$. The deficiency of randomness has the benefit that its exponent is a universal test, where $\mathbf{d}(x|p) =^+ \log \mathbf{t}_p(x)$, where $\mathbf{t}_p$ is a universal lower-computable p-test. This is why randomness deficiency is a good candidate as an outlier score. For the universal semi-measure $\mathbf{m}$, we have that $\mathbf{d}(x|\mathbf{m}) = O(1)$, thus giving evidence that $\mathbf{m}$ is a universal apriori distribution, as there is no refutation to statement x is generated by $\mathbf{m}$.

Randomness deficiency can be generalized to infinite sequences and points in computable metric spaces, as shown in Chapter 13. For the infinite sequences case, $\mathbf{D}(\alpha|P) = \sup_n -\log P(\alpha[0..n]|P) - \mathbf{K}(\alpha[0..n])$. Like the strings case $\mathbf{D}(\alpha|P) =^+ \log \mathbf{t}_P(\alpha)$, where $\mathbf{t}_P$ is a universal lower computable P-test.

Obvious limitation of $\mathbf{d}$ is that is non-computable, there is no computable algorithm that can even approximate $\mathbf{d}$. This is due to $\mathbf{d}$ being defined using $\mathbf{K}$ which has uncomputability properties.

7.1 Definition of Quantum Randomness Deficiency

This chapter extends the definition of randomness deficiency to pure and mixed quantum state. In [Gác01], the quantum notion of randomness deficiency was introduced. This quantum randomness deficiency measures the algorithmic atypicality of a pure or mixed quantum state ρ with respect to a second quantum mixed state σ. Mixed states σ are used to model random mixtures $\{p_i\}$ of pure states $\{|\psi_i\rangle\}$, so quantum randomness deficiency is a score of how atypical a quantum state is with respect to a mixture. We first describe typicality with respect to computable σ, and then generalize to uncomputable σ.

Given a density matrix σ, a σ-test is a lower computable matrx T such that $\mathrm{Tr}T\sigma = 1$. Let $\mathcal{T}_\sigma$ be the set of all σ-tests. If σ is computable, there exists a universal σ test $\mathbf{t}_\sigma$, that is lower computable relative to the number of qubits n, $\mathrm{Tr}\sigma\mathbf{t}_\sigma \leq 1$, and for every lower computable σ test T, $O(1)\mathbf{t}_\sigma > \mathbf{m}(T|\sigma)T$.

This universal test can be computed the following manner, analagously to the classical case (see [Gác21]). A program enumerates all strings p and lower computes $\mathbf{m}(p|\sigma)$. The program then runs p and continues with the outputs as long as p outputs a series of positive semi-definite matrices T_i such that $\mathrm{Tr}T_i\sigma \leq 1$ and $T_i \leq T_{i+1}$. If p outputs something other than this sequence or does not halt, the sequence is frozen. $\mathbf{t}_\sigma = \sum_p \mathbf{m}(p|\sigma) \lim_i T_i$ is the weighted sum of all such outputs of progams p.

Definition 14 (Quantum Randomness Deficiency) *For mixed states σ and ρ, computable σ, $\mathbf{d}(\rho|\sigma) = \log \mathrm{Tr}\, \mathbf{t}_\sigma\rho$.*

The quantum randomness deficiency, among other interpretations, is score of how typical a pure state is with respect to an algorithmically generated quantum source. Indeed, suppose there is a computable probability P over encodings of elementary orthogonal pure states $\{\langle|\psi_i\rangle\rangle\}$ of orthogonal pure states $\{|\psi_i\rangle\}$, with corresponding density matrix $\sigma = \sum_i P(\langle|\psi_i\rangle\rangle)\,|\psi_i\rangle\,\langle\psi_i|$.

Then there is a lower-computable σ-test $T = \sum_i 2^{\mathbf{d}(\langle|\psi_i\rangle\rangle|P)} |\psi_i\rangle\langle\psi_i|$ with $O(1)\mathbf{t}_\sigma > T$. Thus $\mathbf{d}(|\psi_i\rangle|\sigma) >^+ \mathbf{d}(\langle|\psi_i\rangle\rangle|P)$, giving high scores to pure states $|\psi_i\rangle$ which are atypical of the source. In general the $\mathbf{d}(|\phi\rangle|\sigma)$ score for arbitrary $|\phi\rangle$ will be greater than a combination of $\mathbf{d}(\cdot|P)$ scores, with $\mathbf{d}(|\phi\rangle|\sigma) >^+ \log\sum 2^{\mathbf{d}(\langle|\psi_i\rangle\rangle|P)}|\langle\phi|\psi_i\rangle|^2$. In fact $\mathbf{d}$ is equivalent to the classical definition of randomness deficiency when σ is purely classical, i.e. only diagonal.

Exercise 7 *For an orthogonal sequence of elementary states $|1\rangle, |2\rangle, |3\rangle, \ldots$, prove that $\mathbf{d}(|i\rangle||j\rangle) = \infty$ for $i \neq j$.*

Exercise 8 *Show that $\mathbf{d}(\sigma|2^{-n}\mathbf{1}) =^+ n - \mathbf{Hg}(\sigma)$. This mirrors the classical case, where for a uniform measure U_n over strings $x \in \{0,1\}^n$, $\mathbf{d}(x|U_n) = n - \mathbf{K}(x|n)$.*

Theorem 40 *For diagonal $\sigma = \sum_i p(i)|i\rangle\langle i|$, $\mathbf{d}(|i\rangle|\sigma) =^+ \mathbf{d}(i|p)$.*

Proof. The positive semi-definite matrix $T = \sum_i 2^{\mathbf{d}(i|p)}|i\rangle\langle i|$ is a σ-test, so $T \stackrel{*}{<} \mathbf{t}_\sigma$ and thus $\mathbf{d}(|i\rangle|\sigma) >^+ \log\langle i|T|i\rangle =^+ \mathbf{d}(i|p)$. The function $t(i) = \langle i|\mathbf{t}_\sigma|i\rangle$ is a lower computable p-test, so $\mathbf{d}(i|P) >^+ \mathbf{d}(|i\rangle|\sigma)$. □

The following theorem shows that randomness deficiency $\mathbf{d}(\rho|\sigma)$ parallels the classical definition of randomness decificiency, $\mathbf{d}(x|P) = \log \mathbf{m}(x)/P(x)$.

Theorem 41 ([GÓ1]) *Relativized to elementary invertible σ, $\log\mathbf{d}(\rho|\sigma) =^+ \log \mathrm{Tr}\rho\sigma^{-1/2}\boldsymbol{\mu}\sigma^{-1/2}$.*

Proof. The matrix $\sigma^{1/2}\mathbf{t}_\sigma\sigma^{1/2}$ is a lower-computable semi density matrix, so $\mathbf{t}_\sigma \stackrel{*}{<} \sigma^{-1/2}\boldsymbol{\mu}\sigma^{-1/2}$. This implies $\mathrm{Tr}\mathbf{t}_\sigma\rho \stackrel{*}{<} \mathrm{Tr}\rho\sigma^{-1/2}\boldsymbol{\mu}\sigma^{-1/2}$. □

Exercise 9 ([GÓ1]) *Relativized to elementary invertible σ, show that $\mathbf{d}(\rho|\sigma) =^+ \log T\rho$, where*

$$T = \sum_{|\psi\rangle} \frac{\mathbf{m}(|\psi\rangle)|\psi\rangle\langle\psi|}{\langle\psi|\sigma|\psi\rangle}.$$

Exercise 10 ([GÓ1]) *Relativized to elementary invertible σ, show that $\mathbf{d}(\rho|\sigma) =^+ \log S\rho$, where*

$$S = \sum_F \frac{\mathbf{m}(F)F}{\mathrm{Tr}F\sigma},$$

where F varies over all elementary density matrices.

7.1.1 Uncomputable Mixed States

We now extend $\mathbf{d}$ to uncomputable σ. For uncomputable σ, $\mathcal{T}_\sigma$ is not necessarily enumerable, and thus a universal lower computable randomness test does not necessarily exist, and cannot be used to define the σ deficiency of randomness. So in this case, the deficiency of randomness is instead defined using an aggregation of σ-tests, weighted by their lower algorithmic probabilities. The lower algorithmic probability of a lower computable matrix σ is $\underline{\mathbf{m}}(\sigma|x) = \sum\{\mathbf{m}(q|x) : q \text{ lower computes } \sigma\}$. Let $\mathfrak{T}_\sigma = \sum_{\nu\in\mathcal{T}_\sigma} \underline{\mathbf{m}}(\nu|n)\nu$.

Definition 15 (Quantum Randomness Deficiency of Uncomputable States) *The randomness deficiency of ρ with respect to σ is $\mathbf{d}(\rho|\sigma) = \log \mathrm{Tr}\mathfrak{T}_\sigma\rho$.*

If σ is computable, then Definition 15 equals Definition 14. By definition, $\mathfrak{T}_\sigma$ is universal, since for every lower computable σ-test ν, $\underline{\mathbf{m}}(\nu)\nu < \mathfrak{T}_\sigma$.

Theorem 42 *For n qubit density matrices σ, ρ, ν, and ξ, $\mathbf{d}(\sigma|\rho) + \mathbf{d}(\nu|\xi) <^+ \mathbf{d}(\sigma \otimes \nu|\rho \otimes \xi)$.*

Proof.

$$\begin{aligned}
\mathbf{d}(\sigma|\rho) + \mathbf{d}(\nu|\xi) &= \log \mathrm{Tr} \sum_{\rho' \in \mathcal{T}_\rho} \underline{\mathbf{m}}(\rho')\rho'\sigma + \log \mathrm{Tr} \sum_{\xi' \in \mathcal{T}_\xi} \underline{\mathbf{m}}(\xi')\xi'\nu \\
&= \log \mathrm{Tr} \left(\left(\sum_{\rho' \in \mathcal{T}_\rho} \underline{\mathbf{m}}(\rho')\rho' \right) \otimes \left(\sum_{\xi' \in \mathcal{T}_\xi} \underline{\mathbf{m}}(\xi')\xi' \right) \right) (\sigma \otimes \nu) \\
&= \log \mathrm{Tr} \left(\sum_{\rho' \in \mathcal{T}_\rho, \xi' \in \mathcal{T}_\xi} \underline{\mathbf{m}}(\rho')\underline{\mathbf{m}}(\xi')\rho' \otimes \xi' \right) (\sigma \otimes \nu) \\
&<^+ \log \mathrm{Tr} \left(\sum_{\rho' \in \mathcal{T}_\rho, \xi' \in \mathcal{T}_\xi} \underline{\mathbf{m}}(\rho' \otimes \xi')\rho' \otimes \xi' \right) (\sigma \otimes \nu) \\
&<^+ \log \mathrm{Tr} \left(\sum_{\kappa \in \mathcal{T}_{\rho \otimes \xi}} \underline{\mathbf{m}}(\kappa)\kappa \right) (\sigma \otimes \nu) \\
&=^+ \mathbf{d}(\sigma \otimes \nu|\rho \otimes \xi).
\end{aligned}$$

□

7.2 Conservation Over Quantum Operations

Conservation of randomness was introduced by L. A. Levin, culminating to [Lev84]. It states that deterministic or randomnized cannot increase the randomness deficiency of a finite or infinite sequence. In [Ǵ21], this was generalized to computable metric spaces, and is detailed in Chapter 20. The conservation statement for deterministic processes A is

$$\mathbf{d}(Ax|Ap) <^+ \mathbf{d}(x|p),$$

where $Ap(x) \sum_{y:A(y)=x} p(y)$. The additive constant is proportional to the Kolmogorov complexity of the probability p. Thus given a typical member x of a population, there is no method A that "to single out" this data point from the population. The only way for this to occur is for A to encode x. For example take a random uniform sample of n bit strings, at let x be random, with $\mathbf{K}(x) =^+ n$. Since this string is incompressible, there exist no simple means to separate it from the sample. This section shows conservation of randomness with respect to quantum mechanics. If a state is typical of a source mixed state, there is no physical means to transform it to be atypical.

Proposition 4 *For n qubit semi-density matrix ν, relativized to a finite set of elementary matrices $\{M_i\}$,*
$\underline{\mathbf{m}}(\sum_i M_i^* \nu M_i|n) \overset{*}{>} \underline{\mathbf{m}}(\nu|n)$.

Proof. For every string q that lower computes ν, there is a string q_M of the form rq, that lower computes $\sum_i M_i^* \nu M_i$. This string q_M uses the helper code r to take the intermediary outputs ξ_i of q and outputs the intermediary output $\sum_i M_i^* \xi_i M_i$. Since q_M has access to $\{M_i\}$ on the auxilliary tape, $\mathbf{m}(q_M|n) \overset{*}{>} \mathbf{m}(q|n)$.

$$\begin{aligned}
\underline{\mathbf{m}}(\nu|n) &= \sum \{\mathbf{m}(q|n) : q \text{ lower computes } \nu\} \\
&\overset{*}{<} \sum \{\mathbf{m}(q_M|n) : q \text{ lower computes } \nu\} \\
&\overset{*}{<} \sum \{\mathbf{m}(q'|n) : q' \text{ lower computes } \sum_i M_i^* \nu M_i\} \\
&\overset{*}{<} \underline{\mathbf{m}} \left(\sum_i M_i^* \nu M_i / n \right).
\end{aligned}$$

□

The following theorem shows conservation of randomness with respect to elementary quantum operations. It generalizes Theorems 2 and 3 from [Eps19c].

Theorem 43 (Randomness Conservation) *Relativized to elementary quantum operation ε, for semi-density matrices ρ, σ, $\mathbf{d}(\varepsilon(\rho)|\varepsilon(\sigma)) <^+ \mathbf{d}(\rho|\sigma)$.*

Proof. Since the universal Turing machine is relativized to ε, there is an elementary Kraus operator $\{M_i\}$ that can be computed from ε where $\varepsilon(\xi) = \sum_i M_i \xi M_i^*$. If ν is a $\sum_i M_i \rho M_i^*$-test, with $\nu \in \mathcal{T}_{\sum_i M_i \rho M_i^*}$, then $\sum_i M_i^* \nu M_i$ is a ρ-test, with $\sum_i M_i^* \nu M_i \in \mathcal{T}_\rho$. This is because by assumption $\mathrm{Tr}\nu \sum_i M_i \rho M_i^* \leq 1$. So by the cyclic property of trace $\mathrm{Tr} \sum_i M_i^* \nu M_i \rho \leq 1$. Therefore since $\sum_i M_i^* \nu M_i$ is lower computable, $\sum_i M_i^* \nu M_i \in \mathcal{T}_\rho$. From Proposition 4, $\underline{\mathbf{m}}(\sum_i M_i^* \nu M_i|n) \overset{*}{>} \underline{\mathbf{m}}(\nu|n)$. So we have the following inequality

$$\begin{aligned}
\mathbf{d}\left(\sum_i M_i \sigma M_i^* | \sum_i M_i \rho M_i^* \right) &= \log \sum_{\nu \in \mathcal{T}_{\sum_i M_i \rho M_i^*}} \underline{\mathbf{m}}(\nu|n) \mathrm{Tr}\nu \sum_i M_i \sigma M_i^* \\
&<^+ \log \sum_{\nu \in \mathcal{T}_{\sum_i M_i \rho M_i^*}} \underline{\mathbf{m}} \left(\sum_i M_i^* \nu M_i | n \right) \mathrm{Tr} \sum_i M_i^* \nu M_i \sigma \\
&<^+ \mathbf{d}(\sigma|\rho).
\end{aligned}$$

□

7.3 A Quantum Outlier Theorem

As discussed at the beginning of this chapter, outliers can be seen as generated from another process. In other words, outliers can be seen as product of errors or contamination of white noise into the generative model. However recent results [Eps21b] have proved a surprising consequence in statistics: that all sampling algorithms produce outliers. The longer the sampling method operates, the higher the outlier score of a datapoint that appears. In [Eps23b, Eps23c], this was generalized to all physical processes. Thus there is something intrinsic in statistics that results in the ubiquitous nature of anomalies. Note by outlier, we mean a finite or infinite sequence with high randomness deficiency.

These results can be derived from the fact that large sets of sequences with low randomness deficiency are exotic, in that they have high mutual information with the halting sequence. In this section we prove a quantum analog, in that a large collection of quantum states with low quantum deficiency also have high mutual information with the halting sequence. By "large collection" we mean the quantum projections of large rank. Such projections must have "outlier" states in their images otherwise they are exotic. Thus quantum coding schemes that use projections such ad Schumacher Compression must communicate using outlier quantum states. The classical and quantum theorems are analogous, but their proofs are very different!

Theorem 44 ([Eps23e]) *Relativized to an n qubit mixed state σ, for elementary 2^m rank projector P, $3m - 2n <^{\log} \max_{|\phi\rangle \in \text{Image}(P)} \mathbf{d}(|\phi\rangle |\sigma) + \mathbf{I}(\langle P\rangle; \mathcal{H})$.*

Proof. We relativize the universal Turing machine to $\langle\sigma\rangle$ and $(3m - 2n)$. Thus it is effectively relativized to m, n, and σ. Let elementary probability measure Q and $d \in \mathbb{N}$ realize $\mathbf{Ks}(P)$, where $d = \max\{\mathbf{d}(P|Q), 1\}$. Without loss of generality we can assume that the support of Q is elementary projections of rank 2^m. There are $d2^{n-m+2}$ rounds. For each round we select an σ-test T, that is of dimension 1, $\text{Tr}\sigma T \leq 1$, and for a certain Q-probability of projections B, $\text{Tr}TB$ is large. We now describe the selection process.

Select a random test T to be $2^{m-2} |\psi\rangle \langle\psi|$, where $|\psi\rangle$ is an n qubit state chosen uniformly from the unit sphere, with distribution Λ.

$$\mathbf{E}[\text{Tr}T\sigma] = 2^{m-2} \int \text{Tr} \langle\psi| \sigma |\psi\rangle \, d\Lambda = 2^{m-2}\text{Tr}\sigma \int |\psi\rangle \langle\psi| \, d\Lambda = 2^{m-n-2}\text{Tr}\sigma = 2^{m-n-2}.$$

Thus the probability that T is a σ-test is $\geq 1 - 2^{m-n-2}$. Let I_m be an n-qubit identity matrix with only the first 2^m diagonal elements being non-zero. Let $K_m = I - I_m$. Let $p = 2^{m-n}$ and $\hat{T} = T/2^{m-2}$. For any projection B of rank 2^m,

$$\begin{aligned}
&\Pr(\text{Tr}B\hat{T} \leq .5p) \\
=&\Pr(\text{Tr}I_m\hat{T} \leq .5p) \\
=&\Pr(\text{Tr}K_m\hat{T} \geq 1 - .5p) \\
\mathbf{E}[\text{Tr}K_m\hat{T}] &= 1 - p \\
\Pr(\text{Tr}K_m\hat{T} \geq 1 - .5p) &\leq (1 - p)/(1 - .5p) \\
\Pr(\text{Tr}B\hat{T} \geq .5p) &= 1 - \Pr(\text{Tr}K_m\hat{T} \geq 1 - .5p) \\
&\geq 1 - (1 - p)/(1 - .5p) \\
&= .5p/(1 - .5p) \geq .5p \\
\Pr(\text{Tr}BT \geq 2^{2m-n-3}) &\geq .5p.
\end{aligned}$$

Let Ω be the space of all matrices of the form $2^{m-2} |\phi\rangle \langle\phi|$. Let R be the uniform distribution over Ω. Let $[A, B]$ be 1 if $\text{Tr}AB > 2^{2m-n-3}$, and 0 otherwise. By the above equations, for all $A \in \text{Support}(Q)$, $\int_\Omega [A, B] dR(B) \geq .5p$. So $\sum_A \int_\Omega [A, B] Q(A) dR(B) \geq .5p$. For Hermitian matrix A, $\{A\}$ is 1 if $\text{Tr}A\sigma \leq 1$, and 0 otherwise. So $\int_\Omega \{A\} dR(A) \geq (1 - p2^{-2})$. Let $f = \max_T \{T\} \sum Q(A)[T, A]$.

So

$$
\begin{aligned}
.5p &\leq \sum_A \int_\Omega [A,B]Q(A)dR(B) \\
&= \sum_A \int_\Omega \{B\}Q[A,B](A)dR(B) + \sum_A \int_\Omega (1-\{B\})[A,B]Q(A)dR(B) \\
&\leq \sum_A \int_\Omega \{B\}[A,B]Q(A)dR(B) + \int_\Omega (1-\{B\})dR(B) \\
&\leq \sum_A \int_\Omega \{B\}[A,B]Q(A)dR(B) + p2^{-2} \\
p/4 &\leq \sum_A \int_\Omega \{B\}[A,B]Q(A)dR(B) = \int_\Omega \left(\{B\} \sum_A [A,B]Q(A) \right) dR(B) \leq \int_\Omega f dR(B) \\
p/4 &\leq f.
\end{aligned}
$$

Thus for each round i, the lower bounds on f proves there exists a one dimensional matrix $T_i = 2^{m-2} |\psi\rangle \langle\psi|$ such that $\mathrm{Tr} T_i \sigma \leq 1$ and $\sum_R \{Q(R) : \mathrm{Tr} T_i R \geq 2^{2m-n-3}\} \geq p/4 = 2^{m-n-2}$. Such a T_i is selected, and the the Q probability is conditioned on those projections B for which $[T_i, B] = 0$, and the next round starts. Assuming that there are $d2^{n-m+2}$ rounds, the Q measure of projections B such there does not exist a T_i with $[T_i, B] = 1$ is

$$\leq (1-p/4)^{d2^{n-m+2}} \leq e^{-d}.$$

Thus there exists a T_i such that $[T_i, P] = 1$, otherwise one can create a Q test t that assigns e^d to all projections B where there does not exist T_i with $[T_i, B] = 1$, and 0 otherwise. Then $t(P) = e^d$ so

$$1.44d < \log t(P) <^+ \mathbf{d}(P|Q,d) <^+ d + \mathbf{K}(d).$$

This is a contradiction, because without loss of generality, one can assume d is large. Let $T_i = 2^{m-2} |\psi\rangle \langle\psi|$ with $[T_i, P] = 1$. Let $|\phi\rangle = P|\psi\rangle / \sqrt{\langle\psi| P |\psi\rangle}$. So $\langle\phi| T_i |\phi\rangle \geq 2^{2m-n-3}$ and $|\phi\rangle$ is in the image of P. Thus by Lemma 3,

$$
\begin{aligned}
2m-n &<^+ \log \langle\phi| T_i |\phi\rangle \\
2m-n &<^+ \log \max_{|\phi\rangle \in \mathrm{Image}(P)} \langle\phi| T_i |\phi\rangle \\
2m-n &<^+ \max_{|\phi\rangle \in \mathrm{Image}(P)} \mathbf{d}(P|\sigma) + \mathbf{K}(T_i) \\
2m-n &<^+ \max_{|\phi\rangle \in \mathrm{Image}(P)} \mathbf{d}(P|\sigma) + (n-m) + \log d + \mathbf{K}(d) + \mathbf{K}(Q) \\
2m-n &<^+ \max_{|\phi\rangle \in \mathrm{Image}(P)} \mathbf{d}(P|\sigma) + (n-m) + \mathbf{Ks}(P) \\
3m-2n &<^{\log} \max_{|\phi\rangle \in \mathrm{Image}(P)} \mathbf{d}(P|\sigma) + \mathbf{I}(P;\mathcal{H}).
\end{aligned}
$$

Note that due to the fact that the left hand side of the equation is $(3m-2n)$ and it has log precision, this enables one to condition the universal Turing machine to $(3m-2n)$. □

7.3.1 Computable Projections

Theorem 44 is in terms of elementary described projecctions and can be generalized to arbitrarily computable projections. For a matrix M, let $\|M\| = \max_{i,j} |M_{i,j}|$ be the max norm. A program

$p \in \{0,1\}^*$ computes a projection P of rank ℓ if it outputs a series of rank ℓ projections $\{P_i\}_{i=1}^{\infty}$ such that $\|P - P_i\| \leq 2^{-i}$. For computable projection operator P, $\mathbf{I}(P;\mathcal{H}) = \min\{\mathbf{K}(p) - \mathbf{K}(p|\mathcal{H}) : p \text{ is a program that computes } P\}$.

Corollary 11 ([Eps23e]) *Relativized to an n qubit mixed state σ, for computable 2^m rank projector P, $3m - 2n <^{\log} \max_{|\phi\rangle \in \text{Image}(P)} \mathbf{d}(|\phi\rangle\,|\sigma) + \mathbf{I}(\langle P\rangle;\mathcal{H})$.*

Proof. Let p be a program that computes P. There is a simply defined algorithm A, that when given p and σ, outputs P_n such that $\max_{|\psi\rangle \in \text{Image}(P)} \mathbf{d}(|\psi\rangle\,|\sigma) =^{+} \max_{|\psi\rangle \in \text{Image}(P_n)} \mathbf{d}(|\psi\rangle\,|\sigma)$. Thus by Lemma 1, one gets that $\mathbf{I}(P_n;\mathcal{H}) <^{+} \mathbf{I}(P;\mathcal{H})$. The corollary follows from Theorem 44. □

Chapter 8

Quantum Information

For a pair of random variables, $\mathcal{X}$, $\mathcal{Y}$, their mutual information is defined to be $I(\mathcal{X} : \mathcal{Y}) = \mathcal{H}(\mathcal{X}) + \mathcal{H}(\mathcal{Y}) - \mathcal{H}(\mathcal{X}, \mathcal{Y}) = \mathcal{H}(\mathcal{X}) - \mathcal{H}(\mathcal{X}/\mathcal{Y}) = \sum_{x,y} p(x, y) \log p(x, y)/p(x)p(y)$. This represents the amount of correlation between $\mathcal{X}$ and $\mathcal{Y}$. It is the amount of information, in bits, obtained about $\mathcal{Y}$ when observing $\mathcal{X}$. Another intrepretation is that the mutual information between $\mathcal{X}$ and $\mathcal{Y}$ is the reduction in uncertainty of $\mathcal{X}$ after being given access to $\mathcal{Y}$. If the two variables are independent, then their mutual information is zero. Another property of mutual information is that deterministic (or randomized) processing cannot increase information, with

$$\mathbf{I}(\mathcal{X} : f(\mathcal{Y})) \leq \mathbf{I}(\mathcal{X} : \mathcal{Y}).$$

Quantum mutual information between two subsystems described by states ρ_A and ρ_B of a composite system described by a joint state ρ_{AB} is $I(A : B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB})$, where S is the Von Neumman entropy. Quantum mutual information measures the correlation between two quantum states. It is defined as the amount of work that is required to erase the correlations completely. Quantum mutual information is monotonic under quantum operations, with

$$I(A : \epsilon(B)) \leq I(A : B).$$

As stated in Chapter 4, The algorithmic information between two strings is defined to be $\mathbf{I}(x : y) = \mathbf{K}(x) + \mathbf{K}(y) - \mathbf{K}(x, y)$. By definition, it measures the amount of compression two strings achieve when grouped together. A particular important property of $\mathbf{I}$ is that observes information non growth properties ([Lev84]),

$$\mathbf{I}(x : f(y)) <^{+} \mathbf{I}(x : y).$$

In this chapter we show that, like the above definitions, algorithmic mutual information of quantum states $\mathbf{I}$ obeys processing non-growth laws, with respect to quantum channels, ε, with

$$\mathbf{I}(\sigma : \varepsilon(\rho)) <^{+} \mathbf{I}(\sigma : \rho).$$

However there does exist properties of $\mathbf{I}$ which do not hold in the classical setting. In particular most pure and mixed quantum states contain no self information. This is a consequence of the vastness of high dimensional Hilbert spaces. This has far reaching applications, including that most measurements produce white noise, as detailed in Chapter 9.

8.1 Definition of Quantum Algorithmic Information

The three definitions above are based off the difference between a joint aggregate and the separate parts. Another approach is to define information between two semi-density matrices as the deficiency of randomness over $\boldsymbol{\mu} \otimes \boldsymbol{\mu}$, with the mutual information of σ and ρ being $\mathbf{d}(\sigma \otimes \rho | \boldsymbol{\mu} \otimes \boldsymbol{\mu})$.

This is a counter argument for the hypothesis that the states are independently chosen according to the universal semi-density matrix $\boldsymbol{\mu}$. This parallels the classical algorithmic case, where $\mathbf{I}(x : y) =^+ \mathbf{d}((x,y)|\mathbf{m} \otimes \mathbf{m}) =^+ \mathbf{K}(x) + \mathbf{K}(y) - \mathbf{K}(x,y)$. However to achieve the conservation inequalities, a further refinement is needed, with the restriction of the form of the $\boldsymbol{\mu} \otimes \boldsymbol{\mu}$ tests. Let $\mathcal{C}_{C\otimes D}$ be the set of all lower computable matrices $A \otimes B$, such that $\mathrm{Tr}(A \otimes B)(C \otimes D) \leq 1$. Let $\mathfrak{C}_{C\otimes D} = \sum_{A\otimes B \in \mathcal{C}_{C\otimes D}} \underline{\mathbf{m}}(A \otimes B|n) A \otimes B$.

Definition 16 (Information) *The mutual information between two semi-density matrices σ, ρ is defined to be $\mathbf{I}(\sigma : \rho) = \log \mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(\sigma \otimes \rho)$.*

Up to an additive constant, information is symmetric.

Theorem 45 $\mathbf{I}(\sigma : \rho) =^+ \mathbf{I}(\rho : \sigma)$.

Proof. This follows from the fact that for every $A \otimes B \in \mathcal{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}$, the matrix $B \otimes A \in \mathcal{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}$. Furthermore, since $\underline{\mathbf{m}}(A \otimes B|n) \stackrel{*}{=} \underline{\mathbf{m}}(B \otimes A|n)$, this guarantees that $\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(\sigma\otimes\rho) \stackrel{*}{=} \mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(\rho\otimes\sigma)$, thus proving the theorem. □

Exercise 11 *Show that relativized to an orthogonal sequence of elementary states $|1\rangle, |2\rangle, |3\rangle, \ldots$, enumerated by strings $i, j, k \in \{0,1\}^n$,*

1. $\mathbf{I}(|k\rangle : |i\rangle) <^+ \mathbf{I}(|j\rangle : |i\rangle) + \mathbf{K}(k|j,n)$.
2. $\mathbf{I}(i : j|n) <^{\log} \mathbf{I}(|i\rangle : |j\rangle)$.
3. $\mathbf{I}(|i\rangle : |j\rangle) <^+ \mathbf{I}(i : j|n) + \mathbf{I}(i, j : \chi|n)$.
4. $\mathbf{K}(i|n) <^+ \mathbf{I}(|i\rangle : |i\rangle) <^+ \mathbf{K}(i|n) + \mathbf{I}(i : \chi|n)$.
5. $\mathbf{I}(|i\rangle : |i\rangle) <^+ 4n/3$.
6. $\mathbf{I}(|i\rangle : |j\rangle) <^+ \mathbf{I}(|i\rangle : |i\rangle) + \mathbf{I}(i, j : \chi|n)$.

8.2 Paucity of Self-Information

All strings of high Kolmogorov complexity have high self information, with $\mathbf{I}(x : x) =^+ \mathbf{K}(x)$. However the situation is much different in the quantum world, with respect to the definition of mutual information of quantum mixed states σ and ρ introduced in this chapter: $\mathbf{I}(\sigma : \rho)$. In this section we show that most pure and mixed states have low self information.

8.2.1 Pure States

Almost all pure states $|\psi\rangle$ have low $\mathbf{I}(|\psi\rangle : |\psi\rangle)$. In fact this is the case for most quantum states, where for most n qubit pure states $|\psi\rangle$,

$$\mathbf{Hg}(|\psi\rangle) \approx n, \quad \mathbf{I}(|\psi\rangle : |\psi\rangle) \approx 0.$$

This has to do with the huge expanse of high dimensional Hilbert spaces versus the discretionary power of $\boldsymbol{\mu} \times \boldsymbol{\mu}$ tests. The following theorem states that the information between two elementary states is not more than the combined length of their descriptions.

Theorem 46 *For elementary ρ and σ, $\mathbf{I}(\rho : \sigma) <^+ \mathbf{K}(\rho|n) + \mathbf{K}(\sigma|n)$.*

Proof. Assume not. Then for any positive constant c, there exists semi-density matrices ρ and σ, such that

$$c\mathbf{m}(\rho|n)\mathbf{m}(\sigma|n)2^{\mathbf{I}(\rho:\sigma)} = c\text{Tr}\mathbf{m}(\rho|n)\mathbf{m}(\sigma|n)\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(\rho\otimes\sigma) > 1.$$

By the definition of $\boldsymbol{\mu}$, $\mathbf{m}(\rho|n)\rho \overset{*}{<} \boldsymbol{\mu}$ and $\mathbf{m}(\sigma|n)\sigma \overset{*}{<} \boldsymbol{\mu}$. Therefore by the definition of the Kronecker product, there is some positive constant d such that for all ρ and σ, $d\mathbf{m}(\rho|n)\mathbf{m}(\sigma|n)(\rho\otimes\sigma) < (\boldsymbol{\mu}\otimes\boldsymbol{\mu})$, and similarly

$$d\text{Tr}\mathbf{m}(\rho|n)\mathbf{m}(\sigma|n)\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(\rho\otimes\sigma) < \text{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(\boldsymbol{\mu}\otimes\boldsymbol{\mu}).$$

By the definition of $\mathfrak{C}$, it must be that $\text{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}\boldsymbol{\mu}\otimes\boldsymbol{\mu} \leq 1$. However for $c = d$, there exists a ρ and a σ, such that

$$\text{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}\boldsymbol{\mu}\otimes\boldsymbol{\mu} > d\text{Tr}\mathbf{m}(\rho|n)\mathbf{m}(\sigma|n)\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(\rho\otimes\sigma) > 1,$$

causing a contradiction. □

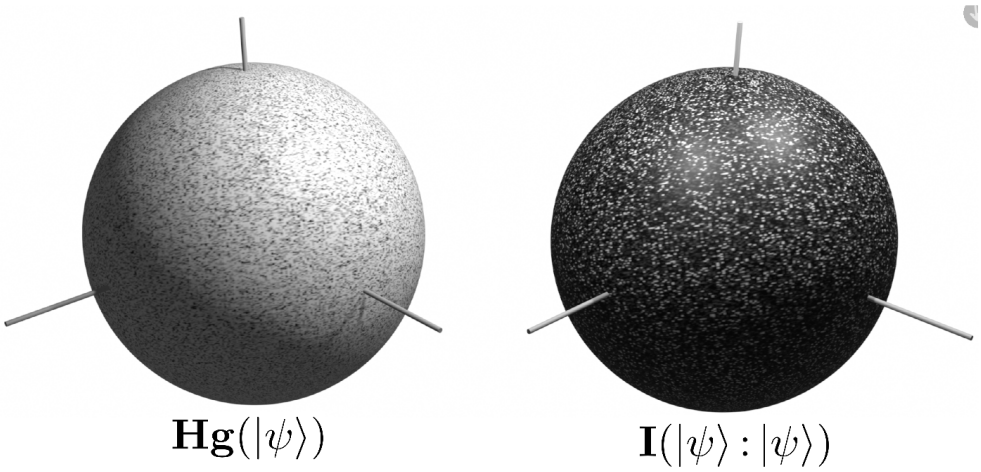


Figure 8.1: The left figure shows the uniform distribution unit sphere over n qubit space and the associated $\mathbf{Hg}(|\phi\rangle)$ values, which are overwhelmingly at the max n. The right figure shows the $\mathbf{I}(|\psi\rangle : |\psi\rangle)$ score over the same distribution. Most states have negligible self information.

The following theorem shows that $\mathbf{Hg}(|\psi\rangle)$ is high for most states and $\mathbf{I}(|\psi\rangle : |\psi\rangle)$ is low, as shown in Figure 8.1.

Theorem 47 ([Eps19b]) *Let Λ be the uniform distribution on the unit sphere of $\mathcal{H}_N$, where $N = 2^n$.*

(1) $\mathbf{Hg}(I/N) =^+ n$,

(2) $\mathbf{I}(I/N : I/N) <^+ 0$,

(3) $\int 2^{-\mathbf{Hg}(|\psi\rangle)} d\Lambda \overset{*}{=} N^{-1}$,

(4) $\int 2^{\mathbf{I}(|\psi\rangle:|\psi\rangle)} d\Lambda <^+ 0$.

Proof.

(1) follows from $\mathbf{Hg}(I/N) =^+ -\log \mathrm{Tr}\boldsymbol{\mu} I/N =^+ \log N - \log \mathrm{Tr}\boldsymbol{\mu} =^+ \log N$.

(2) This is due to Theorem 46, with $\mathbf{I}(I/N : I/N) <^+ 2\mathbf{K}(I/N|n) <^+ 0$.

(3) We use the fact that $\rho = \int |\psi\rangle \langle\psi| \, d\Lambda = I/N$, because $\mathrm{Tr}\rho = 1$, and $\langle\psi| \rho |\psi\rangle = \langle\phi| \rho |\phi\rangle$. Thus $\int 2^{-\mathbf{Hg}(|\psi\rangle)} d\Lambda \stackrel{*}{=} \int \mathrm{Tr}\boldsymbol{\mu} |\psi\rangle \langle\psi| \, d\Lambda \stackrel{*}{=} \mathrm{Tr}\boldsymbol{\mu} \int |\psi\rangle \langle\psi| \, d\Lambda \stackrel{*}{=} N^{-1}$.

(4) We use the proof of Theorem 18, which states $\int |\psi\rangle\langle\psi| \otimes |\psi\rangle \langle\psi| \, d\Lambda = \int |\psi\psi\rangle\langle\psi\psi| \, d\Lambda = \binom{N+1}{2}^{-1} P$, where P is the projection onto the space of pure states $|\psi\psi\rangle$. So

$$\begin{aligned}
\int 2^{\mathbf{I}(|\psi\rangle : |\psi\rangle)} d\Lambda &= \int \mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} |\psi\rangle\langle\psi| \otimes |\psi\rangle \langle\psi| \, d\Lambda \\
&= \mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \int |\psi\rangle\langle\psi| \otimes |\psi\rangle \langle\psi| \, d\Lambda \\
&= \mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \binom{N+1}{2}^{-1} P \\
&\stackrel{*}{<} \mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} N^{-2} I \\
&\stackrel{*}{=} 2^{\mathbf{I}(I/N : I/N)} \\
&<^+ 0.
\end{aligned}$$

□

8.2.2 Mixed States

The results of the previous section can be extended to mixed states. Given a uniform measure over mixed states, an overwhelming majority of such states contain no algorithmic self information. Let Λ be the uniform distribution of the unit sphere of $\mathcal{H}_N$, where $N = 2^n$.

Definition 17 (Uniform Distribution over Mixed States) *Fix any number $M \in \mathbb{N}$. Let the M-simplex be*

$$\Delta_M = \{(p_i)_{1 \leq i \leq M} | p_i \geq 0, p_1 + \cdots + p_M = 1\}.$$

Let η be any distribution over Δ_M. Let

$$\mu\left(\sum_{i=1}^{M} p_i |\psi_i\rangle \langle\psi_i|\right) = \eta(p_1, \ldots, p_M) \prod_{i=1}^{M} \Lambda(|\psi_i\rangle),$$

Theorem 48 $\int 2^{\mathbf{I}(\sigma:\sigma)} d\mu(\sigma) <^+ 0$.

Proof.

$$
\begin{aligned}
&\int 2^{\mathbf{I}(\sigma:\sigma)} d\mu(\sigma) \\
=&\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \int_{\Delta_M}\int_{\Lambda_1}\cdots\int_{\Lambda_M}\left(\sum_{i=1}^{M} p_i \left|\psi_i\right\rangle\left\langle\psi_i\right|\right)\otimes\left(\sum_{i=1}^{M} p_i \left|\psi_i\right\rangle\left\langle\psi_i\right|\right) d\Lambda_1\ldots d\Lambda_M d\eta(p_1,\ldots,p_M) \\
=&\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \int_{\Delta_M}\int_{\Lambda_1}\cdots\int_{\Lambda_M}\left(\sum_{i,j=1}^{M} p_i p_j \left|\psi_i\right\rangle\left\langle\psi_i\right|\otimes\left|\psi_j\right\rangle\left\langle\psi_j\right|\right) d\Lambda_1\ldots d\Lambda_M d\eta(p_1,\ldots,p_M) \\
=&\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \int_{\Delta_M}\int_{\Lambda}\sum_{i=1}^{M} p_i^2 \left|\psi\psi\right\rangle\left\langle\psi\psi\right| d\Lambda d\eta(p_1,\ldots,p_M) \\
&+\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \int_{\Delta_M}\int_{\Lambda_1}\int_{\Lambda_2}\sum_{i,j\in\{1,\ldots,M\},i\neq j} 2p_i p_j \left|\psi_1\right\rangle\left\langle\psi_1\right|\otimes\left|\psi_2\right\rangle\left\langle\psi_2\right| d\Lambda_1 d\Lambda_2 d\eta(p_1,\ldots,p_M).
\end{aligned}
$$

The first term is not greater than

$$
\begin{aligned}
&\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \int_{\Delta_M}\int_{\Lambda}\sum_{i=1}^{M} p_i \left|\psi\psi\right\rangle\left\langle\psi\psi\right| d\Lambda d\eta(p_1,\ldots,p_M) \\
=&\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \int_{\Delta_M}\sum_{i=1}^{M} p_i \left(\int_{\Lambda}\left|\psi\psi\right\rangle\left\langle\psi\psi\right| d\Lambda\right) d\eta(p_1,\ldots,p_M) \\
=&\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \int_{\Lambda}\left|\psi\psi\right\rangle\left\langle\psi\psi\right| d\Lambda.
\end{aligned}
$$

At this point, reasoning from the proof of Theorem 47 can be used to show that this term is $O(1)$. The second term is not greater than

$$
\begin{aligned}
&\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \int_{\Delta_M}\int_{\Lambda_1}\int_{\Lambda_2}\left(\sum_i p_i\right)\left(\sum_i p_i\right)\left|\psi_1\right\rangle\left\langle\psi_1\right|\otimes\left|\psi_2\right\rangle\left\langle\psi_2\right| d\Lambda_1 d\Lambda_2 d\eta(p_1,\ldots,p_M) \\
=&\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \int_{\Delta_M}\int_{\Lambda_1}\int_{\Lambda_2}\left|\psi_1\right\rangle\left\langle\psi_1\right|\otimes\left|\psi_2\right\rangle\left\langle\psi_2\right| d\Lambda_1 d\Lambda_2 d\eta(p_1,\ldots,p_M) \\
=&\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}} \int_{\Lambda_1}\int_{\Lambda_2}\left|\psi_1\right\rangle\left\langle\psi_1\right|\otimes\left|\psi_2\right\rangle\left\langle\psi_2\right| d\Lambda_1 d\Lambda_2 \\
=&\mathrm{Tr}\mathfrak{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(I/N\otimes I/N).
\end{aligned}
$$

Again, at this point, reasoning from the proof of Theorem 47 can be used to show that this term is $O(1)$.

8.3 Information Nongrowth

Classical algorithmic information non-growth laws asserts that the information between two strings cannot be increased by more than a constant depending on the computable transform f, with $\mathbf{I}(f(x) : y) < \mathbf{I}(x : y) + O_f(1)$ (Theorem 2). Conservation inequalities have been extended to probabilistic transforms, infinite sequences and points in computable metric spaces. The following theorem shows information non-growth in the quantum case; information cannot increase under quantum operations, the most general type of transformation that a mixed or pure quantum state

can undergo. The following theorem shows information nongrowth with respect to elementary quantum operations. It generalizes Theorems 5 and 10 from [Eps19c].

Theorem 49 (Information Conservation) *Relativized to elementary quantum operation ε, for semi-density matrices ρ, σ, $\mathbf{I}(\varepsilon(\rho) : \sigma) <^+ \mathbf{I}(\rho : \sigma)$.*

Proof. Since the universal Turing machine is relativized to ε, there is an elementary Kraus operator $\{M_i\}$ that can be computed from ε where $\varepsilon(\xi) = \sum_i M_i \xi M_i^*$. Given density matrices A, B, C and D, we define $\mathbf{d}'(A \otimes B|C \otimes D) = \log \mathfrak{C}_{C\otimes D} A \otimes B$. Thus $\mathbf{I}(\sigma : \rho) = \mathbf{d}'(\sigma \otimes \rho|\boldsymbol{\mu} \otimes \boldsymbol{\mu})$. The semi-density matrix $\sum_i M_i \boldsymbol{\mu} M_i^*$ is lower semicomputable, so therefore $\sum_i M_i \boldsymbol{\mu} M_i^* \stackrel{*}{<} \boldsymbol{\mu}$ and also $(\sum_i M_i \boldsymbol{\mu} M_i^* \otimes \boldsymbol{\mu}) \stackrel{*}{<} \boldsymbol{\mu} \otimes \boldsymbol{\mu}$. So if $E \otimes F \in \mathcal{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}$ then $\mathrm{Tr}(E \otimes F)(\boldsymbol{\mu} \otimes \boldsymbol{\mu}) \leq 1$, implying that $\mathrm{Tr}(E \otimes F)(\sum_i M_i \boldsymbol{\mu} M_i^* \otimes \boldsymbol{\mu}) < O(1)$. Thus there is a positive constant c, where $c(E \otimes F) \in \mathcal{C}_{(\sum_i M_i \boldsymbol{\mu} M_i^*)\otimes\boldsymbol{\mu}}$. So we have

$$\begin{aligned}
\mathbf{d}'\left(\sum_i M_i \sigma M_i^* \otimes \rho|\boldsymbol{\mu} \otimes \boldsymbol{\mu}\right) &= \log \sum_{E\otimes F \in \mathcal{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}} \underline{\mathbf{m}}(E \otimes F|n)\mathrm{Tr}(E \otimes F)(\sum_i M_i \sigma M_i^* \otimes \rho) \\
&<^+ \log \sum_{E\otimes F \in \mathcal{C}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}} \underline{\mathbf{m}}(c(E \otimes F)|n)\mathrm{Tr}c(E \otimes F)\left(\sum_i M_i \sigma M_i^* \otimes \rho\right) \\
&<^+ \mathbf{d}'\left(\sum_i M_i \sigma {M_i}^* \otimes \rho|\sum_i M_i \boldsymbol{\mu} M_i^* \otimes \boldsymbol{\mu}\right).
\end{aligned}$$

Using the reasoning of the proof of Theorem 43 on the elementary Kraus operator $\{M_i \otimes I\}$ and $\mathbf{d}'$, where $\mathcal{C}$ replaces $\mathcal{T}$, we have that

$$\mathbf{d}'\left(\sum_i M_i \sigma M_i^* \otimes \rho|\sum_i M_i \boldsymbol{\mu} M_i^* \otimes \boldsymbol{\mu}\right) <^+ \mathbf{d}'(\sigma \otimes \rho|\boldsymbol{\mu} \otimes \boldsymbol{\mu}).$$

Therefore we have that

$$\begin{aligned}
\mathbf{I}\left(\sum_i M_i \sigma M_i^* : \rho\right) &= \mathbf{d}'\left(\sum_i M_i \sigma M_i^* \otimes \rho|\boldsymbol{\mu} \otimes \boldsymbol{\mu}\right) \\
&<^+ \mathbf{d}'\left(\sum_i M_i \sigma M_i^* \otimes \rho|\sum_i M_i \boldsymbol{\mu} M_i^* \otimes \boldsymbol{\mu}\right) \\
&<^+ \mathbf{d}'(\sigma \otimes \rho|\boldsymbol{\mu} \otimes \boldsymbol{\mu}) =^+ \mathbf{I}(\sigma : \rho).
\end{aligned}$$

□

8.4 Algorithmic No-Cloning Theorem

The no-cloning theorem states that every unitary transform cannot clone an arbitrary quantum state. However some unitary transforms can clone a subset of pure quantum states. For example, given basis states $|1\rangle, |2\rangle, |3\rangle, \ldots$ there is a unitary transform that transforms each $|i\rangle|0\rangle$ to $|i\rangle|i\rangle$. In addition, there exists several generalizations to the no-cloning theorem, showing that imperfect clones can be made. In [BH96], a universal cloning machine was introduced that can clone an

arbitrary state with the fidelity of 5/6. Theorem 18 shows a generalization of the no-cloning theorem using Gács complexity.

Given the information function introduced in this chapter, a natural question to pose is whether a considerable portion of pure states can use a unitary transform to produce two states that share a large amount of shared information. The following theorem answers this question in the negative. It states that the amount of information created between states with a unitary transform is bounded by the self information of the original state.

Theorem 50 ([Eps19b]) *Let $C|\psi\rangle|0^n\rangle = |\phi\rangle|\varphi\rangle$, where C is an elementary unitary transform. Relativized to C, $\mathbf{I}(|\phi\rangle : |\varphi\rangle) <^+ \mathbf{I}(|\psi\rangle : |\psi\rangle)$.*

Proof. We have the inequalities

$$\mathbf{I}(|\phi\rangle : |\varphi\rangle) <^+ \mathbf{I}(|\phi\rangle|\varphi\rangle) : |\phi\rangle|\varphi\rangle) <^+ \mathbf{I}(|\psi\rangle|0^n\rangle : |\psi\rangle|0^n\rangle) <^+ \mathbf{I}(|\psi\rangle : |\psi\rangle),$$

where the first inequality is derived using partial trace, the second inequality is derived using the unitary transform C, and the third inequality is derived by appending of an environment, all constituting quantum operations, whose conservation of information is proven in Theoerem 49. □

Theorem 50, combined with the paucity of self-information in pure states (Theorem 47) shows that only a very sparse set of pure states can, given any unitary transform, can duplicate algorithmic information.

8.5 Purification

Every mixed state can be considered a reduced state of a pure state. The purification process is considered physical, so the extended Hilbert space in which the purified state resides in can be considered the existing environment. It should therefore be possible to regard our system with its mixed state as part of a larger system in a pure state. In this section we proof that the purifications of two mixed states will contain more information than the reduced states.

Purification occurs in the following manner, starting with a density matrix $\rho = \sum_{i=1}^{n} p_i |i\rangle\langle i|$. A copy of the space is defined with orthonormal basis $\{|i'\rangle\}$. In this instance the purification of ρ is $|\psi\rangle = \sum_{i=1}^{n} \sqrt{p_i}|i\rangle \otimes |i'\rangle$. For a density matrix ρ of size n, let $\mathcal{P}_\rho^m$ be the set of purifications of ρ of dimension $m \geq 2n$.

Corollary 12 *For all $|\psi_\sigma\rangle \in \mathcal{P}_\sigma^n$, $|\psi_\rho\rangle \in \mathcal{P}_\rho^n$, $\mathbf{d}(\sigma|\rho) <^+ \mathbf{d}(|\psi_\sigma\rangle \,|\, |\psi_\rho\rangle)$.*

Corollary 13 *For all $|\psi_\sigma\rangle \in \mathcal{P}_\sigma^n$, $|\psi_\rho\rangle \in \mathcal{P}_\rho^n$, $\mathbf{I}(\sigma : \rho) <^+ \mathbf{I}(|\psi_\sigma\rangle : |\psi_\rho\rangle)$.*

This all follows from conservation of randomness (Theorem 43) and information (Theorem 49) over quantum operations, which includes the partial trace function.

Chapter 9

Quantum Measurements

In quantum mechanics, the measurement postulate states that the measured valued obtained will be the eigenvalue of the observation operator. After the measurement is made, the wave function collapses to the eigenspace associated with the eigenvalue. The probability of seeing a particular eigenvalue is proportionate to how much the wavefunction "overlaps" with the corresponding eigenspace.

There is an inconsistency associated with observations in quantum mechanics. Whereas the evolution of the wave function is deterministic (via a unitary transform), the wave function collapse is a probabilistic operation. Furthermore, the exact time of the observation is not clear.

The Copenhagen interpretation scedes completeness, demarcating a quantum domain and an "observer's" domain. Interaction between the two regions results in wavefunction collapses and transference of information. Another interpretation is the Many Worlds Theory, detailed in Chapter **??**.

A central research is how to quantify the amount of information that an observer can obtain. Limitations were discovered during the founding days of quantum mechanics. The Heisenberg uncertainty principle states that an accurate position measurement of a particle will result in a lot of uncertainty about its momentum, and vice versa.

This chapter deals with how to quantify the information that is acquired in measurements using Algorithmic Information Theory. Surprisingly, we show that given an observable operator, for an overwhelming majority of quantum states, white noise (or the empty signal) is produced.

This is one of central results of the manuscript. It indicates future research possibilities of how to quantify information (or the lack thereof) in quantum mechanics. For example, one could adapt these results to quantify the information content of measurements in quantum field theory

9.1 Definition of Measurements

In quantum mechanics, measurements are modeled by POVMs, which stands for positive operator valued measure. A POVM E is a finite or infinite set of positive definite matrices $\{E_k\}$ such that $\sum_k E_k = \mathbf{1}$. For a given semi-density matrix σ, a POVM E induces a semi measure over integers, where $E\sigma(k) = \mathrm{Tr}E_k\sigma$. This can be seen as the probability of seeing measurement k given quantum state σ and measurement E. An elementary POVM E has a program q such that $U(q)$ outputs an enumeration of $\{E_k\}$, where each E_k is elementary. A quantum instrument with respect to POVM E, is a quantum operation Φ_E that takes a state σ to a set of outcomes and their probabilities, $\Phi_E(\sigma) = \sum_k E(\sigma(k)) \left|k\right\rangle \left\langle k\right|$.

9.2 Typicality and Measurements

Theorem 51 shows that measurements can increase only up to a constant factor, the deficiency of randomness of a quantum state with respect to another quantum state. The classical deficiency of randomness of a probability with respect to a another probability is denoted as follows. This definition is well known in the literature, one appearance is

Definition 18 (Deficiency, probabilities (Folklore)) *For probabilities p and q over $\{0,1\}^\infty$, $\mathbf{d}(q|p) = \log \sum_x q(x)\mathbf{m}(x)/p(x)$.*

Note that in the following theorem, $\mathbf{d}(E\sigma|E\rho)$ term represents the classical deficiency of randomness of a semimeasure $E\sigma$ with respect to a computable probability measure $E\rho$. The term $\mathbf{d}(\sigma|\rho)$ is from Definition 15.

Theorem 51 ([Eps19b]) *For density matrices σ, ρ, relativized to elementary ρ and POVM E, $\mathbf{d}(E\sigma|E\rho) <^+ \mathbf{d}(\sigma|\rho)$.*

Proof. $2^{\mathbf{d}(E\sigma|E\rho)} = \sum_k (\mathrm{Tr}E_k\sigma)\mathbf{m}(k|n)/(\mathrm{Tr}E_k\rho) = \mathrm{Tr}(\sum_k(\mathbf{m}(k|n)/\mathrm{Tr}E_k\rho)E_k)\sigma = \mathrm{Tr}\nu\sigma$, where the matrix $\nu = (\sum_k(\mathbf{m}(k|n)/\mathrm{Tr}E_k\rho)E_k)$ has $\nu \in \mathcal{T}_\rho$, since ν is lower computable and $\mathrm{Tr}\nu\rho \leq 1$. So $2^{\mathbf{d}(\sigma|\rho)} \geq \underline{\mathbf{m}}(\nu|n)\mathrm{Tr}\nu\sigma = \underline{\mathbf{m}}(\nu|n)2^{\mathbf{d}(E\sigma|E\rho)}$. Since $\underline{\mathbf{m}}(\nu|n) \stackrel{*}{>} 1$, $\mathbf{d}(E\sigma|E\rho) <^+ \mathbf{d}(\sigma|\rho)$.

9.3 Information and Measurements

Given two mixed states σ and ρ and POVM E, the mutual information between the probabilities of $E\sigma$ and $E\rho$, from Definition 3, is $\mathbf{I}_{\mathrm{Prob}}(E\sigma : E\rho)$. The following theorem states that given two states, the classical (algorithmic) information between the probabilities generated by two quantum measurements is less, up to a logarithmic factor, than the information of the two states. Thus $\mathbf{I}$ represents an upper bound on the amount of classical algorithmic information that can be extracted between two states.

Theorem 52 *Relative to POVMS E and F, $\mathbf{I}_{\mathrm{Prob}}(E\sigma : F\rho) <^{\log} \mathbf{I}(\sigma : \rho)$.*

Note than since the universal Turing machine is relativized to E and F, all $\mathbf{K}$ and $\mathbf{m}$ are conditioned to the number of qubits n. Quantum instruments with respect to POVMs E and F produces two mixed states $\Psi_E(\sigma) = \sum_{i=1}^m E_i(\sigma)\,|i\rangle\langle i|$ and $\Psi_F(\rho) = \sum_{j=1}^m F_j(\rho)\,|j\rangle\langle j|$, where, without loss of generality, m can be considered a power of 2. By Theorem 8, the (i,i)th entry of $\boldsymbol{\mu}$ is $\stackrel{*}{=} \mathbf{m}(i)$, so $\mathcal{T}_{ij} = 2^{\mathbf{K}(i)+\mathbf{K}(j)-O(1)}\,|i\rangle\langle i|\,|j\rangle\langle j|$ is a $\boldsymbol{\mu}\otimes\boldsymbol{\mu}$ test, with $\mathrm{Tr}\mathcal{T}_{i,j}(\boldsymbol{\mu}\otimes\boldsymbol{\mu}) < 1$. So, using the fact that

$x/\log x$ is convex, and conservation of information Theorem 49,

$$\begin{aligned}
\mathbf{I}(\sigma:\rho) &>^+ \mathbf{I}(\Psi_E(\sigma):\Psi_F(\rho)) \\
&>^+ \log \sum_{i,j} \mathbf{m}(\mathcal{T}_{i,j})\mathcal{T}_{i,j}\Psi_E(\sigma)\otimes\Psi_F(\rho) \\
&>^+ \log \sum_{i,j} 2^{\mathbf{K}(i)+\mathbf{K}(j)}\mathbf{m}(i,j,\mathbf{K}(i)+\mathbf{K}(j))E_i(\sigma)F_j(\rho) \\
&>^+ \log \sum_{i,j} 2^{\mathbf{I}(i:j)-\mathbf{K}(\mathbf{I}(i:j))}E_i(\sigma)F_j(\rho) \\
&>^+ \log \sum_{i,j} 2^{\mathbf{I}(i:j)}\mathbf{I}(i:j)^{-O(1)}E_i(\sigma)F_j(\rho) \\
&>^{\log} \log \sum_{i,j} 2^{\mathbf{I}(i:j)}E_i(\sigma)F_j(\rho) \\
&>^{\log} \mathbf{I}_{\mathrm{Prob}}(E\sigma:F\rho).
\end{aligned}$$

Exercise 12 *For density matrices ρ and σ, and $i,j\in\mathbb{N}$, relativized to POVMS E and F, show that* $\mathbf{I}(i:j)+\log E_i(\rho)F_j(\sigma) <^{\log} \mathbf{I}(\rho:\sigma)$.

9.4 Algorithmic Contents of Measurements

This sections shows the limitations of the algorithmic content of measurements of pure quantum states. Theorem 53 says that given a measurement apparatus E, the overwhelming majority of pure states, when measured, will produce classical probabilities with no self-information, i.e. random noise. Theorem 3 shows that there is no randomized way to process the probabilities to produce more self-information, i.e. process the random noise. This is independent of the number of measurement outcomes of E.

To prove this result, we need to define an upper-information term $\mathbf{I}_{\mathrm{upper}}$ that is defined using *upper computable* tests. We say a semi-density matrix ρ is upper computable if there a program $q\in\{0,1\}^*$ such that when given to the universal Turing machine U, outputs, with or without halting, a finite or infinite sequence of elementary matrices ρ_i such that $\rho_{i+1}\preceq\rho_i$ and $\lim_{i\to\infty}\rho_i=\rho$. If U reads $\leq\|q\|$ bits on the input tape, then we say q upper computes ρ. The upper probability of an upper computable mixed state σ is defined by $\overline{\mathbf{m}}(\sigma|x)=\sum\{\mathbf{m}(q|x): q \text{ upper computes } \sigma\}$.

Let $\mathcal{G}_{C\otimes D}$ be the set of all upper computable matrices (tests) of the form $A\otimes B$, where $\mathrm{Tr}(A\otimes B)(C\otimes D)\leq 1$. Let $\mathfrak{G}_{C\otimes D}=\sum_{A\otimes B\in\mathcal{G}_{C\otimes D}}\overline{\mathbf{m}}(A\otimes B|n)(A\otimes B)$ be an aggregation of upper computable $C\otimes D$ tests of the form $A\otimes B$, weighted by their upper probability (which is a summation of *lower-computable* $\mathbf{m}$ measures of strings).

Definition 19 *The upper information between semi-density matrices A and B is* $\mathbf{I}_{\mathrm{upper}}(A:B)=\log \mathrm{Tr}\mathfrak{G}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(A\otimes B)$.

Proposition 5 $\mathbf{I}_{\mathrm{upper}}(I/2^n:I/2^n)=O(1)$.

Proof. $1\geq \mathrm{Tr}\mathfrak{G}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(\boldsymbol{\mu}\otimes\boldsymbol{\mu}) \stackrel{*}{>} \mathrm{Tr}\mathfrak{G}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(I/2^n\otimes I/2^n) \stackrel{*}{>} 2^{\mathbf{I}(I/2^n:I/2^n)}$. □

Lemma 13 *Let Λ be the uniform distribution on the unit sphere of an n qubit space. Let μ be the uniform distribution over mixed states introduced in Definition 17.*

- $\int 2^{\mathbf{I}_{\text{upper}}(|\psi\rangle:|\psi\rangle)} d\Lambda = O(1)$,
- $\int 2^{\mathbf{I}_{\text{upper}}(\sigma:\sigma)} d\mu(\sigma) = O(1)$.

Proof. The proof follows identically to that of Theorems 47 and 48, with reference to Proposition 5. The main benefit in using $\mathbf{I}_{\text{upper}}$ is that it achieves an additive majorization over the information of the probabilities resultant from applying measurements. This is in contrast to the standard definition of information, $\mathbf{I}$, which achieves only logarithmic majorization (see Theorem 52).

Lemma 14 ([Eps21a]) *Relativized to POVM E, $\mathbf{I}_{\text{Prob}}(E\sigma{:}E\sigma) <^{+} \mathbf{I}_{\text{upper}}(\sigma{:}\sigma)$.*

Proof. Note that all complexity terms are relativized to the number of qubits n, due to the relativization of E. Since $z(k) = \text{Tr}\boldsymbol{\mu}E_k$ is lower semi-computable and $\sum_k z(k) < 1$, $\mathbf{m}(k) \overset{*}{>} \text{Tr}\boldsymbol{\mu}E_k$, and so $1 > 2^{\mathbf{K}(k)-O(1)}\text{Tr}\boldsymbol{\mu}E_k$. So $\nu_{i,j} = 2^{\mathbf{K}(i)+\mathbf{K}(j)-O(1)}(E_i \otimes E_j) \in \mathcal{G}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}$ is an upper-computable $(\boldsymbol{\mu}\times\boldsymbol{\mu})$-test , with $\overline{\mathbf{m}}(\nu_{i,j}) \overset{*}{>} \mathbf{m}(i,j)$.

$$\begin{aligned}
\mathbf{I}_{\text{upper}}(\sigma{:}\sigma) &= \log \sum_{A\otimes B\in\mathcal{G}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}} \overline{\mathbf{m}}(A\otimes B)(A\otimes B)(\sigma\otimes\sigma) \\
&>^{+} \log \text{Tr}\sum_{ij} \nu_{i,j}\overline{\mathbf{m}}(\nu_{i,j})(\sigma\otimes\sigma) \\
&>^{+} \log \sum 2^{\mathbf{K}(i)+\mathbf{K}(j)}\mathbf{m}(i,j)E\sigma(i)E\sigma((j) \\
&>^{+} \mathbf{I}_{\text{Prob}}(E\sigma{:}E\sigma).
\end{aligned}$$

□

Remark 5 *The following theorems state that given a quantum measurement, for an overwhelming majority of pure or mixed quantum states, white noise (or the empty signal) will be produced. They are a part of the central results of the manuscript.*

Theorem 53 ([Eps21a]) *Let Λ be the uniform distribution on the unit sphere of an n qubit space. Relativized to POVM E, $\int 2^{\mathbf{I}_{\text{Prob}}(E|\psi\rangle:E|\psi\rangle)} d\Lambda = O(1)$.*

Proof. By Lemma 14, $2^{\mathbf{I}_{\text{upper}}(|\psi\rangle:|\psi\rangle)} \overset{*}{>} 2^{\mathbf{I}_{\text{Prob}}(E|\psi\rangle:E|\psi\rangle)}$. From Lemma 13, $\int 2^{\mathbf{I}_{\text{upper}}(|\psi\rangle:|\psi\rangle)} d\Lambda = O(1)$. The integral $\int 2^{\mathbf{I}_{\text{Prob}}(E|\psi\rangle:E|\psi\rangle)} d\Lambda$ is well defined because $2^{\mathbf{I}_{\text{Prob}}(E|\psi\rangle:E|\psi\rangle)} = \text{Tr}\sum_{i,j}\mathbf{m}(i,j)\nu_{i,j}(|\psi\rangle\langle\psi|\otimes|\psi\rangle\langle\psi|)$, for some matrices $\nu_{i,j}$ which can be integrated over Λ. □

Theorem 54 *Relativized to POVM E, $\int 2^{\mathbf{I}_{\text{Prob}}(E\sigma:E\sigma)} d\mu(\sigma) = O(1)$.*

Proof. By Lemma 14, $2^{\mathbf{I}_{\text{upper}}(\sigma:\sigma)} \overset{*}{>} 2^{\mathbf{I}_{\text{Prob}}(E\sigma:E\sigma)}$. From Lemma 13, $\int 2^{\mathbf{I}_{\text{upper}}(\sigma:\sigma)} d\mu(\sigma) = O(1)$. □

9.5 PVMs

However the measurement process has a surprising consquence, whereas for most states, an initial measurement produces no signal, the subsequent wave function collapse causes a massive uptake in algorithmic signal strength of the states. Thus a second measurement will produce a valid signal.

Quantum measurements is also of the form of PVMs, or projection value measures. A PVM $P = \{\Pi_i\}$ is a collection of projectors Π_i with $\sum_i \Pi_i = I$, and $\mathrm{Tr}\Pi_i\Pi_j = 0$ when $i \neq j$. When a measurement occurs, with probability $\langle\psi|\,\Pi_i\,|\psi\rangle$, the value i is measured, and the state collapses to

$$|\psi'\rangle = \Pi_i\,|\psi\rangle\,/\sqrt{\langle\psi|\,\Pi_i\,|\psi\rangle}.$$

Further measurements of $|\psi'\rangle$ by P will always result in the i measurement, so $P\,|\psi'\rangle\,(i) = 1$.

Let F be a PVM of 2^{n-c} projectors, of an n qubit space and let Λ_F be the distribution of pure states when F is measured over the uniform distribution Λ. Thus Λ_F represents the F-collapsed states from Λ. Note that if F has two few projectors, it lacks discretionary power to produce a meaningful signal when the states are in distribution Λ_F.

Theorem 55 $n - 2c <^+ \log \int 2^{\mathbf{I}_{\mathrm{Prob}}(F:|\psi\rangle:F|\psi\rangle|n)} d\Lambda_F$.

Proof. Note that $\int \langle\psi|\,\Pi_i\,|\psi\rangle\,d\Lambda = \mathrm{Dim}(\Pi_i)2^{-n}$. Furthermore, let $\kappa \subset \{1,\ldots,2^{n-c}\}$ be the set of numbers $a \in \kappa$ such that $\mathbf{K}(a|n) >^+ n - c$. So $|\kappa| \stackrel{*}{>} 2^{n-c}$. We have that if $\langle\psi|\,\Pi_i\,|\psi\rangle = 1$ then $\mathbf{I}_{\mathrm{Prob}}(F\,|\psi\rangle : F\,|\psi\rangle\,|n) = \mathbf{I}_{\mathrm{Prob}}(j \mapsto [i=j] : j \mapsto [i=j]|n) = \mathbf{I}(i:i|n) =^+ \mathbf{K}(i|n)$.

$$\begin{aligned}
&\int 2^{\mathbf{I}(F:|\psi\rangle:F|\psi\rangle)} d\Lambda_F \\
&= \sum_{i=1}^{2^{n-c}} \mathrm{Dim}(\Pi_i)2^{-n}2^{\mathbf{K}(i)} \\
&\stackrel{*}{>} \sum_{i\in\kappa} \mathrm{Dim}(\Pi_i)2^{-n}2^{n-c} \\
&\stackrel{*}{>} |\kappa|2^{-n}2^{n-c} \\
&\stackrel{*}{>} 2^{n-2c}.
\end{aligned}$$

□

9.6 Quantum Decoherence

The following letter of Einstein to Born (April 1954) illustrated the problem of superposition of quantum macrosystems.

> *Let Ψ_1 and Ψ_2 be two solutions to the Same Schrödinger equation... When the system is a macrosystem and when Ψ_1 and Ψ_2 are 'narrow with respect to position, then in by far the greater number of cases this is no longer true $\Psi_{12} = \Psi_1 + \Psi_2$. Narrowness with respect to macrocoordinates is not only independent of the principles of quantum mechanics, but is, moreover, incompatible with them.*

This letter brings up the astonishing fact that observables on the microscale and absent from everyday experiments. In fact, *quantum decoherence* and *einselection* show that such superpositions are highly fragile and decay exponentially fast. The root cause of this phenomena is caused by interactions between a system and environment. A closed system assumption is a fundamental obstacle to the study of the transition of the quantum domain to the classical domain.

In this light, the setup is a (microscopic) system and (macroscopic) environment. Given joint Hamiltonian dynamics between the system and environment, there are two main consequences.

1. The effective disappearance of coherence, the source of quantum interference effects, from the system.

2. The dynamical definition of preferred "pointer states", which are unchanged by the system/environment dynamics.

The phenomena of (1) is called *decoherence* (see [Sch10] for an extensive overview). The phenomena (2) is called *einselection*, short for Environment INduced Selection [Zur03]. In Einselection, the system-environment Hamiltonian "selects" a set of prefered quasi-classical "pointer states" which do not decohere. Einselection explains why we only observe a few "classical" quantities such as momentum and positon, and not superpositions of these pointer states.

We begin our explanation with a two state case, which can be generalized to arbitrary number of pointer states. Suppose the system is described by a superposition of two quantum states $|\psi_1\rangle$ and $|\psi_2\rangle$ which for example can be thought of as two localization of two positions x_1 and x_2 in a double slit experiment. The system/environment interaction results in

$$\begin{aligned} |\psi_1\rangle |E_0\rangle &\to |\psi_1\rangle |E_1\rangle \\ |\psi_2\rangle |E_0\rangle &\to |\psi_2\rangle |E_2\rangle . \end{aligned}$$

So the state of the environment evolves according to the state of the system. Now if the system is in a superposition of $|\psi_1\rangle$ and $|\psi_2\rangle$, we get the dynamics

$$\frac{1}{\sqrt{2}}(|\psi_1\rangle + |\psi_2\rangle) |E_0\rangle \to \frac{1}{\sqrt{2}}(|\psi_1\rangle |E_1\rangle + |\psi_2\rangle |E_2\rangle)$$

The reduced density matrix of system (with the environment traced out) is

$$\frac{1}{2} \left(|\psi_1\rangle \langle\psi_1| + |\psi_2\rangle \langle\psi_2| + |\psi_1\rangle \langle\psi_2| \langle E_2|E_1\rangle + |\psi_2\rangle \langle\psi_2| \langle E_1|E_2\rangle\right) .$$

The last two terms correspond to the interference between the state $|\psi_1\rangle$ and $|\psi_2\rangle$. If the environment recorded the position of the particle, then $|E_1\rangle$ and $|E_2\rangle$ will be approximately orthogonal. In fact, it can be shown that in many dynamics, $\langle E_1|E_2\rangle \leq e^{-t/\tau}$, where t is the time of the interaction and τ is a positive constant. In this case

$$\rho \approx \frac{1}{2} \left(|\psi_1\rangle \langle\psi_1| + |\psi_2\rangle \langle\psi_2|\right) .$$

Thus virtually all coherence between the two states $|\psi_1\rangle$ and $|\psi_2\rangle$ is lost. The states $|\psi_1\rangle$ and $|\psi_2\rangle$ are called invariant to the dynamics, and will not undergo decoherence. They are called "pointer states" because they induce an apparatus with a pointer mechanism to be orientated at a particular angle. Einselection preserves "pointer states" but superpositions of them are fragile and do not survive the dynamics with the system.

9.6.1 Predictability Sieve

In general, there is not a clear division between pointer and non-pointer states. Instead one can use a score to measure how much of the state has been preserved. The interaction of pointer states with the environment is predictable; they are effectively classical states. However a state that is heavily decohered is unpredictable. Let $|\psi\rangle$ be an initial pure state, and $\rho_{|\psi\rangle}(t)$ be the density matrix of the system state after interacting with the environment for time t. The loss of predictability caused by the environment can be measured in the following two measures

- $\varsigma^T_{|\psi\rangle}(t) = \mathrm{Tr}\rho^2_{|\psi\rangle}(t)$.
- $\varsigma^S_{|\psi\rangle}(t) = S(\rho_{|\psi\rangle}(t))$.

The first measure, uses squared trace of the density matrix whereas the second measure uses von Neumann entropy. The first measure will start at 1 and then decrease proportionately to much much the state decoheres. This is similarly true for the von Neumann entropy predictability sieve, except the measure starts at 0.

In this section we introduce an algorithmic predictability sieve ς^A. Assume a basis of 2^n pointer states. Let the system be $|\psi\rangle$, an arbitrary pure state. We consider the limit of interacting with the environment as time approaches infinity. In this idealized case, the decoherence $|\psi\rangle\langle\psi|$ into a classical probability, with the off-diagonal terms turned to 0. Let $p_{|\psi\rangle}$ be the classical probability that σ decoheres to, with $p_{|\psi\rangle}(i) = |\psi\rangle\langle\psi|(ii)$.

Definition 20 (Algorithmic Predictability Sieve) $\varsigma^A(|\psi\rangle) = \mathbf{I}_{\mathrm{Prob}}(p_{|\psi\rangle} : p_{|\psi\rangle}|n)$.

Thus, ς^A is the self information of the probability measure induced by the diagonal of the density matrix $|\psi\rangle\langle\psi|$. Note that this self information is relativized to n, that is the universal Turing machine U has n on an auxiliary tape. We first show that, on average, pointer states $|i\rangle$ have high algorithmic predictability.

Theorem 56

$$\frac{1}{2^n}\sum_{i=1}^{2^n}\varsigma^A(|i\rangle) =^+ n.$$

Proof. By the definition of $\mathbf{I}_{\mathrm{Prob}}$, we have that $\varsigma^A(|i\rangle) =^+ \mathbf{K}(i|n)$. Since $\mathbf{K}(i)$ is a prefix free code, $\frac{1}{2^n}\sum_{i=1}^{2^n}\mathbf{K}(i) \geq H(u_n) = n$, where $H(u_n)$ is the entropy of the uniform distribution u_n over strings of length n. Furthermore, the upper bound comes from the fact that $\max_{i\in\{1,\ldots,2^n\}}\mathbf{K}(i|n) <^+ n$.□

We now show that an overwhelming majority of pure states over the pointer basis decohere into algorithmic white noise. Due to algorithmic conservation inequalities (see Theorem 3), there is no (even probabilisitic) method of processing this white noise to produce a signal. Thus superpositions of pointer bases will produce garbage that can't be measured. The following corollary to Theorem 53 follows from the fact that there is a POVM E, where $E_i = |i\rangle\langle i|$ with $E_i|\psi\rangle = p_{|\psi\rangle}(i)$.

Corollary 14 *Let Λ be the uniform distribution on the unit sphere of an n qubit space.*

$$\int 2^{\varsigma^A(|\psi\rangle)}d\Lambda = O(1).$$

Apriori distributions which are close to Λ also have this property.

Proposition 6 *Let Γ be a distribution over n qubit pure states such that $\Gamma(|\psi\rangle) \leq 2^c\Lambda(|\psi\rangle)$ for all $|\psi\rangle$.*

$$\log\int 2^{\varsigma^A(|\psi\rangle)}d\Gamma <^+ c.$$

Proof. Using Corollary 14,

$$\int 2^{\varsigma^A(|\psi\rangle)}d\Gamma \leq 2^c\int 2^{\varsigma^A(|\psi\rangle)}d\Lambda \stackrel{*}{<} 2^c.$$

□

Chapter 10

Infinite Quantum Spin Chains

A qubit abstracts the properties of a single spin 1/2 particle. A complex system can be described by the collection of qubits, which model properties of superposition and entanglement. It can be convenient to consider a system's *thermodynamic limit*, which is the limit of a system of infinite size. In this chapter we use an infinite quantum spin chain as our model. In the study of infinite quantum spin chains one can make a distinction between local and global effects. In addition, one does not need to consider boundary conditions.

A Martin Löf random sequence is the accepted definition in AIT for a random infinite sequence. Can one define a quantum Martin Löf infinite quantum state? This chapter shows that this can be answered in the affirmative, and even landmark theorems in AIT like the Levin-Schnorr theorem can transfer over to the quantum domain.

We first review Martin Löf random sequences. A Martin Löf test is an effective null set of the form $\bigcap_n G_n$, where the measure of open set G_n of the Cantor space goes toward zero. An infinite sequence passes a Martin Löf test if it is not contained in its null set. A Martin Löf random infinite sequence passes all Martin Löf tests. Let MLR be the set of Martin Löf random sequences.

In [NS19], the set of random infinite quantum states was introduced, which we call NS random states. Just like the classical setting, a NS random state passes allso-called NS tests. An NS test is a quantum analog to Martin Löf tests, and it is defined by projections instead of open sets.

10.1 Infinite Quantum Bit Sequences

Before we introduce NS random sequences, we revisit the notion of C^* algebras and functional states. A C^* algebra, $\mathcal{M}$, is a Banach algebra and a function $* : \mathcal{M} \to \mathcal{M}$ such that

- For every $x \in \mathcal{M}$, $x^{**} = x$.
- For every $x, y \in \mathcal{M}$, $(x + y)^* = x^* + y^*$ and $(xy)^* = y^*x^*$.
- For every $\lambda \in \mathbb{C}$ and $x \in \mathcal{M}$, $(\lambda x)^* = \overline{\lambda} x^*$.
- For all $x \in \mathcal{M}$, $\|x^*x\| = \|x\|\|x^*\|$.

A C^* algebra $\mathcal{M}$ is unital if it admits a multiplicative identity $\mathbf{1}$. A state over unital $\mathcal{M}$ is a positive linear functional $Z : \mathcal{M} \to \mathbb{C}$ such that $Z(\mathbf{1}) = 1$. States are used to define NS random sequences. The set of states of $\mathcal{M}$ is denoted by $S(\mathcal{M})$. A state is tracial if $Z(x^*x) = Z(xx^*)$, for all $x \in \mathcal{M}$.

The C^* algebra over matrices of size 2^k over $\mathbb{C}$ is denoted by $\mathcal{M}_k$. Each state $\rho \in S(\mathcal{M}_k)$, can be identified with a density matrix S such that $\rho(X) = \mathrm{Tr}SX$, for all $X \in \mathcal{M}$. States that cannot

be represented as the convex combination of other states are called pure states. Otherwise they are called mixed states. States are used interchangeably with density matrices, depending on the context. The tracial state $\tau_n \in S(\mathcal{M}_n)$ corresponds to the matrix $2^{-n} I_{2^n}$. The algebra $\mathcal{M}_\infty$ is the direct limit of the ascending sequence of $\mathcal{M}_n$. . A state $Z \in S(\mathcal{M}_\infty)$ over $\mathcal{M}_\infty$ can be seen as a sequence of density matrices $\{\rho_n\}$ that are coherent under partial traces, with $\mathrm{Tr}_{\mathcal{M}_{n+1}} \rho_{n+1} = \rho_n$. We use $Z\restriction n$ to denote the restriction of state Z to the algebra $\mathcal{M}_n$. There is a unique tracial state $\tau \in S(\mathcal{M}_\infty)$, where $\tau\restriction n = \tau_n$. A projection $p \in \mathcal{M}_\infty$ is a self adjoint positive element such that $p = p^2$. A special projection $p \in \mathcal{M}_n$ is a projection represented by an elementary matrix.

10.1.1 NS Randomness

An NS Σ^0_1 set is a computable sequence of special projections $\{p_i\}$ in $\mathcal{M}_\infty$ with $p_i \leq p_{i+1}$ over all i. For state ρ and NS Σ^0_1 set G, $\rho(G) = \sup_i \rho(p_i)$.

We define NS tests. But initially, we will provide the definition for the classical Martin Löf random sequence, to provide a point of reference. A classical Martin Löf test, is a sequence $\{U_n\}$ of uniformly Σ^0_1 sets of infinite sequences $U_n \subseteq \{0,1\}^\infty$ such that $\mu(U_n) \leq 2^{-n}$. An infinite sequence $\alpha \in \{0,1\}^\infty$ is Martin-Löf random if there is no Martin Löf test $\{U_n\}$ such that $\alpha \in \bigcap_n U_n$. There is a universal Martin Löf test $\{V_n\}$ such that if $\alpha \notin \bigcap_n V_n$, then α is random.

Mirroring the classical case, a NS test is an effective sequence of NS Σ^0_1 sets $\langle G^r \rangle$ such that $\tau(G^r) \leq 2^{-r}$. Unlike a classical test, which can either pass or fail a sequence, a NS test can pass a quantum state up to a particular order. For $\delta \in (0,1)$, state $Z \in S(\mathcal{M}_\infty)$ fails test $\langle G^r \rangle$ at order δ if $Z(G^r) > \delta$ for all r. Otherwise Z passes the test at order δ. We says Z passes a NS test if it passes it at all orders $\delta \in (0,1)$. A state is NS random if it passes every NS test at every order.

Theorem 57 ([NS19]) *There exists a universal NS test $\langle R^n \rangle$, where for each NS test $\langle G^k \rangle$ and each state Z and for each n there exists a k such that $Z(R^n) \geq Z(G^k)$.*

Proof. Let $\langle G^k_n \rangle_{n=1}^\infty$ be an enumeration of NS tests, performed analgously to the classical case (see [Gác01]). Furthermore let $G^{e,m} = \langle p_r^{e,m} \rangle_{r \in \mathbb{N}}$. For each $k, n \in \mathbb{N}$, let $q^n_k = \bigvee_{e+n+1 \leq k} p_k^{e,e+n+1}$. Thus $q^n_k \leq q^n_{k+1}$ and $\tau(q^n_k) \leq \sum_e \tau(p_k^{e,e+n+1}) \leq 2^{-n}$. The universal NS test is $R^n = \langle q^n_k \rangle_{k \in \mathbb{N}}$. Since $\tau(R^r) \leq 2^{-n}$, $\langle R^n \rangle$ is a NS test. For a set e,

$$\rho(R^n) = \sup_k \rho(q^n_k) \geq \sup_k \rho(p_k^{e,n+e+1}) = \rho(G^{e,n+e+1}).$$

□

A state Z is NS random if it passes the test $\langle R^n \rangle$. More information about $\langle R^n \rangle$ can be found in [NS19].

Exercise 13 ([NS19]) *Let $S_{n,i}$ be the subspace of $\mathbb{C}^{2^n}$ generated by $|x\rangle$ where $x \in \{0,1\}^n$ has $x[i] = 1$. So for any state Z on M_∞, the real $Z(S_{n,i})$ is the probability that a measurement of the ith qubit of its initial segment $Z\restriction n$ return 1. Prove that for NS random state Z,*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i<n} Z(S_{n,i}) = 1/2.$$

Exercise 14 ([NS19]) *Let $Z \in \{0,1\}^\infty$. Show that Z is ML random iff Z viewed as an element of $S(M_\infty)$ is NS random.*

10.2 Closure Properties

The set of NS random sequences has closure properties over (possibly noncomputable) convex combinations, as shown in the following theorem.

Theorem 58 *Every convex combination $Z = \sum_i \alpha Z_i$ of NS random states Z_i, with $\sum_i \alpha_i = 1$ and $\alpha_i \geq 0$, is NS random.*

Proof. Given an NS test $\langle G^r \rangle = \langle p_t^r \rangle$, there exists a NS test $\langle H^r \rangle$ such that for all states Z, $\inf_r Z(H^r) \geq \inf_r Z(G^r)$ and $H^r \supseteq H^{r+1}$. This is by setting H^r equal to $\bigvee_{i \geq r} G^i$. More formally, $\langle H^r \rangle = \langle q_t^r \rangle$, where $q_t^r = \bigvee_{i=1}^t p_t^{r+i}$. Thus there exists a universal NS test $\langle L^r \rangle$ such that $L^r \supseteq L^{r+1}$. Assume that Z is not NS random. Then

$$\lim_{r \to \infty} Z(L^r) > 0$$
$$\lim_{r \to \infty} \sum_i \alpha_i Z_i(L^r) > 0$$
$$\sum_i \alpha_i \lim_{r_i \to \infty} Z_i(L^{r_i}) > 0.$$

So there exists an i such that $\lim_{r \to \infty} Z_i(L^r) > 0$, and thus Z_i is not NS random. □

10.3 Gács Complexity and NS Random Sequences

In this section, we characterize NS random states in terms of Gács complexity, **Hg**.

Lemma 15 *Given state $Z \in \mathcal{M}_\infty$, and program p that enumerates infinite set $A \subseteq \mathbb{N}$, then $\sup_{n \in \mathbb{N}} n - \mathbf{Hg}(Z \upharpoonright n) <^+ \sup_{n \in A} n - \mathbf{Hg}(Z \upharpoonright n) + \mathbf{K}(p)$.*

Proof. There exists a program p' of size $\|p\| + O(1)$ that outputs a list $\{a_n\} \subseteq A$ such that $n < a_n$. For a given a_n, $\sigma = 2^{n - a_n} \boldsymbol{\mu}_n \otimes I_{a_n - n}$ is a lower computable $2^{a_n} \times 2^{a_n}$ semi-density matrix. There is a program $q = q' \langle a_n, n \rangle$ that lower computes σ where q' is helper code that uses the encodings of a_n and n. By the universal properties of $\boldsymbol{\mu}$, we have the inequality $\mathbf{m}(q|a_n)\sigma \stackrel{*}{<} \boldsymbol{\mu}_{a_n}$. So, using properties of partial trace,

$$\begin{aligned}
a_n + \log \mathbf{m}(q|a_n) \mathrm{Tr} \sigma Z \upharpoonright a_n &<^+ a_n + \log \mathrm{Tr} \boldsymbol{\mu}(Z \upharpoonright a_n) \\
a_n + \log \mathrm{Tr} 2^{n - a_n} (\boldsymbol{\mu}_n \otimes I_{a_n - n}) Z \upharpoonright a_n - \mathbf{K}(q|a_n) &<^+ a_n + \log \mathrm{Tr} \boldsymbol{\mu}(Z \upharpoonright a_n) \\
n + \log \mathrm{Tr} (\boldsymbol{\mu}_n \otimes I_{a_n - n}) Z \upharpoonright a_n - \mathbf{K}(\langle n, a_n \rangle | a_n) &<^+ a_n + \log \mathrm{Tr} \boldsymbol{\mu}(Z \upharpoonright a_n) \\
n + \log \mathrm{Tr} (\boldsymbol{\mu}_n \mathrm{Tr}_n Z \upharpoonright a_n) - \mathbf{K}(p'|a_n) &<^+ a_n + \log \mathrm{Tr} \boldsymbol{\mu}(Z \upharpoonright a_n) \\
n - \mathbf{Hg}(Z \upharpoonright n) &<^+ a_n - \mathbf{Hg}(Z \upharpoonright a_n) + \mathbf{K}(p).
\end{aligned}$$

So $\sup_{n \in \mathbb{N}} n - \mathbf{Hg}(Z \upharpoonright n) <^+ \sup_{a_n \in \{a_n\}} a_n - \mathbf{Hg}(Z \upharpoonright a_n) + \mathbf{K}(p) <^+ \sup_{n \in A} n - \mathbf{Hg}(Z \upharpoonright n) + \mathbf{K}(p)$. □

Theorem 59 *Suppose for state Z, and for infinite enumerable set $A \subseteq \mathbb{N}$, $\sup_{n \in A} n - \mathbf{Hg}(Z \upharpoonright n) < \infty$. Then Z is NS random.*

Proof. Suppose Z is not NS random. Let $L^r = \langle p_t^r \rangle$ be the universal NS test. So $\text{Rank}(p_n^r) \leq 2^{n-r}$. Thus $\inf_r Z(L^r) = \delta > 0$. For each r, there exists an n such that $\text{Tr}(p_n^r z_n) \geq \delta$, where $z_n = Z \upharpoonright n$. Since $2^{r-n} p_n^r$ is a computable semi-density matrix given n and r, $\mathbf{m}(r|n)2^{r-n}p_n^r \stackrel{*}{<} \boldsymbol{\mu}$. So $\mathbf{m}(r|n)2^{r-n}\delta \stackrel{*}{<} \text{Tr}\boldsymbol{\mu} z_n$, which implies that $\mathbf{Hg}(Z \upharpoonright n) <^+ n - r + \mathbf{K}(r|n)$. Since this property holds for all $r \in \mathbb{N}$, $\sup_n n - \mathbf{Hg}(Z \upharpoonright n) = \infty$. From Lemma 15, $\sup_{n \in A} n - \mathbf{Hg}(Z \upharpoonright n) = \infty$. □

10.4 Encodings of States

Let $[Z] \in \{0,1\}^\infty$ be an encoding of the state Z described as follows. For each n, let $e(n,m)$ be the mth enumeration of a pair (p,k) consisting of a special projection p of $\mathcal{M}_n$ and a rational $0 \leq k \leq 1$. For $[Z]$, the ith bit, where $i = 2^n m$ for maximum n, corresponds to 1 if and only if $\text{Tr}pZ\upharpoonright n > k$, where (p,k) is the pair enumerated by $e(n,m)$. We say that state $Z \in \mathcal{QH}$ if and only if the halting sequence can be computed from $[Z]$. By the Independence Postulate (see Chapter **??**), states in $\mathcal{QH}$ are exotic and non-physical.

10.5 Quantum Unitary Complexity

In this section, the communication complexity of a quantum state is introduced. In a canonical algorithmic information theory example, Alice wants to send a single text message x to Bob. Alice sends a program p to B such that $x = U(p)$, where U is a fixed universal Turing machine. The cost of the transmission is the length of p. Alice can minimize cost by sending $\mathbf{K}(x)$ bits to Bob, where $\mathbf{K}$ is the Kolmogorov complexity function. We now look at the quantum case. Suppose that Alice wants to send a (possibly mixed) n qubit quantum state σ to Bob. Alice has access to two channels, a quantum channel and a classical channel. Alice can choose to send $m \leq n$ qubits ρ on the quantum channel and P regular bits p on the classical channel, representing an encoding of unitary operation V, and m, where $U(p) = (V,m)$. Bob, upon receiving ρ and p, constructs the unitary operation V, and then applies it to ρ (tensored with $|0^{n-m}\rangle$) to produce $\sigma' = V(|0^{n-m}\rangle \rho \langle 0^{n-m}|)V^*$. Bob is not required to produce σ exactly. Instead the precision of the attempt is measured by trace distance between σ and σ'.

A quantum unitary pair (V,m) consists of two parts, (1) an elementary unitary transform over $\mathcal{Q}_n$, and (2) the number of qbits $m \leq n$ that are connected to the quantum input. $\mathcal{C}_{n,m}$ be the set of all quantum pairs over n qbits with an input size of m. The precision is measured using the trace distance D.

Definition 21 *For density matrix σ, the quantum unitary complexity at accuracy ϵ is* $\mathbf{Huc}^\epsilon(\sigma) = \min\{\mathbf{K}(V,m) + m \colon (V,m) \in \mathcal{C}_{n,m}, \xi \in S(\mathcal{M}_m), D(\sigma, V(|0^{n-m}\rangle \xi \langle 0^{n-m}|)V^*) < \epsilon\}$.

Lemma 16 *Let R_j be the smallest subspace spanned by pure states produced by quantum unitary pairs $(V,z) \in \mathcal{C}_{n,z}$ with $\mathbf{K}((V,z)) + z < j$. Then $Dim(R_j) < 2^j$.*

Proof. Let $b(y,z)$ be the number of programs of length y that outputs an quantum unitary pair $(V,z) \in \mathcal{C}_{n,z}$. Let $b(y) = \sum_{z \leq n} b(y,z)$

$$\begin{aligned}
\mathrm{Dim}(R_j) &\leq \sum_{y+z<j} b(y,z)2^z \\
&= 2^j \sum_{y+z<j} b(y,z)2^{z-j} \\
&< 2^j \sum_{y,z} b(y,z)2^{-y} \\
&= 2^j \sum_{y} b(y)2^{-y} \\
&\leq 2^j.
\end{aligned}$$

□

The following theorem uses **Huc** conditioned on n, which means the universal Turing machine U has n on an auxiliary tape.

Theorem 60 $\mathbf{Hg}(\sigma) <^+ \mathbf{Huc}^\epsilon(\sigma|n) - \log(1-\epsilon)$.

Proof. Let quantum unitary pair $(V,m) \in \mathcal{C}_{n,m}$ and input $\rho \in S(\mathcal{M}_m)$ realize $\mathbf{Huc}^\epsilon(\sigma|n)$, where $D(\sigma,\xi) < \epsilon$, where $\xi = V(|0^{n-m}\rangle \rho \langle 0^{n-m}|)V^*$. Let I_k be the projection to k length inputs. So for $\nu = VI_mV^*$, $\mathrm{Tr}\nu\xi = 1$. By the properties of trace distances, this implies that $\mathrm{Tr}\nu\sigma > 1-\epsilon$. We define the following lower computable semi-density matrix,

$$\lambda = \sum_{m=1}^{n} 2^{-m} \sum_{(V,m)\in\mathcal{C}_{n,m}} \mathbf{m}((V,m)|n)VI_mV^*,$$

where I_m is the projection onto m length states. So

$$2^{-m}\mathbf{m}((V,m)|n)\nu \leq \lambda \stackrel{*}{<} \boldsymbol{\mu}.$$

This implies that

$$\mathbf{m}((V,m)|n)2^{-m}(1-\epsilon) \leq \mathbf{m}((V,m)|n)2^{-m}\mathrm{Tr}\sigma\nu \stackrel{*}{<} \mathrm{Tr}\boldsymbol{\mu}\sigma.$$

So

$$\mathbf{Hg}(\sigma) <^+ \mathbf{K}((V,m)|n) + k - \log(1-\epsilon) =^+ \mathbf{Huc}^\epsilon(\sigma|n) - \log(1-\epsilon).$$

□

Theorem 61 *Let $|\psi\rangle$ vary over pure n qubit states and $Q_{\epsilon,n,m} = \max_{|\psi\rangle} \mathbf{Huc}^\epsilon(|\psi^m\rangle)$. Then we have that* $\log \binom{2^n+m-1}{m} <^+ Q_{\epsilon,n,m} <^+ \mathbf{K}(n,m) + \log \binom{2^n+m-1}{m}$.

Proof. Let $M = \binom{2^n+m-1}{m}$. As seen in the proof of Theorem 18, for pure n qubit states $|\psi\rangle$, the subspace S of pure states $|\psi^m\rangle$ has dimension M. Let T be the subspace of $\mathcal{Q}_{nm}$ spanned by the basis states $|B(i)\rangle |0^{mn-\lceil \log M \rceil}\rangle$, where $B(i)$ is the binary representation of the number i, ranging from 1 to M. Let V be an elementary unitary transform that maps T to S. Thus every pure state

$|\psi^n\rangle$ of S can be reproduced by a pure state in T and a quantum unitary group $(V, \lceil \log M \rceil)$ of complexity $<^+ \mathbf{K}(n, m)$. Therefore $C_{\epsilon,n,m} <^+ \mathbf{K}(n, m) + \log M$. Let R_j be the subspace spanned by pure states produced by quantum unitary groups (V, k) with $\mathbf{K}(V, k) + k < j$. By Lemma 16, $\mathrm{Dim}(R_j) < 2^j$. Thus there is pure state $|\psi^m\rangle$ that is orthogonal to the subspace $R_{\lceil \log M \rceil - 1}$. Thus $C_{\epsilon,n,m} >^+ \log \binom{2^n+m-1}{m}$. □

10.6 Initial Segment Incompressibility

Due to Levin and Schnorr, [Lev74, Sch71] α is random iff there is an r such that for all n, $\mathbf{K}(\alpha_{\leq n}) \geq n - r$, where $\alpha_{\leq n}$ is a prefix of α of size n, and $\mathbf{K}$ is prefix free Kolmogorov complexity. In this section, we prove a quantum analog to this result. We show that NS states that are NS random have incompressible prefixes with respect to quantum operation complexity. Otherwise the quantum states are in $\mathcal{QH}$. Theorem 62 builds upon the proof of the Theorem 4.4 in [NS19] using quantum unitary complexity **Huc**.

Theorem 62 *Let Z be a state on $\mathcal{M}_\infty$.*

1. *Let $1 > \epsilon > 0$, and suppose Z passes each NS test at order $1 - \epsilon$. Then there is an r where for all n, $\mathbf{Huc}^\epsilon(Z\restriction_n) > n - r$.*

2. *Let $1 > \epsilon > 0$ be lower computable and Z fails some NS test at order $1 - \epsilon$. Then either $Z \in \mathcal{QH}$ or for all r, there is an n where $\mathbf{Huc}^{\sqrt{\epsilon}}(Z\restriction_n) < n - r$.*

Proof. (1). Let $\mathbf{K}_t(x)$ be the smallest program to produce x in time t. Let $s(n, r, t)$ be the set of pure n qubit states $|\psi\rangle \in S(\mathcal{M}_n)$ such that there exists a quantum unitary pair $(V, m) \in \mathcal{C}_{n,m}$ and pure state $|\phi\rangle \in S(\mathcal{M}_m)$ such that $|\psi\rangle \langle\psi| = V(|0^{n-m}\rangle |\phi\rangle \langle\phi| \langle 0^{n-m}|)V^*$ and $\mathbf{K}_t((V, m))+m \leq n-r$. Let $p(n, r, t)$ be the orthogonal projection in $\mathcal{M}_n$ with minimum $\tau(p(n, r, t))$ such that $\rho(p(n, r, t)) = 1$ for all $\rho \in s(n, r, t)$. Let $p(r, t) = \sup_{n \leq t} p(n, r, t)$. So $p(r, t)$ is in $\mathcal{M}_t$, $p(r, t) \leq p(r, t + 1)$, and $p(r, t)$ is computable from r and t. This is because one can compute a collection of sets W_n of quantum unitary pairs $(V, m) \in \mathcal{C}_{n,m}$, with $n \leq t$, and $\mathbf{K}_t((V, m)) + m \leq n - r$. So we have that $p(r, t) = \sup_{n \leq t} \sup_{(V,m) \in W_n} V(|0^{n-m}\rangle I_m \langle 0^{n-m}|)V^*$, where I_m is the m qubit identity matrix.

Let $b(y, n, z)$ be the number of programs of length y which output an quantum unitary pair in

$\mathcal{C}_{n,z}$. Let $b(y,n) = \sum_{z\le n} b(y,n,z)$. So

$$\begin{aligned}
\text{Range}(p(n,r,t)) &\le \sum_{y+z\le n-r} b(y,n,z)2^{z} \\
\tau(p(n,r,t)) &\le \sum_{y+z\le n-r} b(y,n,z)2^{z-n} \\
&\le \sum_{y+z\le n-r} b(y,n,z)2^{-y-r} \\
&\le \sum_{y=1}^{n-r} b(y,n)2^{-y-r} \\
\tau(p(r,t)) &\le \sum_{n=1}^{\infty} \tau(p(n,r,t)) \\
&\le \sum_{n=1}^{\infty}\sum_{y=1}^{n-r} b(y,n)2^{-y-r} \\
&= 2^{-r}\sum_{n=1}^{\infty}\sum_{y=1}^{n-r} b(y,n)2^{-y} \le 2^{-r}.
\end{aligned}$$

So for NS Σ_1^0 set G^r enumerated by the sequence $\{p(r,t)\}_t$, $\langle G^r\rangle$ is a NS test. For each r suppose there is an n such that $\mathbf{Huc}^{\epsilon}(Z\restriction n) \le n-r$. So there is a quantum unitary pair $(V,z) \in \mathcal{C}_{n,z}$ and input $\rho \in S(\mathcal{M}_z)$ such that $\mathbf{K}((V,z)) + z \le n-r$ and $D(Z\restriction_n, \xi) < \epsilon$, where $\xi = V\left|0^{m-n}\right\rangle \rho \left\langle 0^{n-m}\right| V^*$. So ξ is in the range of $p(n,r,t)$ for some t and so $\text{Tr}\xi p(n,r,t) = 1$. This implies $1-\epsilon < Z(p(n,r,t)) \le Z(p(r,t)) \le Z(G^r)$. Since this is for all r, Z fails the test at order $1-\epsilon$.

(2). Let $\mathbf{bb}(n)$ be the longest running time of a halting program of length $\le n$. Let $\langle L^r\rangle$ be the universal NS test, where each L^r is enumerated by $\{p_t^r\}$, with $p_t^r \in \mathcal{M}_{n(r,t)}$. Assume there is an infinite number of r where $\text{Tr}Z\restriction n(r,\mathbf{bb}(r/2))p^r_{\mathbf{bb}(r/2)} > 1-\epsilon$. Fix one such r and let $n = n(r,\mathbf{bb}(r/2))$, and $p = p^r_{\mathbf{bb}(r/2)}$. Projection p has eigenvectors $\{|u_i\rangle\}$ and kernel spanned by $\{|v_i\rangle\}$. Thus $2^{-r} \ge \tau(p)$. Let $p' \ge p$ with $p' \in \mathcal{M}_n$ such that each u_i is in the range of p' and $\{|v_i\rangle\}_{i=1}^k$ is in the range of p' such that k is minimized such that $\tau(p') = 2^{-r}$. Thus $\text{Tr}Z\restriction n(p') > 1-\epsilon$. The eigenvectors of p' are $\{|w_i\rangle\}_{i=1}^{2^{n-r}}$ and its kernel is spanned by the vectors $\{|y_i\rangle\}_{i=1}^{2^n-2^{n-r}}$. Let $z' = \text{Proj}(Z\restriction n; p')$ be a density matrix with eigenvalues $x_i \in \mathbb{R}$ corresponding to eigenvectors $|w_i\rangle$, where Proj is defined in Proposition 7. For $i \in [1,2^n]$, let $B(i) \in \{0,1\}^*$ be an encoding of n bits of the number i, with $B(1) = |0^{(n)}\rangle$, $B(2) = |10^{(n-1)}\rangle$, and $B(2^n) = |1^{(n)}\rangle$. Let U be a $2^n \times 2^n$ unitary matrix, of the form $U = \sum_{i=1}^{2^{n-r}} |B(i)\rangle\langle w_i| + \sum_{i=1}^{2^n-2^{n-r}} |B(i+2^{n-r})\rangle\langle y_i|$.

Proposition 7 ([NS19]) *Let* $\text{Proj}(s;h) = \frac{1}{\text{Tr}[sh]}hsh$. *Let p be a projection in M_n and σ be a density matrix in M_n. If $\alpha = \text{Tr}p\sigma$ and $\sigma' = \text{Proj}(\sigma;p)$ then $D(\sigma,\sigma') \le \sqrt{1-\alpha}$.*

Proof. Let $|\psi_\sigma\rangle$ be a purification of σ. Then $\alpha^{-\frac{1}{2}} p\,|\psi_\sigma\rangle$ is a purification of σ'. Uhlmann's theorem states $F(\sigma,\sigma') \ge \alpha^{-\frac{1}{2}}\langle\psi_\sigma|\,p\,|\psi_\sigma\rangle = \alpha^{\frac{1}{2}}$, where F is fidelity, with $F(\sigma,\sigma') = \text{Tr}\sqrt{\sqrt{\sigma'}\sigma\sqrt{\sigma'}}$. Thus the proposition follows from $D(\sigma,\sigma') \le \sqrt{1-F(\sigma,\sigma')}$. □

For the diagonal $2^{n-r}\times 2^{n-r}$ matrix σ with entries $\{x_i\}_{i=1}^{2^{n-r}}$, $z' = U(|0^r\rangle\sigma\langle 0^r|)U^*$. By Proposition 7, since $1-\epsilon < \text{Tr}(p'Z\restriction n)$ and $z' = \text{Proj}(z_n;p')$, it must be that $D(z', Z\restriction n) < \sqrt{\epsilon}$. Thus using

the quantum unitary pair $(U, n-r)$,

$$\begin{aligned}\mathbf{Huc}^{\sqrt{\epsilon}}(Z\upharpoonright n) &\leq \mathrm{Dim}(\sigma) + \mathbf{K}((V, n-r))\\ &<^+ n - r + \mathbf{K}(n, r)\\ &<^+ n - r + \mathbf{K}(\mathbf{bb}(r/2), r)\\ &<^+ n - r + r/2 + \mathbf{K}(r)\\ &<^+ n - r/3.\end{aligned}$$

Thus for every r there exists an n where $\mathbf{Huc}^{\sqrt{\epsilon}}(Z\upharpoonright n) < n-r$. This is because the additive constant of the above equation is not dependent on r.

Otherwise there is some R where for all $r \geq R$, and $q < \mathbf{bb}(r/2)$, $\mathrm{Tr}Z_{n(r,q)}p_q^r \leq 1-\epsilon$. Thus given R, $\langle L^r\rangle$, $[Z]$, and a lower enumeration of ϵ, one can iterate through each $r \geq R$ and return an s such that $\mathrm{Tr}Z_{n(r,s)}p_s^r > 1-\epsilon$.. This number s has the property that $s \geq \mathbf{bb}(r/2)$, and can be used to compute the prefix of the halting sequence over all programs of length $\leq r/2$ as every such program that will halt will do so in less than s steps. Thus the halting sequence is computable relative to $[Z]$ and thus $Z \in \mathcal{QH}$. □

Corollary 15 *Let state $Z \notin \mathcal{QH}$. Then Z is NS random iff for all $0 < \epsilon < 1$, there is an r, where for all n, $\mathbf{Huc}^{\epsilon}(Z \upharpoonright n) > n - r$.*

Proof. Assume Z is NS random. Then for all $0 < \epsilon < 1$, Z passes each NS test at order $1-\epsilon$. Then by Theorem 62 (1), for all $0 < \epsilon < 1$ there is an r where for all n, $\mathbf{Huc}^{\epsilon}(Z\upharpoonright n) > n-r$. Assume Z is not NS random. Then there is some rational $0 < \delta < 1$ such that Z fails some NS test at order $1-\delta$. Then by Theorem 62 (2), for $\epsilon = \sqrt{\delta}$, for all r, there is an n where $\mathbf{Huc}^{\epsilon}(Z\upharpoonright n) < n - r$. □

10.6.1 Prefix-Free Quantum Kolmogorov Complexity

Comparable result to Theorem 62 have been achieved in [Bho21b], using a definition called prefix-free quantum Kolmogorov complexity, **QK**. Like **Huc**, the definition of **QK** uses the summation of two parts: a classical component and a quantum component. We recall that U is the (classical and prefix free) universal Turing machine.

Definition 22 $\mathbf{QK}^{\epsilon}(\sigma) = \min\{\|x\| + \log \mathrm{Rank}(P) : U(x) = \textit{Projection } P \textit{ and } \mathrm{Tr}P\sigma > \epsilon\}$.

Theorem 63 ([Bho21b]) *Assume state $\rho \in M_\infty$ does not compute $\emptyset''$. Then ρ is NS random iff for all $\epsilon \in (0,1)$ there exists c such that for all n, $\mathbf{QK}^{\epsilon}(\rho \upharpoonright n) > n - c$.*

10.7 Quantum Ergodic Sources

In [Bru78], Brudno proved that for ergodic measures η over bi-infinite sequences, for η-almost all sequences, the rate of the Kolmogorov complexity of their finite prefixes approaches the entropy rate of η. Therefore the average compression rate of sequences produced by η is not more than its entropy rate. In [BKM+06], a quantum version of Brudno's theorem was introduced relating, in a similar fashion, Von Neumann entropy and complexity with respect to a quantum Turing machine. The results provide two bounds with respect to two variants: an approximate-scheme complexity and a finite accuracy complexity.

In this subsection we provide a quantum variant of Brudno's theorem with respect to quantum unitary complexity **Huc**. Differently from the quantum Turing machine results, the bounds provided below are for almost all n, invariant to the accuracy term ϵ.

We define the quasilocal C^* algebra $\mathcal{A}_\infty$, which differs only from $\mathcal{M}_\infty$ in that it is a doubly infinite product space over $\mathbb{Z}$. In particular, $\mathcal{A}$ is the C^* algebra of qubits, i.e. 2×2 matrices acting on $\mathbf{C}^2$. For finite $\Lambda \subset \mathbb{Z}$, $\mathcal{A}_\Lambda = \bigotimes_{z \in \Lambda} \mathcal{A}_z$.

The quasilocal C^* algebra $\mathcal{A}_\infty$ is defined to be the norm closure of $\bigcup_{\Lambda \subset \mathbb{Z}} \mathcal{A}_\Lambda$. For states Ψ over $\mathcal{A}_\infty$, we use $\Psi \restriction n$ to denote Ψ restricted to the finite subalgebra $\mathcal{A}_{\{1,\ldots,n\}}$ of $\mathcal{A}_\infty$. The right shift T is a $*$-automorphism on $\mathcal{A}_\infty$ uniquely defined by its actions on local observables $T : a \in \mathcal{A}_{\{m,\ldots,n\}} \mapsto \mathcal{A}_{\{m+1,\ldots,n+1\}}$. A quantum state Ψ is stationary if for all $a \in \mathcal{A}_\infty$, $\Psi(a) = \Psi(T(a))$. The set of shift-invariant states on $\mathcal{A}_\infty$ is convex and compact in the weak* topology. The extremal points of this set are called ergodic states. The mean entropy of a quantum state Ψ is $\lim_{n \to \infty} \frac{1}{n} S(\Psi \restriction n)$, where S is von Neumann's entropy.

10.7.1 Quantum Unitary Complexity

Theorem 64 *Let Ψ be an ergodic state with mean entropy h. For all $\delta > 0$, for almost all n, there is an orthogonal projector $P_n \in \mathcal{A}_n$ such that for all $\epsilon > 0$,*

1. *$\Psi \restriction n(P_n) > 1 - \delta$.*
2. *For all one dimensional projectors $p \leq P_n$, $\mathbf{Huc}^\epsilon(p)/n \in (h - \delta, h + \delta)$.*

Proof. Let $\delta' < \delta'' < \delta$. From [BDK+05], there is a sequence of projectors $P'_n \in \mathcal{A}_n$ where for almost all n, $\Psi \restriction n(P'_n) > 1 - \delta'$, for all one dimensional projectors $p' \leq P'_n$, $2^{-n(h+\delta')} < \Psi \restriction n(p') < 2^{-n(h-\delta')}$, and $2^{n(h-\delta')} < \mathrm{Tr} P'_n < 2^{n(h+\delta')}$. Let S'_n be the subspace that P'_n projects onto. Let R_n be the smallest subspace spanned by all pure states produced by a quantum unitary pair (V, g) where $\mathbf{K}((V, g)) + g < n(h - \delta'')$. Let Q_n be the projector onto R_n. By Lemma 16, $\mathrm{Dim}(R_n) < 2^{n(h-\delta'')}$. Let S_n be the largest subspace of S'_n that is orthogonal to R_n. Let P_n be the orthogonal projector onto S_n. So for sufficiently large n, $\Psi \restriction n(P_n) \geq \Psi \restriction n(P'_n) - \mathrm{Dim}(R_n) 2^{-n(h-\delta')} > 1 - \delta' - 2^{n(h-\delta'')} 2^{-n(h-\delta')} = 1 - \delta' - 2^{n(\delta'-\delta'')} > 1 - \delta$, for large enough n.

By definition, since P_n is orthogonal to R_n, for all ϵ, for all one dimensional projectors $p \leq P_n$, $\mathbf{Huc}^\epsilon(p) \geq n(h - \delta'') > n(h - \delta)$. Furthermore, all such p can be produced from an elementary quantum unitary pair $(V, \lceil n(h(+\delta')) \rceil)$ that maps $\lceil n(h(+\delta')) \rceil$ qbit pure states and padded zeros with unitary transform V. Therefore for large enough n, $\mathbf{Huc}^\epsilon(p) \leq \mathbf{K}((V, \lceil n(h + \delta') \rceil)) + \lceil n(h + \delta') \rceil <^+ \mathbf{K}(\Psi, n, h) + \lceil n(h + \delta') \rceil < n(h + \delta)$. □

10.7.2 Gács Complexity

Theorem 65 *Let Ψ be an ergodic state with mean entropy h. For all $\delta > 0$, for almost all n, there is an orthogonal projector $P_n \in \mathcal{A}_n$ such that $\Psi \restriction n(P_n) > 1 - \delta$ and for all one dimensional projections $p \leq P_n$, $\mathbf{Hg}(p)/n <^+ h + \delta$.*

Proof. From [BDK+05], there is a sequence of projectors $P_n \in \mathcal{A}_n$ where for almost all n, $\Psi \restriction n(P_n) > 1 - \delta$, and $2^{n(h-\delta)} < \mathrm{Tr} P_n < 2^{n(h+\delta)}$. Let $p \leq P_n$ be a one dimensional projection. $\mathbf{Hg}(p) =^+ -\log \mathrm{Tr} p \boldsymbol{\mu} <^+ \log \mathrm{Tr} p \mathbf{m}(P_n | n(P_n 2^{-n(h+\delta)})) <^+ n(h + \delta) + \mathbf{K}(\delta, \Psi)$. So for large enough n, $\mathbf{Hg}(p)/n <^+ h + \delta + \mathbf{K}(\delta, \Psi)/n <^+ \mathbf{Hg}(p)/n <^+ h + \delta$. □

10.8 Measurement Systems

We note that pre-measures are of the form $\gamma : \{0,1\}^* \to \mathbb{R}_{\geq 0}$, where $\gamma(x) = \gamma(x0) + \gamma(x1)$. By Carathéodory's Extension Theorem, each such pre-measure can be uniquely extended to a measure Γ over $\{0,1\}^\infty$. In Chapter 9, measurements of finite collections of qubits are studied. This section deals with measurement systems, which can be applied to infinite quantum states.

Definition 23 (Meaurement System ([Bho21a])) *An α-computable measurement system $B = \{(|b_0^n\rangle, |b_1^n\rangle)\}$ is a sequence of orthonormal bases for $\mathcal{Q}_1$ such that each $|b_i^n\rangle$ is elementary and the sequence $\langle |b_1^n\rangle, |b_0^n\rangle\rangle_{n=1}^\infty$ is α-computable.*

Note that the above definition can be generalized to a sequence of PVMs. We now define the application of a measurement system B to an infinite quantum state Z which produces a pre-measure p. Let ρ_n be the density matrix associated with $Z\restriction n$. For the first bit, we use the standard definition of measurement, where

$$p(i) = \mathrm{Tr}\, |b_i^1\rangle \langle b_i^1| \, \rho_1.$$

Given ρ_2, if i is measured on the first bit, then the resulting state would be

$$\rho_2^i = \frac{(|b_i^1\rangle \langle b_i^1| \otimes I)\rho_2(|b_i^1\rangle \langle b_i^1| \otimes I)}{\mathrm{Tr}(|b_i^1\rangle \langle b_i^1| \otimes I)\rho_2}$$

So

$$\begin{aligned} p(ij) &= p(i)p(j|i) \\ &= \left(\mathrm{Tr}\, |b_i^1\rangle \langle b_i^1| \, \rho_1\right) \mathrm{Tr}\left(I \otimes |b_j^2\rangle \langle b_j^2|\right) \left(\frac{\left(|b_i^1\rangle \langle b_i^1| \otimes I\right) \rho_2 \left(|b_i^1\rangle \langle b_i^1| \otimes I\right)}{\left(|b_i^1\rangle \langle b_i^1| \otimes I\right) \rho_2} \right) \end{aligned}$$

Since $\mathrm{Tr}_2 \rho_2 = \rho_1$, $\mathrm{Tr}\, |b_i^1\rangle \langle b_i^1| \, \rho_1 = \mathrm{Tr}\left(|b_i^1\rangle \langle b_i^1| \otimes I\right) \rho_2$. Therefore

$$p(ij) = \mathrm{Tr}\rho_2 \left(|b_i^1 b_j^2\rangle \langle b_i^1 b_j^2|\right).$$

More generally for $x \in \{0,1\}^n$, we define the pre-measure p to be

$$p(x) = \mathrm{Tr}\rho_n \, |\otimes_{i=1}^n b_{x_i}^i\rangle \langle \otimes_{i=1}^n b_{x_i}^i| \, ;$$

It is straightforward to see that p is a pre-measure, with $p(x) = p(x0) + p(x1)$. Let μ_Z^B be the measure over $\{0,1\}^\infty$ derived from the described pre-measure, using measurement system B and state Z. We recall that MLR is the set of Martin Löf random sequences.

Definition 24 (Bhojraj Random) *A state Z is Bhojraj Random if for any computable measurement system B, $\mu_Z^B(\mathrm{MLR}) = 1$.*

Theorem 66 ([Bho21a]) *All NS Random states are Bhojraj Random states,*

Proof. Let state Z be NS random. Let $\{\rho_n\}$ be the density matrices associated with Z. Suppse not. Then there is $\delta \in (0,1)$ and computable measurement system $B = \{|b_0^n\rangle, |b_1^n\rangle\}_{n=1}^\infty$ where $\mu_Z^B(\{0,1\}^\infty \setminus \mathrm{MLR}) > \delta$. Let $\{S^m\}$ be a universal ML test. Without loss of generality, this test is of the form

$$S^m = \bigcup_{m \leq i} [\![A_i^m]\!],$$

where $[\![A_i^m]\!] \subseteq [\![A_{i+1}^m]\!]$, and $A_i^m\{\tau_1^{m,i}, \ldots, \tau_{k^{m,i}}^{m,i}\} \subset \{0,1\}^i$ for some $0 \leq k^{m-i} \leq 2^{i-m}$. Thus $\mu(S^m) \leq 2^{-m}$, where μ is the uniform distribution over $\{0,1\}^\infty$. We define an NS test as follows. For all m and i, with $m \leq i$, let $\tau_a = \tau_a^{m,i}$ and define the special projection

$$p_i^m = \sum_{a \leq k^{m,i}} |\otimes_{q=1}^i b_{\tau_\alpha[q]}^q\rangle \langle \otimes_{q=1}^i b_{\tau_\alpha[q]}^q| .$$

We define $P^m = \{p_i^m\}_{m \leq i}$ we have that $\langle P^m \rangle$ is an NS Test. The special tests p_i^m is uniformly computable in i and m since B and A_i^m are uniformly computable in i and m. Since $[\![A_i^m]\!] \subseteq [\![A_{i+1}^m]\!]$, $\text{Range}(p_i^m) \subseteq \text{Range}(p_{i+1}^m)$. So P^m is an NS Σ_1^0 set for all m. Since $k^{m,i} \leq 2^{i-m}$ for all m and i, this implies $\tau(P^m) \leq 2^{-m}$ for all m.

For all m, $\{0,1\}^\infty \setminus \text{MLR} \subseteq S^m$. Since by assumption $\mu_Z^B(\{0,1\}^\infty \setminus \text{MLR}) > \delta$, for all m there exists $i(m) > m$ such that

$$\mu_Z^B([\![A_{i(m)}^m]\!]) > \delta.$$

Fix an m and $i = i(m)$ and let $A_i^m = \{\tau_1, \ldots, \tau_{k^{m,i}}\}$, where $k^{m,i} \leq 2^{i-m}$. Let μ be the pre-measure associated with μ_Z^B. So we have

$$\delta < \sum_{a \leq k^{m,i}} \mu(\tau_a) = \sum_{a \leq k^{m,i}} \text{Tr}\rho_i \, |\otimes_{q=1}^i b_{\tau[q]}^q\rangle \langle \otimes_{q=1}^i b_{\tau[q]}^q| .$$

So we see that for all m there is an i such that

$$\delta < \text{Tr}\rho_i p_i^m \leq Z(P^m).$$

So $\inf_m Z(P^m) > \delta$, contradicting that Z is NS random. □

Theorem 67 ([Bho21a]) *There are states that are Bhojraj random and not NS Random.*

10.9 NS Solovay States

A NS Solovay test is a sequence of NS Σ_0^1 sets $\langle G_n \rangle$ such that $\sum_n \tau(G^n) < \infty$. A state Z fails a quantum NS test $\langle G^r \rangle$ at order $\delta \in (0,1)$ if there is an infinite number of $r \in R$ such that $\inf_{r \in R} Z(G^r) > \delta$. Otherwise state Z passes the quantum NS test at order δ. A quantum state Z is NS Solovay random if it passes all NS Solovay tests at all orders. The following theorem shows the equivalence of NS randomness and NS Solovay randomness with respect to every order δ. Given a special projection p, NS Σ_0^1 set $Q = \{q_n\}$, and state Z, we define $Z(p \setminus Q) = \inf_n Z(p \setminus q_n)$. In [Bho21a], it was proven that NS randomness is equivalent to NS Solovay randomness.

Proposition 8 *Given a special projection p, NS Σ_0^1 set Q, and state Z, $Z(p) - Z(Q) \leq Z(p \setminus Q) \leq Z(p)$.*

The proof is straightforward.

Theorem 68 *If a state Z fails an NS test at order δ then it fails an NS Solovay test at order δ.*

Proof. Assume that state Z fails a NS test $\langle G^r \rangle$ at order δ. Since $\sum_r \tau(G^r) \leq 1$, and each G^r is an NS Σ_1^0 set, $\langle G^r \rangle$ is a NS Solovay test. Furthermore since $\inf_r Z(G^r) \geq \delta$, there exists an infinite number of r such that $Z(G^r) > \delta$. Thus Z fails a NS Solovay test at order δ. □

Theorem 69 *For all $\delta' < \delta$, if a state Z fails an NS Solovay test at order δ then it fails an NS test at order δ'.*

Proof. Assume state Z fails NS Solovay test $\langle G^r \rangle$ at order δ. Given $\langle G^r \rangle$, where $G^r = \langle p_n^r \rangle_{n\in\mathbb{N}}$, we construct an NS test $\langle H^r \rangle$ as follows. There exists an m such that $\sum_{n>m} \tau(G^n) \leq 1$. Fix r. Enumerate all unordered sets of $r+1$ natural numbers $\{D_n^r\}_{n\in\mathbb{N}}$, $D_n^r \subset \mathbb{N}$, with infinite repetition.

$$H^r = \{q_n^r\}, q_n^r = \bigvee_{\ell<n} q_\ell^r \bigvee \left(\bigwedge_{t\in D_n^r} p_n^t \right).$$

Each H^r can be seen to be an NS Σ_1^0 set. In addition, $1 \geq \sum_{m<n} \tau(G^n) \geq 2^r \tau(H^r)$, so $\langle H^r \rangle$ is an NS test. For each r, $Z(H^r) > \delta'$. Assume not. Then there exists a k such that $Z(H^k) \leq \delta'$. Since Z fails $\langle G^r \rangle$ at order δ, there exists an infinite number of $r \in R$ and $n_r \in \mathbb{N}$ such that $Z(p_{n_r}^r) \geq \delta''$, for some $\delta' < \delta'' < \delta$. We reorder the NS Solovay test $\langle G^r \rangle$ such that r ranges over solely R. Let $z^r = p_{n,r}^r$. Let $D_{n,\ell}$ be the set of all unordered subsets of $\{1,\ldots,n\}$ of size ℓ. For $\ell > n$ let $F_{n,\ell} = \emptyset$. Let

$$F_{n,\ell} = \left(\bigvee_{A\in D_{n,\ell}} \bigwedge_{r\in A} z^r \right) \setminus \bigvee_{s>\ell} F_{n,s}.$$

So for all $n \in \mathbb{N}$, using Proposition 8,

$$\begin{aligned} & n(\delta'' - \delta') \\ \leq & \sum_{r=1}^{n} (Z(z^r) - Z(H^k)) \\ \leq & \sum_{r=1}^{n} Z(z^r \setminus H^k) \\ \leq & \sum_{r=1}^{n} Z\left(\bigvee_{s=1}^{k} F_{n,s} \wedge z^r \right). \end{aligned}$$

Let $F_{n,s,r} = F_{n,s} \wedge z^r$, with for a fixed $s \leq k$, $\sum_{i=1}^{n} Z(F_{n,s,i}) \leq s$.

$$\begin{aligned} & n(\delta'' - \delta') \\ \leq & \sum_{r=1}^{n} Z\left(\bigvee_{s=1}^{k} F_{n,s,r} \right) \\ = & \sum_{s=1}^{k} \sum_{r=1}^{n} Z\left(F_{n,s,r} \right) \\ \leq & \sum_{s=1}^{k} s = O(k^2). \end{aligned}$$

This is a contradiction for large enough n. □

Corollary 16 *A quantum state is NS random if and only if it is NS Solovay random.*

Chapter 11

Conclusions About Quantum Mechanics

11.1 Signals from Classical and Quantum Sources

Information non-growth laws say information about a target source cannot be increased with randomized processing. In classical information theory, where I is mutual information, and g is a random function, one has

$$I(g(X):Y) \leq I(X:Y).$$

where g is a randomized function, X and Y are random variables, and I is the mutual information function. Thus processing a channel at its output will not increase its capacity. Information conservation carries over into the algorithmic domain, with the inequalities (as seen in Chapter 2)

$$\mathbf{I}(f(x):y) <^+ \mathbf{I}(x:y); \qquad \mathbf{I}(f(a);\mathcal{H}) <^+ \mathbf{I}(a;\mathcal{H}).$$

These inequalities ensure target information cannot be obtained by processing. If for example the second inequality was not true, then one can potentially obtain information about the halting sequence $\mathcal{H}$ with simple functions. Obtaining information about $\mathcal{H}$ violates the Independence Postulate, discussed in Chapter **??**. Information nongrowth laws can be extended to signals [Eps23a] which can be modeled as probabilities over $\mathbb{N}$ or Euclidean spac. In [Eps23a] probabilities over $\{0,1\}^\infty$ and T_0 second countable topologies were also studied. The "signal strength" of a probability p over $\mathbb{N}$ is measured by its self information.

$$\mathbf{I}_{\text{Prob}}(p:p) = \log \sum_{i,j} 2^{\mathbf{I}(i:j)} p(i)p(j).$$

A signal, when undergoing randomized processing f (see Section 2.2), will lose its cohesion. Thus any signal going through a classical channel will become less coherent.

$$\mathbf{I}_{\text{Prob}}(f(p):f(p)) <^+ \mathbf{I}_{\text{Prob}}(p:p).$$

In Euclidean space, probabilities that undergo convolutions with probability kernels will lose self information. For example a signal spike at a random position will spread out when convoluted with the Gaussian function, and lose self information. The above inequalities deal with classical transformations. One can ask, is whether, quantum information processing can add new surprises to how information signals occur and evolve.

One can start with the prepare-and-measure channel, also known as a Holevo-form channel. Alice starts with a random variable X that can take values $\{1, \ldots, n\}$ with corresponding probabilities $\{p_1, \ldots, p_n\}$. Alice prepares a quantum state, corresponding to density matrix ρ_X, chosen from $\{\rho_1, \ldots, \rho_n\}$ according to X. Bob performs a measurement on the state ρ_X, getting a classical outcome, denoted by Y. Though it uses quantum mechanics, this is still a classical channel $X \to Y$. So using the above inequality, cohesion will deteriorate regardless of $X's$ probability, with

$$\mathbf{I}_{\text{Prob}}(Y : Y) <^{+} \mathbf{I}_{\text{Prob}}(X : X).$$

There remains a second option, constructing a signal directly from a mixed state. This involves constructing a mixed state, i.e. density matrix σ, and then performing a measurement E on the state, inducing the probability $E\sigma(k) = \text{Tr}\sigma E_k$. However from Theorem 4, for elementary (even enumerable) probabilities $E\sigma$,

$$\mathbf{I}_{\text{Prob}}(E\sigma : E\sigma) <^{+} \mathbf{K}(\sigma, E).$$

Thus for simply defined density matrices and measurements, no signal will appear. So experiments that are simple will result in simple measurements, or white noise. However it could be that a larger number of uncomputable pure or mixed states produce coherent signals. Theorems 53 and 54 say otherwise, in that the POVM measurement E of a vast majority of pure and mixed states will have negligible self-information. Thus for uniform distributions Λ and μ over pure and mixed states (see Section 8.2.2),

$$\int 2^{\mathbf{I}_{\text{Prob}}(E|\psi\rangle : E|\psi\rangle)} d\Lambda = O(1); \qquad \int 2^{\mathbf{I}_{\text{Prob}}(E\sigma : E\sigma)} d\mu(\sigma) = O(1).$$

The measure μ is a uniform measure over mixed states. This can be seen as a consequence of the vastness of Hilbert spaces as opposed to the limited discriminatory power of quantum measurements. In addition, there could be non-uniform distributions of pure or mixed states that could be of research interest. In quantum decoherence, a quantum state becomes entangled with the environment, losing decoherence. The off diagonal elements of the mixed state become dampened, as the state becomes more like a classical mixture of states. Let p_σ be the idealized classical probability that σ decoheres to, with $p_\sigma(i) = \sigma_{ii}$. Corollary 14 states that for an overwhelming majority of pure or mixed states σ, p_σ is noise, that is, has negligible self-information.

$$\int 2^{\mathbf{I}_{\text{Prob}}(p_{|\psi\rangle} : p_{|\psi\rangle})} d\Lambda = O(1); \qquad \int 2^{\mathbf{I}_{\text{Prob}}(p_\sigma : p_\sigma)} d\mu(\sigma) = O(1).$$

This is to be expected, with one supporting fact being for an n qubit space, $i \in \{1, \ldots, 2^n\}$, $\mathbf{E}_\Lambda[p_{|\psi\rangle}(i)] = 2^{-n}$. With Algorithmic Information Theory, this result was taken this fact one step further, showing that $p_{|\psi\rangle}$ has no (in the exponential) self-algorithmic information and cannot be processed by deterministic or randomized means to produce a more coherent signal. In addition, it appears a more direct proof of the first decoherence inequality could be possible.

However the measurement process has a surprising consquence, in that the wave function collapse causes an massive uptake in algorithmic signal strength. Let F be a PVM (defined in Chapter 9), of size 2^{n-c}, of an n qubit space and let Λ_F be the distribution of pure states when F is measured over the uniform distribution Λ. Thus Λ_F represents the F-collapsed states from Λ. Theorem 55 states

$$n - 2c <^{\log} \log \int 2^{\mathbf{I}_{\text{Prob}}(F|\psi\rangle : F|\psi\rangle)} d\Lambda_F.$$

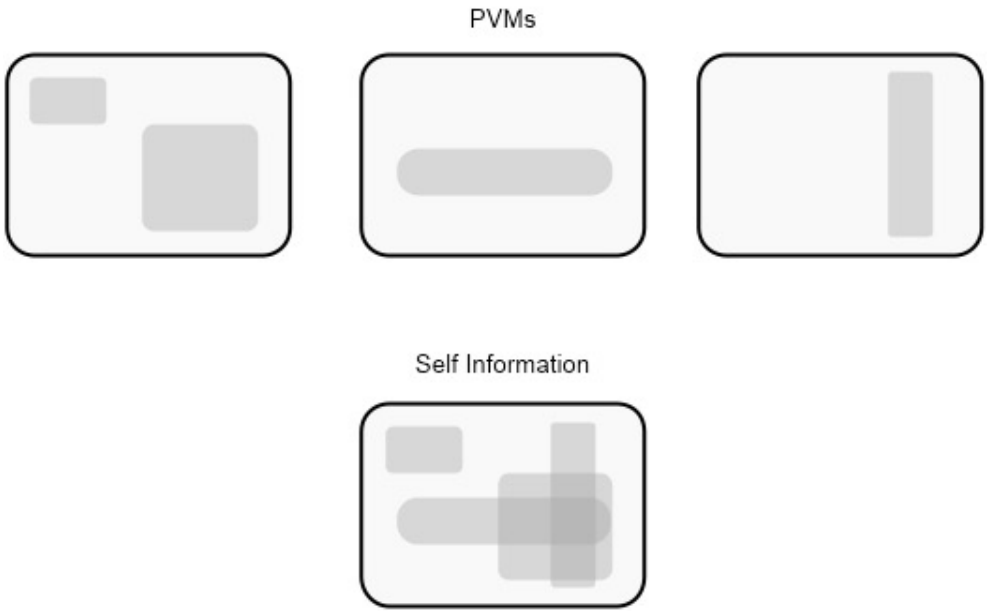


Figure 11.1: Each box on the top row represents an n qubit Hilbert space, with the shaded rectangles being the subspaces of the PVM projectors. Thus there are three PVMs. The self-information majorizes these subspaces, inversely weighted by the PVM's complexity.

11.2 Apriori Distributions

To avoid the pitfall of a signalless distribution that only produces white noise, we can conjecture a new apriori distribution for quantum states that is not signalless. Note that we are dealing with measures over the density operator space and not directly with density operators because we are measuring properties, such as self-information, over all possible (pure or mixed) states. Research into distributions over operators can be found in [SBC01]. Properties of this apriori distribution can be discerned by working backwards. Indeed, suppose there are a set of (possibly infinite) systems $\{|\psi_i\rangle\}$, where for each system $|\psi_i\rangle$, a measurement occurs, producing a discernable signal. By Theorem 52, this implies the states $|\psi_i\rangle$ have high $\mathbf{I}(|\psi_i\rangle : |\psi_i\rangle)$, where $\mathbf{I}$ is the information function between mixed states introduced in Definition 16. Thus any universal quantum apriori distribution over these systems must be weighted toward states with high self information. One candidate is an probability measure ξ over pure states where

$$\xi(|\psi\rangle) \propto 2^{\mathbf{I}(|\psi\rangle:|\psi\rangle)}.$$

However this area of research is still ongoing. Another clue to this universal quantum apriori distribution is the measurement operation, which as shown above, causes an uptake in signal strength. Take a PVM measurement F, which procures a value i from a state $|\psi\rangle$, projecting to a new state $|\psi'\rangle$. $P|\psi'\rangle(i) = 1$. By Exercise 12, this new state $|\psi'\rangle$ has self information

$$\mathbf{K}(i) <^{\log} \mathbf{I}(|\psi'\rangle : |\psi'\rangle).$$

The error term is on the order of $\mathbf{K}(P)$. Most of the measurement values i of P will be random, i.e. have large $\mathbf{K}(i)$ (just look at the Kolmorogov complexity of the first 2^n numbers!). Thus simple

quantum measurements increase the self information of most measured quantum states (see Figure 11.1). So this fact, and Theorem 47, leads us to the following conclusion.

> *Take a distribution over density operators, such as Λ, where an overwhelming majority of states have negligible self-information. When each such state in its support is measured with a simple apparatus, the result is new a distribution where most of the states have substantial self-information.*

However, the situation is reversed for quantum channels. A quantum state that is transformed by a quantum operation will not increase in self-information. So by Theorem 49, we get the following claim, where equality occurs if the quantum operation is a unitary transform.

> *Given any distribution over density operators, if all the density matrices its support are transformed by a simple quantum operation, then the resultant distribution will give more measure to mixed states with less self-information.*

Thus simple measurements with many operators can only increase self-information, simple quantum operations can only decrease self-information, and simple unitary transforms leave the self-information unaltered. If the operation is complex, then nothing so far has been proven.

11.3 Measurements Before Information Cloning

The no-cloning theorem states that every unitary transform cannot clone an arbitrary quantum state. However there is the possibility of copying information from a subset of states. By "copying information", we mean that two measurements of two states will produce two values that are similar. More formally, the information cloned from a state $|\psi\rangle$ relative to unitary transform U, and POVMs E and F is,

$$\mathbf{I}_{\text{Clone}}(|\psi\rangle) = \mathbf{I}_{\text{Prob}}(E\,|\phi_1\rangle : F\,|\phi_2\rangle), \text{ where } U\,|\psi\rangle\,|0\rangle = |\phi_1\rangle\,|\phi_2\rangle\,.$$

This represents the shared signal strength between $|1\rangle$ and $|2\rangle$ when the states[1] were created from a unitary transform U of $|\psi\rangle$ tensored with an ancillia state $|0\rangle$. Note that by Theorems 50 and 52, cloneable information is less than self information, with

$$\mathbf{I}_{\text{Clone}}(|\psi\rangle) <^{\log} \mathbf{I}(|\psi\rangle : |\psi\rangle).$$

The question is, given an initial distribution over density operators with low expected $\mathbf{I}_{\text{Clone}}$, what sort of transform is required to increase this expectation. In this section, we discuss necessary conditions of this transform. We require the following two assumptions.

Assumption (1): The initial distribution has low expected self information. Theorem 47 shows there is a large set of natural distributions that have this property. Any distribution Ω that is less than $2^c\Lambda$ will have $\log \int 2^{\mathbf{I}(|\psi\rangle:|\psi\rangle)} d\Omega <^+ c$. Another way to intrepret this assumption is through parmeterized distributions. Let P be a probability over parameters θ over pure state distribution, $\Gamma(|\psi\rangle\,|\theta)$. The distribution is balanced, where $\int \Gamma(|\psi\rangle\,|\theta) dP(\theta) = \Lambda(|\psi\rangle)$. Then because of Theorem 47,

$$P(\{\theta : \mathbf{E}_{|\psi\rangle\sim\Gamma(\cdot|\theta)}[\mathbf{I}(|\psi\rangle : |\psi\rangle)] \geq m\}) \leq 2^{-m+1}.$$

[1] Note this definition can be generalized to arbitrary states, with $\mathbf{I}_{\text{Prob}}(E\text{Tr}_2\sigma : F\text{Tr}_1\sigma)$, where $\sigma = \varepsilon(|\psi\rangle)$, for quantum operation ε.

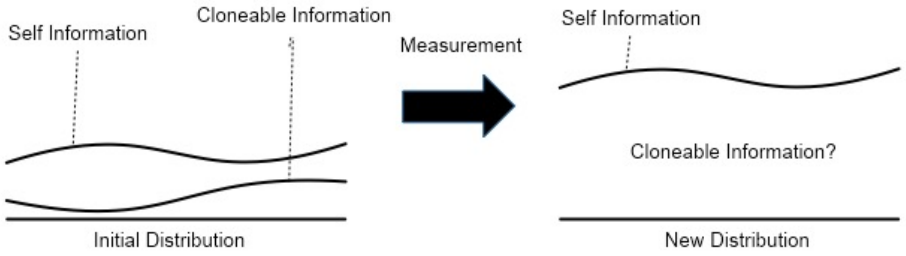


Figure 11.2: The intial distribution has low self information and cloneable information. A measurement increases the self information and potentially increases the cloneable information.

Assumption (2): The universal Turing machine is relativized to all the transforms and operators. This assumption states that for a system, the operations are known quantities. This is congruent with quantum information theory, in which actors are seen to compute unitary transforms or quantum operations. It is asumed that these actors have knowledge of the transforms.

How do you create a distribution with high expected $\mathbf{I}_{\text{Clone}}$, where most states can have cloneable information? Any transform that increases cloneable information must increase self-information. However Theorem 49, along with assumption (2) bars quantum operations as a means to create self-information, as the complexity of the quantum operations is $O(1)$. Thus the only way to potentially increase self-information is to perform a measurement, which as Theorems 52 and 55 show, often times cause an uptake in self-information (see Figure 11.2). This is also discussed in the quotes of Section 11.2. Thus we get the following claim.

> *Measurements are required to produce distributions over quantum states that have cloneable information.*

For example, take the starting distribution to be the uniform measure over pure states, Λ. Let $E = F = \{|i\rangle\langle i|\}$ be POVM measurements over projectors to the basis states and let U be any unitary transform such that $U|i\rangle|0\rangle = |i\rangle|i\rangle$ for $i \in \{1, \ldots, 2^n\}$. By Theorems 47 and 50, we have that

$$\int 2^{\mathbf{I}_{\text{Clone}}(|\psi\rangle)} d\Lambda = O(1).$$

Now suppose we apply the measurement $G = E$ to Λ, producing a new distribution Λ_G concentrated evenly among the basis states, where $\Lambda_G(|i\rangle) = 2^{-n}$. Thus we have that $\mathbf{I}_{\text{Clone}}(|i\rangle) = \mathbf{I}_{\text{Prob}}(E|i\rangle : F|i\rangle) = \mathbf{K}(i)$. Since there are $2^{n-O(1)}$ basis states $|i\rangle$ where $n <^+ \mathbf{K}(i)$, we have the following uptake in cloneable information.

$$n <^+ \log \int 2^{\mathbf{I}_{\text{Clone}}(|\psi\rangle)} d\Lambda_G.$$

Other such applications can be seen as generalizations from this extreme example. Future work involves determining how tightly self information covers cloneable information.

Part III

Thermodynamics

Chapter 12

Introduction to Algorithmic Thermodynamics

Classical thermodynamics is the study of substances and changes to their properties such as volume, temperature, and pressure. Substances, such as a gas or a liquid, is modeled as a point in a phase space. The phase space, $\mathcal{X}$, is modeled by a computable metric space, [HR09], and a volume measure μ, is modeled by a computable (not necessarily probabilistic) positive measure over $\mathcal{X}$. The continuous dynamics are modeled by a one dimensional transformation group G^t, indexed by $t \in \mathbb{R}$. Due to Louville's theorem, the dynamics are measure-preserving, where $\mu(G^t A) = \mu(A)$, for all Borel sets $A \subseteq \mathcal{X}$. Discrete dynamics are modelled by functions $\mathcal{X} \mapsto \mathcal{X}$, and considerable attention is spent on the ergodic case.

Whether quantum or classical, the known laws of physics are reversible. Thus the dynamics G of our system are also reversible, in that if $\beta = G^t\alpha$, then there is some t' such that the original state can found with $\alpha = G^{t'}\beta$. Thus if given a set of particles with position and velocity, by reversing the velocities, a previous state can be found. This is contradiction to the second law of thermodynamics, which states,

> *The total entropy of a system either increases or remains constant in any spontaneous process; it never decreases.*

This conforms to our experiences of broken vases never reforming. To reconcile this difference, Boltzmann introduced *macro-states*, Π_i, indexed by $i \in \mathbb{N}$, which groups states together by macroscopic parameters, with corresponding Boltzmann entropy $S(\Pi_i) = k_B \ln \mu(\Pi_i)$. By definition, a vast majority of typical states will experience an increase in Boltzmann entropy.

In [Gac94], coarse grained entropy was introduced as an algorithmic update to Boltzmann entropy. This formulation was made to be independent of the choice of parameters of the macro state. In this manuscript, a modified version of coarse grained entropy is introduced. We also model the thermodynamic entropy of a micro-state with algorithmic methods. The micro-state of a system contains the information of the entire physical state. For example, the microstate of a system of N molecules is a point

$$(q_1, \ldots, q_{3N}, p_1, \ldots, p_{3N}) \in \mathbb{R}^{6N}$$

where q_i are the position coordinates and p_i are the momentum coordinates. The set of states, $\mathbb{R}^{3N}$ is a computable metric space. To model the entropy of the state, we use slight variant to algorithmic fine-grained entropy H_μ in [Gac94], using symbol $\mathbf{H}_\mu$. This entropy measure captures

the level of disorder of the state. Continuing the example above, if all the particles are at rest, then the thermodynamic entropy of the state of

$$(q_1, \ldots, q_{3N}, 0, \ldots, 0)$$

is expected to be very low.

> *The evolution of the system will be thermodynamic like if it spends most of the time close to its maximum value, from which it exhibits frequent small fluctuations and rarer large fluctuations.*

In this paper, using the algorithmic definition of thermodynamic entropy, $\mathbf{H}_\mu$, we prove that such fluctuations *have to* occur, and the greater fluctuations are more rare.

Throughout the part on thermodynamics, conditions will be proven for states, sets of states, dynamics, and measures to have high or infinite mutual information in the halting sequence. Assuming the Independence Postulate (see Chapter **??**), such constructs with high information are non-physical and can be considered non-realizable in nature. This also goes for the part of Quantum Mechanics, in particular the Quantum EL Theorem 19 and the Quantum Outlier Theorem 44. The thermodynamics section of this manuscript is arranged as follows.

- **Chapter 13:** Computable metric spaces and their relation to randomness is detailed. This material is the foundation for which algorithmic coarse and fine grained entropy is based upon. This chapter is a modification to the work in [HR09] to abritrary positive measures and dual measure spaces.
- **Chapter 14:** Algorithmic fine grained entropy is introducd. This is a modification to the definition in [Gac94], using computable measure theory. An entropy balance lemma is proven with applications to Maxwell's demon. We detail a result from [Gác21] that algorithmic fine grain entropy is conformant to the addition additive equality analagous to that of string algorithmic information theory.
- **Chapter 15:** In this chapter, algorithmic fine grained entropy is proved to oscillate in the presence of dynamics, regardless of the choice of phase space and volume measure.
- **Chapter 16:** Discrete dynamics are studied in this chapter. It is proved that given a phase space and two different volume measures of it, the algorithmic thermodynamic entropy will oscillate in a synchronized fashion with respect to both volume measures.
- **Chapter 17:** It is shown in the course of dynamics on product spaces, for typical states, the marginal entropies cannot be in sync. This is true for continuous dynamics and discrete ergodic dynamics.
- **Chapter 18:** Algorithmic coarse grained entropy is defined and shown to be an excellent approximation to algorithmic fine grained entropy. Algorithmic coarse grained entropy is proved to oscillate in the presence of dynamics.
- **Chapter 19:** In this chapter, computability properties of algorithmic thermodynamics is studied. It is proved that there is no computable means to select times where dynamics produce states of continually decreasing entropy. It is also proved that algorithmic fine grained entropy cannot be approximated.
- **Chapter 20:** In this chapter a stochastic transition function is defined, in the spirit of stochastic thermodynamics. A conservation inequality is proven.

Chapter 13

Computable Measure Spaces

The results in the manuscript of thermodynamics uses computable metric spaces and computable measure spaces. We use the constructs from[HR09], where in the paper, they noted physics an application of the paper. In [HR09, Gác21], computable metric spaces $\mathcal{X}$ were paired with a space of Borel probability measures $\mathcal{M}(\mathcal{X})$ and a computable measure is defined as a constructive point in $\mathcal{M}(\mathcal{X})$. Computable probabilities can lower compute the measure of open sets and the integral lower semi-continuous functions. A key insight in [HR09] is that there exists morphisms mapping $\mathcal{X}$ to the Cantor space $\{0,1\}^\infty$. This is known as a binary represention, and they are used in the proofs of many theorems in the Thermodynamics section of the manuscript. This chapter introduces some generalizations to [HR09], including generalizing from probability measures to arbitrary nonnegative finite measures, from computable measure spaces to dual computable measure spaces and from binary representations to dual binary representations.

Definition 25 *A computable metric space consists of a triple $(X, \mathcal{S}, d)$, where*

- *X is a separable complete metric space.*
- *$\mathcal{S}$ is an enumerable list of dense ideal points $\mathcal{S}$ in X.*
- *d is a distance metric that is uniformly computable over points in $\mathcal{S}$.*

For $x \in X$, $r \in \mathbb{Q}_{>0}$ a ball is $B(x,r) = \{y : d(x,y) < r\}$. The ideal points induce a sequence of enumerable ideal balls $B_i = \{B(s_i, r_j) : s_i \in \mathcal{S}, r_j \in \mathbb{Q}_{>0}\}$. We have $\overline{\overline{B}}(s_i, r_j) = \{y : d(x,y) \leq r\}$, which might not equal the closure of $B(s_i, r_j)$ if there are isolated points. A sequence of ideal points $\{x_n\} \subseteq Y$ is said to be a fast Cauchy sequence if $d(x_n, x_{n+1}) < 2^{-n}$ for all $n \in \mathbb{N}$. A point x is computable there is a computable fast Cauchy sequence converging to x. Each computable function f between computable metric spaces X and $\mathcal{Y}$ has an algorithm $\mathfrak{A}$ such that if $f(x) = y$ then for all fast Cauchy sequences $\overrightarrow{x}$ for x, $\mathfrak{A}(\overrightarrow{x})$ outputs an encoding of a fast Cauchy sequence for y.

Proposition 9 *For $x \in X$, the following statements are equivalent.*

1. *x is a computable point.*
2. *Each $d(x, s_i)$ are upper semi-computable uniformly in i.*
3. *$d_x = d(x, \cdot) \to \mathbb{R}^+$ is a computable function.*

Exercise 15 *Define computable metric spaces over the Cantor space, Euclidean space, and a product space.*

13.1 Enumerative Lattices

An enumerative lattice is the tuple $(X, \leq, \mathscr{P})$ where $(X, \leq)$ is a complete lattice and $\mathscr{P}$ is a numbered set such that if $x \in X$ then $x = \sup P$, for some $P \subseteq \mathcal{P}$. An element $x \in X$ is called lower-computable if there is some enumeration of $\{p : p \leq x\}$.

Proposition 10 *Let $(X, \leq, \mathscr{P})$ be an enumerative lattice. There is an enumeration of all its lower-computable elements .*

Proof. Using the universal Turing machine, enumerate all enumerable sets. That is, for each enumerable E, there is an i such that $E = \{U(i, n) : n \in \mathbb{N}\}$. Thus for each lower-computable element x has an enumerated set E such that $x = \sup\{p_i : i \in E\}$. □

Definition 26 *Given two enumerative lattices, Y and Z, a function $f : Y \to Z$ is Scott continuous if it is monotonic and $\sup f(\overrightarrow{p}) = f(\sup \overrightarrow{p})$ for every increasing sequence $\overrightarrow{p}$. We say f is bi-lower-computable, if there exists a computable method that given an enumerable sequence $\overrightarrow{p} = (p_{n_1}, p_{n_2}, \dots)$ with $y = \sup \overrightarrow{p}$, outputs a sequence $\overrightarrow{q} = \{q_{n_1}, q_{n_2}, \dots\}$ such that $f(y) = \sup \overrightarrow{q}$.*

Proposition 11 *If a function $f : Y \to Z$ is Scott-continuous and all $f(\sup(p_{n_1}, \dots, p_{n_k}))$ are lower-computable uniformly in $(n_1, \dots, n_k)$, then f is bi-lower-computable.*

Proof. Let $\overrightarrow{p} = (p_{n_1}, p_{n_2}, \dots)$ be a sequence such that $y = \sup \overrightarrow{p} \in Y$. An algorithm works with access to $\overrightarrow{p}$ works as follows. For all k, it lower computes $f(\sup(p_{n_1}, p_{n_2}, \dots)$, which is possible due to the assumption of the Proposition. The supremum of this sequence is $\sup_k f(\sup\{p_{n_1}, \dots, p_{n_k}\})$, which is lower computable due to f being Scott continuous. Thus the enumerated sequence is a lower description of $f(y)$. □

Definition 27 *Given a computable metric space $(X, d, \mathscr{L})$ and an enumerative lattice $(Y, \leq, \mathscr{P})$, we denote $\mathscr{F}$ to be the step functions from X to Y, where*

$$f_{(i,j)}(x) = \begin{cases} p_j & \text{if } x \in B_i \\ \bot & \text{otherwise.} \end{cases}$$

We define $\mathscr{C}(X, Y)$ as the closure of $\mathscr{F}$ under pointwise suprema, with pointwise ordering $\sqsubseteq$. It immediately follows that $(\mathscr{C}(X, Y), \sqsubseteq, \mathscr{F}))$ is an enumerative lattice. A function $f : X \to Y$ is lower-computable if it is a lower-computable element of the enumerative lattice. $(\mathscr{C}(X, Y), \sqsubseteq, \mathscr{F}))$.

Example 4 *The set $\overline{\mathbb{R}}^+ = [0, \infty) \cup \{\infty\}$ has an enumerative lattice $\left(\overline{\mathbb{R}}^+, \leq, \mathbb{Q}^+\right)$ which induces a lattice $\mathscr{S}(X, \overline{\mathbb{R}}^+)$ of positive lower semi-continuous functions from X to $\overline{\mathbb{R}}^+$. Its lower-computable elements are the lower semi-computable functions.*

Definition 28 *A subset A of X is semi-decidable if it is an r.e. open set.*

Proposition 12 *Let (X, d_X, S_X) and (Y, d_y, S_y) be computable metric spaces. A partial function $f : D \subseteq X \to Y$ is computable if and only if the preimages of ieal balls are uniformly r.e. open in D sets. So for all i, $f^{-1}(B_i) = U_i \cap D$ where U_i is an r.e. open set uniformly in i.* □

13.2 Computing with Measures

The computable metric space of all Borel probability measures over X is $\mathcal{M}(X)$. If X is separable and compact then so is $\mathcal{M}(X)$. The ideal points of $\mathcal{M}(X)$ are $\mathscr{D}$, the set of probability measures that are concentrated on finitely many points with rational values. The distance metric on $\mathcal{M}(X)$ is the *Prokhorov metric*, defined as follows.

Definition 29 (Prokhorov metric)

$$\pi(\mu,\nu) = \inf\left\{\epsilon \in \mathbb{R}^+ : \mu(A) \leq \nu(A^\epsilon) \text{ for Borel set } A\right\},$$

where $A^\epsilon = \{x : d(x,A) < \epsilon\}$.

In thermodynamics, the measure function representing the volume is not necessarily a probability measure. Thus the results of [HR09] needs to be extended to nonnegative measures of arbitrary size to prove a result about thermodynamics. Let $(\mathbb{R}^+, d_\mathbb{R}, \mathbb{Q}^+)$ be the computable metric space where $\mathbb{R}^+ = [0,\infty)$ is the complete separable metric space and nonnegative rationals $\mathbb{Q}^+$ consists of the ideal points. The distance function is $d_\mathbb{R}(x,y) = |x-y|$, which is obviously computable over $\mathbb{Q}_{\geq 0}$. The space of nonnegative Borel measures over a computable metric space is the space $\mathfrak{M}(X) = \mathcal{M}(X) \times \mathbb{R}_{\geq 0}$, the product space of the space of probability measures of X, $\mathcal{M}(X)$, with the space of nonnegative reals. We identify a measure with a pair $(\mu, m) \in \mathfrak{M}(X)$ where $\mu \in \mathcal{M}(X)$ is a Borel probability measure over X, and $m \in \mathbb{R}^+$ is the size of the measure.

The distance function of $\mathfrak{M}$ is

$$d_\mathfrak{M}((\mu,m),(\nu,n)) = \max\{\pi(\mu,\nu), d_\mathbb{R}(n,m)\},$$

where π is the Prokhorov metric (see Definition 29). The ideal points of $\mathfrak{M}(X)$ is the set $\mathscr{D}_\mathfrak{M}$ of all finite points with nonnegative rational values. This definition is different from the ideal points in $\mathcal{M}(X)$ in that they don't have to sum to 1. The computable measures of $\mathfrak{M}(X)$ are its constructive points, with respect to a fast Cauchy description.

Proposition 13 *The tuple* $(\mathfrak{M}(X), d_\mathfrak{M}, \mathscr{D}_\mathfrak{M})$ *is a computable metric space.*

Proof. Let (μ_i, v_i) and (μ_j, v_j) be two ideal points of $(\mathfrak{M}(X), d_\mathfrak{M}, \mathscr{D}_\mathfrak{M})$, where μ_i and μ_j are two probability measures over X, assigning rational measure to a finite number of ideal points. In addition $v_i, v_j \in \mathbb{Q}^+$. If U is a r.e. open subset of X, $\mu_i(U)$ is lower-computable uniformly in i and U. This is because of $(s_{n_1}, q_{m_1}), \ldots, (s_{n_k}, q_{m_k})$ are the mass points of μ_i with their weights then $\mu_i(U) = \sum_{s_{n_j} \in U} q_{m_j}$. As all $s_{n_j} \in U$ can be enumerated from a description of U this sum is lower computable. So $\mu_i(B_{i_1} \cup \ldots B_{i_k})$ is lower-computable and $\mu_i(\overline{\overline{B}}_{i_1} \cup \ldots \overline{\overline{B}}_{i_k})$ is upper semi-computable, uniformly in i and $(i_1, \ldots, i_k)$.

We show that $\pi(\mu_i, \mu_j)$ is computable uniformly in (i,j). Since μ_i is an ideal measure concentrated over S_i, we have $\pi(\mu_i,\mu_j) = \inf\{\epsilon \in \mathbb{Q} : \forall A \subset S_i, \mu_i(A) < \mu_j(A^\epsilon) + \epsilon\}$. Since μ_j is an ideal measure and A^ϵ is a finite union of open ideal balls, $\mu_j(A^\epsilon)$ is lower computable, uniformly in ϵ and j, so $\pi(\mu_i,\mu_j)$ is upper computable, uniformly in (i,j). The term $\pi(\mu_i,\mu_j)$ is lower computable, uniformly in (i,j) because $\pi(\mu_i,\mu_j) = \sup\{\epsilon \in \mathbb{Q} : \exists A \subset S_i, \mu_i(A) > \mu_j(A^{\overline{\epsilon}}))\}$, with $A^{\overline{\epsilon}} = \{x : d(x,A) \leq \epsilon\}$, and using the upper semi-computability of $\mu_j(A^{\overline{\epsilon}})$.

In addition, it easy to see that $d_\mathbb{R}(v_i, v_j)$ is computable. Thus the following term is computable.

$$d_\mathfrak{M}((\mu_i,v_i),(\mu_j,v_j)) = \max\{\pi(\mu_i,\mu_j), d_R(v_i,v_j))\}.$$

□

Claim 1 *If $(\mu, m) \in \mathfrak{M}(X)$ is a computable measure, then $(\mu, 1) \in \mathfrak{M}(X)$ is computable as well. This follows from taking the fast Cauchy sequence for (μ, m) and normalizing each ideal point in the series.*

For a metric space X, let τ be the set of all open sets of X. The valuation operator $v : \mathfrak{M}(X) \times \tau \to \mathbb{R}^+$ maps $((\mu, m), U)$ to $m\mu(U)$. More formally, for the first argument, v takes a $\mathfrak{M}(X)$ fast Cauchy sequence to a measure (μ, m), $m \in \mathbb{R}^+$, and a sequence of ideal balls B_i such that $U = \bigcup_i B_i$ and outputs $\{x : x < m\mu(U)\}$.

Proposition 14 *The valuation operator v is bi-lower computable, in the second argument.*

Proof. Since $v((\mu, m), \cdot)$ is Scott-continuous, due to Proposition 11, the proof is satisfied if we show that $v((\mu, m), \cdot)$ is uniformly lower-computable on finite union of ideal balls. For ideal probability measure $\mu_i \in \mathcal{M}(X)$, due to the proof of Proposition 13, $\mu_i(B_{i_1} \cup \cdots \cup B_{i_k})$ is lower computable, uniformly in $(i, i_1, \ldots, i_k)$.

Let $((\mu_{k_n}, m_n))_{n\in\mathbb{N}}$ be a description of a (not necessarily probability) measue $(\mu, m) \in \mathfrak{M}(X)$. Thus $\pi(\mu_{k_n}, \mu) \leq \epsilon_n$ and $|m_n - m| \leq \epsilon_n$, where $\epsilon_n = 2^{-n+1}$. For $n \in \mathbb{N}$ and $U = B(s_{i_1}, q_{j_1}) \cup \cdots \cup B(s_{i_k}, q_{j_k})$ we have

$$U_n = \bigcup_{m\leq k} B(s_{i_m}, q_{j_m} - \epsilon_n).$$

We have $U_{n-1}^{\epsilon_n} \subseteq U_n$ and $U_n^{\epsilon_n} \subseteq U$, where $A^\epsilon = \{x : d(x : A) < \epsilon\}$. We will show that $\mu(U) = \sup_n(\mu_{j_n}(U_n) - \epsilon_n)$. Since $\pi(\mu_{j_n}, \mu) \leq \epsilon_n$, and we have that $\mu_{j_n}(U_n) \leq \mu(U) + \epsilon_n$ for all n, so $\mu(U) \geq \sup_n(\mu_{j_n}(U_n) - \epsilon_n)$. Similarly, we have $\mu(U_{n-1}) \leq \mu_{j_n}(U_n) + \epsilon_n$, for all n. So as $n \to \infty$, $\mu(U) \leq \sup_n(\mu(U_{n-1}) - 2\epsilon_n) \leq \sup_n(\mu_{j_n}(U_n) - \epsilon_n)$. Thus $\mu(U) = \sup_n \mu_{j_n}(U_n) - \epsilon_n$ is lower computable. In addition $m = \sup_n m_n - \epsilon_m$ is lower computable $v((\mu, m), U) = m\mu(U)$ is lower computable, uniformly in $(i, i_1, \ldots, i_k)$. □

Proposition 15 *For measure (μ, m), if m is computable and measure $\mu(B_{i_1} \cup \ldots B_{i_k})$ is uniformly lower computable in $(i_1, \ldots, i_k)$ then (μ, m) is computable.*

Proof. We show that $\pi(\mu_n, \mu)$ is upper computable uniformly in n and then apply Proposition 9. Since $\pi(\mu, \mu) < \epsilon$ iff $\mu_n(A) < \mu(A^\epsilon) + \epsilon$ for all $A \subset S_n$ where S_n is the finite support of μ_n, and $\mu(A^\epsilon)$ is lower computable (as A^ϵ is a finite union of open ideal balls) $\pi(\mu_n, \mu)$ is semi-decidable, uniformly in n and ϵ. Furthermore, for any ideal point $(\mu_n, m_n) \in \mathfrak{M}(\mathcal{X})$, since m is computable $d_{\mathfrak{M}}((\mu_n, m_n), (\mu, m)) = \max\{\pi(\mu_n, \mu), d_{\mathbb{R}}(m_n, m)\}$ is upper computable so Proposition 9 can be applied. Thus one can construct a fast sequence of ideal measures converging to (μ, m). □

For the Cantor space $\{0, 1\}^\infty$ with the standard metric space structure, the ideal balls are the cylinders $x\{0, 1\}^\infty$, for $x \in \{0, 1\}^*$.

Corollary 17 *If a measure $(\mu, m) \in \mathfrak{M}(\{0, 1\}^\infty)$ is computable and m is computable, then the cylinders are uniformly computable.*

Proposition 16 *The integral operator $\int : \mathfrak{M}(X) \times \mathscr{C}(X, \mathbb{R}^+) \to \overline{\mathbb{R}}^+$ is bi-lower computable, in the second argument.*

Proof. Let $(\mu, m) \in \mathfrak{M}(X)$. The integral of a finite supremum of steps functions can be expressed by induction on the number functions, starting with $m \int f_{(i,j)} d\mu = m q_j \mu(B_i)$ and

$$m \int \sup \left\{ f_{(i_1,j_1)}, \ldots, f_{(i_k,j_k)} \right\} d\mu = m q_{j_z} \mu(B_{i_1} \cup \cdots \cup B_{i_k}) + m \int \sup \left\{ f_{(i_1,j'_1)}, \ldots, f_{(i_k,j'_k)} \right\} d\mu$$

where $q_{j_z} = \min\{q_{j_1}, \ldots, q_{j_k}\}$ and $q_{j'_i} = q_{j_i} - q_{j_z}$. Since $f_{(i_z,j'_z)}$ is zero, it can be removed. It is easy to see that m can be computed, and by Proposition 14, the measure of finite balls can be lower computed, uniformly in $(B_{i_1}, \ldots, B_{i_m})$. For any measure (μ, m), the operator $m \int d_\mu : \mathscr{C}(X, \overline{R}^+) \to \overline{\mathbb{R}}^+$ is Scott continuous, so by Proposition 11, the opertor is bi-lower computable. □

Corollary 18 *Let $(f_i)_i$ be a sequence of uniformly computable functions, such that the function $(i, x) \mapsto f_i(x)$ is computable. If f_i is bounded by M_i computable uniformly in i, then the function $((\mu, m), i) \mapsto m \int f_i \mu$ is computable.*

Proof. $f_i + M$ and $M_i - f_i$ are unifomly lower computable, so $m \int f_i d\mu = m \int (f + M_i) d\mu - m M_i = m M_i - m \int (M_i - f_i) d\mu$ is lower and upper computable by Proposition 16.

13.3 Computable Measure Space

Definition 30

1. *A dual measure space $(\mathcal{X}, (\mu, m), (\nu, n))$ is a computable metric space $\mathcal{X}$ and two computable Borel measures, (μ, m), and (ν, n) over $\mathcal{X}$. A measure space $(\mathcal{X}, (\mu, m))$ is dual measure space $(\mathcal{X}, (\mu, m), (\mu, m))$.*

2. *A constructive G_δ-set is a set of the form $\bigcap_n U_n$ where $(U_n)_n$ is a sequence of uniformly r.e. open sets.*

3. *For computable measure space $(\mathcal{X}, (\mu, m))$ and computable metric space $\mathcal{Y}$, a function $f :\subset (\mathcal{X}, (\mu, m)) \to \mathcal{Y}$ is almost computable if it is computable on a constructive G_δ set of measure m.*

4. *A morphism of computable probability spaces $Q : (\mathcal{X}, (\mu, m)) \to (\mathcal{Y}, (\nu, m))$ is an almost computable measure-preserving function $Q : D_Q \subset \mathcal{X} \to \mathcal{Y}$, where $\mu(Q^{-1}(A)) = \nu(A)$ for all Borel sets A. An isomorphism (Q, R) is a pair of morphisms such that $Q \circ R = \mathrm{id}$ on $R^{-1}(D_Q)$ and $R \circ Q = \mathrm{id}$ on $Q^{-1}(D_R)$.*

5. *A dual binary representation of dual computable measure space $(\mathcal{X}, (\mu, m), (\nu, n))$ is a tuple $(\delta, \mu_\delta, \nu_\delta)$ where (μ_δ, m) and (ν_δ, n) are computable (not necesssarily probability) measures on $\{0,1\}^\infty$ and $\delta : (\{0,1\}^\infty, (\mu_\delta, m)) \to (\mathcal{X}, (\mu, m))$ and $\delta : (\{0,1\}^\infty, (\nu_\delta, n)) \to (\mathcal{X}, (\nu, n))$ are surjective morphisms. Denoting $\delta^{-1}(x)$ to be the set of expansion of $x \in X$:*

   - *There is a dense full-measure constructive G_δ-set D of points have a unique expansion.*
   - *$\delta^{-1} : D \to \delta^{-1}(D)$ is computable.*
   - *(δ, δ^{-1}) is an isomorphism.*

6. *A binary representation of computable measure space $(\mathcal{X}, (\mu, m))$ is a dual representation of the dual computable measure space $(\mathcal{X}, (\mu, m), (\mu, m))$.*

7. *A set A is almost decidable with respect to measures (μ, m) and (ν, n) if there are two. r.e. open sets U and V such that $U \subset A$, $V \subseteq A^{\mathscr{C}}$, $U \cup V$ is dense and has full μ and ν measure. We say the elements of a sequence $\{A_i\}$ are uniformly almost decidable with respect to (μ, m) and (ν, n) if there are two sequences $\{U_i\}$ and $\{V_i\}$ of uniformly r.e. sets satisfying the above conditions.*

Exercise 16 *Explicitly define a binary expansion for the space $([0,1], L)$, where L is the Lebesgue measure.*

The follow proof of existence of an almost decidable set is from [GHR11].

Lemma 17 *Let X be $\mathbb{R}$ or $\mathbb{R}^+$ or $[0,1]$. Let (μ, m) and (ν, n) be a computable measures on X. Then there is a sequence of uniformly computable reals $(x_n)_n$ which is dense in X and such that $\mu(\{x_n\}) = \nu(\{x_n\}) = 0$ for all n.*

Proof. Let I be a closed rational interval. Let $M = \max\{m, n\}$. We construct $x \in I$ such that $\mu(\{x\}) = \nu(\{x\}) = 0$. To do this, we construct inductively a nested sequence of closed intervals J_k of μ and ν measure $< M2^{-k+1}$, with $J_0 = I$. Suppose $J_k = [a, b]$ has been constructed, with $\mu(J_k) < M2^{-k+1}$ and $\nu(J_k) < 2^{-k+1}$. Let $m = (b-a)/16$ and $\ell = (b-a)/64$: by the Markov inequality one of the intervals $[a + jm + \ell, a + (j+1)m - \ell]$ $j \in \{0, \ldots, 15\}$ must have μ and ν measure $< M2^{-k}$ and since these measures are upper computable, it can be found effectively, and we denote it J_{k+1}. By enumerating all dydadic intervals $(I_n)_n$, one can constuct $x_n \in I_n$ uniformly. □

Corollary 19 *Let $(\mathcal{X}, (\mu, m), (\nu, m))$ be a dual measure space and $(f_i)_i$ be a sequence of uniformly computable real valued functions on X. There is a sequence of uniformly computable reals $(x_n)_n$ which is dense in $\mathbb{R}$ such that $\mu(\{f_i^{-1}(x_n)\}) = \nu(\{f_i^{-1}(x_n)\}) = 0$ over all i, n.*

Proof. We define the uniformly computable measure (μ_i, m) where $\mu_i = \mu \circ f_i^{-1}$ and (ν_i, n) where $\nu_i = \mu \circ f_i^{-1}$. Define measure (λ, m), $\lambda = \sum 2^{-i}\mu_i$ and (γ, n), $\gamma = \sum 2^{-i}\nu_i$. By Propositio 15, (λ, m) and (γ, m) are computable measures so by Lemma 17 there is a sequence of uniformly computable reals $(x_n)_n$ which is dense in $\mathbb{R}$ such that $\lambda(\{x_n\}) = \gamma(\{x_n\}) = 0$ for all i, n. □

Corollary 20 *There is a sequence of uniformly computable reals $(r_n)_{n\in\mathbb{N}}$ such that $(B(s_i, r_i))$ is a basis of almost decidable balls.*

Proof. Apply Corollary 19 to $(f_i)_i$ defined by $f_i(x) = d(s_i, x)$. □

Every ideal ball can be expressed as a r.e. union of almost decidable balls, and vice-versa. So the two bases are constructively equivalent.

Definition 31 *A set D is an ad-set if it is a finite union of almost decidable balls, with $D = B_{i_1} \cup \ldots B_{i_k}$. We have $\overline{\overline{D}} = \overline{\overline{B}}_{i_1} \cup \ldots \overline{\overline{B}}_{i_k}$, which may differ than the closure of D if there are isolated points.*

Proposition 17 *For computable measure space $(\mathcal{X}, \mu)$ and ad-set D, $\mu(D)$ is computable.*

Proof. This follows from Proposition 15, which implies $\mu(D)$ and $\mu(X \setminus \overline{\overline{B}})$ being lower computable, noting the fact that all almost decidable balls have borders of null measure.

We fix computable measures (μ, m) and (ν, n), and their computable representations. We denote $B(s_i, r_n)$ by B_k where $k = \langle i, n\rangle$ and r_n is the sequence defined in 20. Let $C_k = X \setminus \overline{\overline{B}}(s_i, r_n)$. For $w \in \{0,1\}^*$, the cell $\Gamma(w)$ is defined by $\Gamma(\epsilon) = X$, $\Gamma(w0) = \Gamma(w) \cap C_i$ and $\Gamma(w1) = \Gamma(w) = \cap B_i$, where ϵ is the empty word and $i = \|w\|$. This is an almost decidable set, uniformly in w.

Theorem 70 *Every dual measure space $(\mathcal{X}, (\mu, m), (\nu, n))$ has a dual binary representation.*

Proof. We construct an encoding function $b : D \to \{0,1\}^\infty$, a decoding function $\delta : D_\delta \to X$, and show that δ is a multi binary representation, with $b = \delta^{-1}$.

Let $D = \cap_i B_i \cup C_i$. The set D is a full-measure constructive G_δ-set. Define the computable function $b : D \to \{0,1\}^\infty$ with

$$b(x)_i = \begin{cases} 1 & \text{if } x \in B_i \\ 0 & \text{if } x \in C_i. \end{cases}$$

Let $x \in D$: $\omega = b(x)$ is also characterized by $\{x\} = \cap_i \Gamma(\omega_{0\ldots i-1})$. b can be computed from $\Gamma(\cdot)$. Let (μ_δ, m) and (ν_δ, n) computable measures over $\{0,1\}^\infty$, where $\mu_\delta = \mu \circ b^{-1}$, and $\nu_\delta = \nu \circ b^{-1}$. Let D_δ be the set of binary sequences ω such that $\cap_i \overline{\Gamma(\omega_{0\ldots i-1})}$ is a singleton. The decoding function $\delta : D_\delta \to X$ is defined by

$$\delta(\omega) = x \text{ if } \cap_i \overline{\Gamma(\omega_{0\ldots i-1})} = \{x\}.$$

The next steps are to prove that δ is a surjective morphism. The center and radius of the ball B_i will be s_i and r_i, respectively. We say n is an i-witness for ω if $r_i < 2^{-(n+1)}$, $\omega[i] = 1$, and $\Gamma(\omega[0..i]) \neq \emptyset$. We firts prove that

$$D_\delta = \cap_n \{\omega \in \{0,1\}^\infty : \omega \text{ has a } n\text{-witness}\}.$$

Let $\omega = D_\delta$ and $x = \delta(\omega)$ For every n, $x \in D(s_i, r_i)$ for some i with $r_i \leq 2^{-(n+1)}$. Since $x \in \overline{\Gamma(\omega[0\ldots i])}$, we have $\Gamma(\omega[0\ldots i]) \neq \emptyset$ and $\omega[i] = 1$. So i is an n-witness for ω. Conversely if ω has a n-witness i_n for all n, since $\overline{\Gamma[0\ldots i_n]} \subseteq \overline{\overline{B}}_{i_n}$ with radius going to zero, the sequence $\overline{\Gamma(\omega[0\ldots n])}$ of closed balls has a non-empty intersection, due to the completeness of the space, and it a singleton.

$\delta : D_\delta \to X$ is computable. For each n, find an n-witness i_n of ω: the sequence $(s_{i_n})_n$ is a fast Cauchy sequence converge to $\delta\omega$). In addition, δ is surjective: each $x \in X$ has at least one expansion. We construct by induction a sequence $\omega = \omega[0]\omega[1]\ldots$ such that for all i, $x \in \overline{\Gamma(\omega[0\ldots i])}$. Let $i \geq 0$ andd suppose that $\omega[0\ldots i-1]$ has been constructed. Since $B_i \cup C_i$ is open and dense and $\Gamma(\omega 0\ldots i-1)$ is open, $\overline{\Gamma(\omega_{0\ldots i-1})} = \overline{\Gamma(\omega_{0\ldots i-1}) \cap (B_i \cup C_i)} = \overline{\Gamma\omega_{0\ldots i-1}0} \cup \overline{\Gamma\omega_{0\ldots i-1}1}$, so for some $\omega[i] \in \{0,1\}$, has $x \in \overline{\Gamma(\omega_{0\ldots i})}$. So $x \in \cap_i \overline{\Gamma(\omega_{0\ldots i-1})}$. Since $(B_i)_i$ is a basis and $\omega_i = 1$ whenever $x \in B_i$, ω is an expansion of x.

13.4 Randomnesss

Definition 32 *For a measure $(\mu, m) \in \mathfrak{M}(X)$, a (μ, m) ML randomness test is a sequence of unifomly r.e. open sets $(U_n)_n$, satisfying $m\mu(U_n) \leq 2^{-n}$. The set $\cap_n U_n$ is a null measure set and is called an μ-effective null set. An alternative definition of null sets uses integrals (see [GÁ1]), with a slight modification as measures are being used. Given a measure $(\mu, m) \in \mathfrak{M}(X)$ a μ-randomness test is a (μ, m) computable element of $\mathscr{C}(X, \overline{R}^+)$ such that $m \int t d\mu \leq 1$ Any subset of $\{x \in X : t(x) = \infty\}$ is called a μ-effective null set. The two definitions of null sets are equivalent.*

A point $x \in X$ is (μ, m)-ML random it is in no effective null set. A uniform randomness test is a computable function T from $\mathfrak{M}(X)$ to $\mathscr{C}(X, \overline{\mathbb{R}}^+)$ such that $m \int T^{(\mu,m)} d\mu \leq 1$.

Using proposition 10, let $(H_i)_{i\in\mathbb{N}}$ be an enumeration of all lower computable elements of the enumerative lattice $\mathscr{C}(\mathfrak{M}(X), \mathscr{C}(X, \overline{\mathbb{R}}^+)))$, such as $H_i \sup_k f_\phi$ where $\phi : \mathbb{N}^2 \to \mathbb{N}$ is some recursive dunction and f_n are step functions.

Lemma 18 *There is a computable function $T : \mathbb{N} \times \mathfrak{M}(X) \to \mathscr{C}(X, \overline{R}^+)$ with*

- *For all i, $T_i = T(i, \cdot)$ is a uniform randomness test.*
- *If $\int m H_i((\mu, m)) d\mu < 1$ for some (μ, m), then $T_i(\mu) = H_i(\mu)$.*

Proof. To enumerate only tests, we'd like to be able to semi-decide $m \int \sup_{k<n} f_{\phi(i,k)}((\mu, m)) d\mu < 1$. But $m \sup_{k<n} f_{\phi(i,k)}((\mu, m))$ is only lower computable (relative to (μ, m)). Let $\mathcal{Y}$ be a computable metric space. For an ideal point $s \in Y$ and positive rations q, r, ϵ, define the hat function:

$$h_{q,s,r,\epsilon}(y) = q[1 - [d(y, s) - r]^+/\epsilon]^+,$$

where $[a]^+ = \max\{0, a\}$. This is a continuous function whose value is q in $B(s, r)$ and 0 outside $B(s, r + \epsilon)$. It is easy to see there is a number $(h_n)_{n\in\mathbb{N}}$ of all the hat functions. They are equivalent to step function in the enumerative lattice $\mathscr{C}(Y, \overline{\mathbb{R}}^+)$. The step functions can be constructed as the supremum of such function $f_{(i,j)} = \sup\{h_{q_j,s,r-\epsilon,\epsilon:0<\epsilon<r}\}$ with $B_i = B(s, r)$ and conversely.

We let $Y = \mathfrak{M}(X) \times X$ endowed with the canonical computable metric structure. By "curryfication" it provides functions $h_n \in \mathscr{C}(\in \mathfrak{M}(X), \mathscr{C}(X, \overline{R}^+))$ with which the H_i can be expressed: there is a recursive function $\phi : \mathbb{N}^2 \to \mathbb{N}$ such that for all i, $H_i = \sup_k h_{\phi(i,k)}$.

In addition, $h_n((\mu, m))$ is bound by a constant computable from n and independent of (μ, m). Hence, by Corollary 18, the integration operator $\int : \mathfrak{M}(X) \times \mathbb{N} \to [0, 1]$ which maps $((\mu, m)\langle i_1, \ldots, i_k\rangle)$ to $m \int \sup\{h_{i_1}((\mu, m)), \ldots, h_{i_k}((\mu, m))\} d\mu$ is computable. Thus $T(i, (\mu, m)) = \sup\{H_i^k((\mu, m)) : m \int H_i^k((\mu, m)) \leq 1\}$ where $H_i^k = \sup_{n<k} h_{\phi(i,n)}$. Since $m \int H_i^k((\mu, m)) d\mu$ can be computed from i, k, and a description of (μ, m), T is a computable function from $\mathbb{N} \times \mathfrak{M}(X)$ to $\mathscr{C}(X, \overline{\mathbb{R}}^+)$. □

Theorem 71 *There is a universal uniform randomness test, that is a uniform test $\mathbf{t}$ such that for every uniform test T, there is a constant $c > 0$ with $\mathbf{t} > \mathbf{m}(T)T$.*

Proof. Using Lemma 18, the universal test is defined by $\mathbf{t} = \sum_i \mathbf{m}(i) T_i$: since every T_i is a uniform randomness test, $\mathbf{t}$ is also a uniform randomness test. In addition, for every uniform test T, there is an i such that $T = T_i = H_i$. □

The following corollary is due to [Ǵ21] with the proofs adapted to uniform tests. Assume there is a fixed measure $(\nu, n) \in \mathfrak{M}(\mathcal{Y})$, where $\mathcal{Y}$ is a computable metric space. Let $F : \mathcal{X} \to \mathbb{Z} \cup \{-\infty\}$ be upper computable, where $\mathcal{X}$ is another computable metric space. An F randomness test R is a computable function from $\mathcal{X}$ to $\mathscr{C}(Y, \overline{R}^+)$ such that $n \int R^x d\nu \leq 2^{-F(y)}$.

Corollary 21 *There exists a universal F uniform test $\mathbf{r}$ such that $n \int \mathbf{r}^x d\nu \leq 2^{-F(x))}$ and for every F randomness test R, $\mathbf{m}(R|\langle \overrightarrow{x} \rangle) R_x \overset{*}{<} \boldsymbol{r}_x$.*

Proof. The proof follows analogously to that of Lemma 18, except $T(i,x,(\mu,m)) = \sup\{H_i^k(x,(\nu,n)) : n\int H_i^k xy,(\nu,n)) \leq 2^{-F(x)}\}$ where $H_i^k = \sup_{n<k} h_{\phi(i,n,\langle\overrightarrow{x}\rangle)}$. The term $\phi(i,n,\langle\overrightarrow{x}\rangle)$ is the partial recursive function being given the numbers i and n, and an encoding of a fast Cauchy sequence for $x \in \mathcal{X}$.

Definition 33 (Randomness Deficiency) *We recall that the deficiency of randomness of an infinite sequence $\alpha \in \{0,1\}^\infty$ with respect to a computable measure (P,p) over $\{0,1\}^\infty$ is defined to be*

$$\mathbf{D}(\alpha|(P,p),x) = \log \sup_n \mathbf{m}(\alpha[0..n]|x)/p{\cdot}P(\alpha[0..n]).$$

We have $\mathbf{D}(\alpha|(P,p)) = \mathbf{D}(\alpha|(P,p),\emptyset)$. By [Gác21], $2^{\mathbf{D}}$ is a lower-computable (P,p)-test, in that

$$p\int_{\{0,1\}^\infty} 2^{\mathbf{D}(\alpha|(P,p))} dP(\alpha) = O(1).$$

Proposition 18 *Let $(\{0,1\}^\infty,(\mu,m))$ be a computable measure space. Then $\mathbf{t}_{(\mu,m)}(\alpha) \stackrel{*}{=} 2^{\mathbf{D}(\alpha|(\mu,m))}$.*

Proof. There is a constructive element H of $\mathscr{C}(\in \mathfrak{M}(X), \mathscr{C}(X,\overline{R}^+))$, such that $H((\nu,n),\alpha) = 2^{\mathbf{D}(\alpha|(\mu,m))}$. Thus there is some i, such that $H_i = H$. Furthermore T_i is a uniform test satisfying $T_i((\mu,m)) = 2^{\mathbf{D}(\alpha|(\mu,m))}$ because $m\int T_i(\mu,m)d\mu = m\int H((\mu,m))\mathbf{d}\mu < 1$.

As described in [Gác21], $2^{\mathbf{D}(\alpha|(\mu,m))}$ is a universal lower computable test. Thus $\mathbf{t}_{(\mu,m)}(\alpha) \stackrel{*}{<} 2^{\mathbf{D}(\alpha|(\mu,m))}$.

Proposition 19 *For computable measure space $(X,(\mu,m))$, every random point lies in every r.e. open set of full measure.*

Proof. Let $U = \bigcup_{(i,j)\in E} B(s_i,q_j)$ be a r.e. open set of measure m, with $E \subseteq \mathbb{N}$ being r.e. Let F be the r.e. set $\{(i,k) : \exists j,(i,j) \in Eq_k < q_j\}$. Let

$$U_n = \bigcup_{\langle i,k\rangle\cap[0,n]} B(s_i,q_k) \text{ and } V_n^{\mathscr{C}} = \bigcup_{\langle i,k\rangle\cap[0,n]} \overline{\overline{B}}(s_i,q_k).$$

Then U_n and V_n are r.e. uniformly in n, $U_n \nearrow U$ and $U^{\mathscr{C}} = \cap_n V_n$. As $\mu(U_n)$ is lower semi-computable uniformly in n, a sequence $(n_i)_{i\in\mathbb{N}}$ can be computed such that $m\mu(U_{n_i}) > 1 - 2^{-i}$. Thus $m\mu(V_{n_i}) < 2^{-i}$ and $U^{\mathscr{C}} = \cap_i V_{n_i}$ is a μ-ML test. Thus every (μ,m)-random point is in U. □

Lemma 19 *Let $Q : D \subset X \to \mathcal{Y}$ be a morphism of equal computable measure spaces $(X,(\mu,m))$ and $(\mathcal{Y},(\nu,m))$, with universal tests $\mathbf{t}_{(\mu,m)}$ and $\mathbf{t}_{(\nu,m)}$. Then there is some c with the following properties. If $x \in X$ and $\mathbf{t}_{(\mu,m)}(x) < \infty$, then $Q(x)$ is defined and $\mathbf{t}_{(\nu,m)}(Q(x)) \stackrel{*}{<} c\mathbf{t}_{(\mu,m)}(x)$.*

Proof. Assuming $\mathbf{t}_{(\mu,m)}(x) < \infty$, then x is a random point then $x \in D$, because due to Propostion 19, every random point lies in every r.e. open set of full measure, and D is an intersection of full-measure r.e open sets. Thus $Q(x)$ is defined.

We have that $\mathbf{t}_{(\nu,m)} \circ Q \in \mathscr{C}(X,\overline{R}^+)$ because there is an algorithm that enumerates all finite prefixes of fast Cauchy sequences to Q and enumerates all resultant outputted ideal balls. Then the algorithm sees which outputted ideal balls B are in the values ideals balls (B',v) enumerated by $\mathbf{t}_{(\nu,m)}$. If $B \subseteq B'$, then the algorithm outputs (B,v).

Since $\mu(D) = 1$, $\int \mathbf{t}_{(\nu,m)} \circ Q d\mu$ is well defined. As Q is measure-preserving, $m\int \mathbf{t}_{(\nu,m)} \circ Q du = m\int \mathbf{t}_{(\nu,m)} d\nu \leq 1$. Hence $\mathbf{t}_{(\nu,m)} \circ Q$ is a μ-test, so there exists $c \in \mathbb{N}$ with $\mathbf{t}_{(\nu,m)} \circ Q \stackrel{*}{<} c\mathbf{t}_{(\mu,m)}$. □

Corollary 22 *Let $(Q, R) : (\mathcal{X}, \mu) \rightleftarrows (\mathcal{Y}, \nu)$ be an isomorphism of computable measure spaces, with universal tests $\mathbf{t}_\mu$ and $\mathbf{t}_\nu$. Then there is a $c \in \mathbb{N}$ where $\mathbf{t}_\nu(Q(x)) = \mathbf{t}_\mu(x) \pm c$ and $\mathbf{t}_\mu(R(y)) = \mathbf{t}_\nu(y) \pm c$.*

Chapter 14

Algorithmic Fine Grained Entropy

In this chapter we introduce the central term of algorithmic thermodynamics: algorithmic fine grain entropy $\mathbf{H}_\mu$. We also show some canonical properties of $\mathbf{H}_\mu$, originating from [Ǵ21], with modifications to the proofs as needed to be compatible with Chapter 13. In this chapter and in subsequent ones, we represent (not necessarily probabilistic) measures as μ, dropping the (μ, m) notation.

Definition 34 *Given a measure space $(\mathcal{X}, \mu)$, its corresponding algorithmic fine grained entropy is $\mathbf{H}_\mu(x) = -\log \mathbf{t}_\mu(x)$, where $\mathbf{t}$ is the universal uniform test introduced in Theorem 71.*

The term $\mathbf{H}_\mu$ is bounded from above by $\log \mu(X)$ and can take arbitrary negative values, including infinitely negative values. If x is in a μ constructive nullset then $\mathbf{H}_\mu(x) = -\infty$.

Definition 35 *For measure μ and lower continuous function f over metric space $\mathcal{X}$, $\mu^x f(x) = \int_{x\in\mathcal{X}} f(x) d\mu(x)$.*

Exercise 17 *Prove that if μ is a probability measure, then $\mathbf{H}_\nu(y) >^+ \mathbf{H}_{\mu,\nu}(x, y)$.*

Proof. $2^{-\mathbf{H}_\nu(y)}$ is a test for $\mu \times \nu$, since $\mu^x \nu^y 2^{-\mathbf{H}_\mu(y)} \leq \mu^x 1 = 1$. □

Proposition 20

(1) $\mathbf{H}_\mu(x|\lceil \log \mu(X) \rceil) <^+ \log \mu(X)$.

(2) $\mathbf{H}_\mu(x) <^+ \log \mu(X) + \mathbf{K}(\lceil \log \mu(X) \rceil)$.

Proof.

(1) We use the μ-test $t_\mu(x) = 1/\lceil \mu(X) \rceil$, where $\int t_\mu d\mu \leq 1$. Thus $-\log \mu(X) <^+ \log t_\mu(x) <^+ \log \mathbf{t}_\mu(x|\lceil \mu(X) \rceil)$.

(2) We use Proposition 26.

□

Definition 36 (Computable Transformation Group) *A one dimensional transformation group G^t, parameterized by $t \in \mathbb{R}$ over a measure space $(\mathcal{X}, \mu)$ where each G^t is a homeomorphism of $\mathcal{X}$ onto itself, where $G^t(G^s(x) = G^{t+s}(x)$. And $G^t x$ is continuously simultaneously in x and t. G is measure preserving, where $\mu(G^t(A)) = \mu(A)$, for all Borel sets A. Furthermore there is a program that when given an encoding of a fast Cauchy sequence of $t \in \mathbb{R}$ and $x \in \mathcal{X}$, outputs an encoding of a fast Cauchy sequence of $G^t x$.*

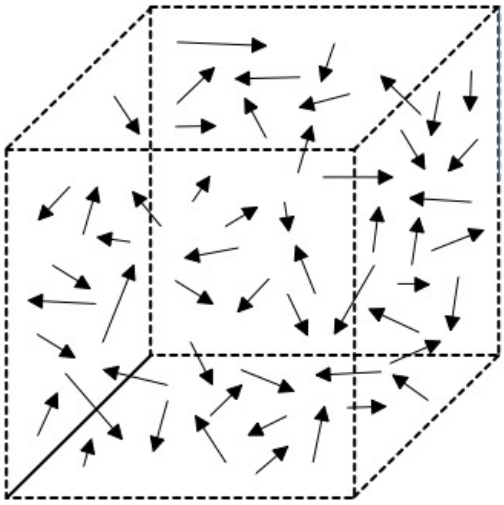


Figure 14.1: A canonical example for algorithmic fine grained entropy. The space consists of the momentum and position of N particles.

Example 5 (Particles in a Box) *We detail a canonical example: particles in a box, as depicted in Figure 14.1. Let there be a box containing N particles. Each particle has a position $p \in [0, M] \times [0, M] \times [0, M]$ and momentum $m \in [-M/2, M/2] \times [-M/2, M/2] \times [-M/2, M/2]$ and no other degrees of freedom. The state space X is the position and volume of all particles, and using the Lebesgue measure μ, the total volume is $\mu(X) = M^{6N}$. A state of this space can be represented as $\omega \in X$, with*

$$(\omega.q_1, \ldots, \omega.q_{3N}, \omega.p_1, \ldots \omega.p_{3N}).$$

When dynamics are applied to such a point, the particles will move from their positions according to their momentum. If ω contains a computable coordinate, either an x, y, z value of momentum or position, then $\mathbf{H}_\mu(\omega) = -\infty$. Indeed suppose it is a position coordinate k, where $\omega.p_k = \epsilon$ for some computable value $\epsilon \in [0, M]$ (it works similarly for a momentum coordinate). We define the test $t(\alpha) = \sum_n [|\alpha.p_n - \epsilon| < 2^{-n-1}]\mathbf{m}(n)2^n/M^{6N}$. Thus since ϵ is computable, so is t, and since $t(\omega) = \infty$, $\mathbf{H}_\mu(\omega) = -\infty$.

Exercise 18 *Prove that relativized to computable probability p over $\{0,1\}^*$, $\mathbf{H}_p(x) =^+ -\mathbf{d}(x|p)$.*

Proposition 21 *For rational $t \in \mathbb{Q}$, $\mathbf{H}_\mu(G^t x) - \mathbf{H}_\mu(x) <^+ \mathbf{K}(t)$.*

Proof. This is because, since G is measure preserving, $\mathbf{t}_\mu(G^{-t}\omega)$ is a μ-test of complexity $\mathbf{K}(t)$. Thus $\mathbf{m}(t)\mathbf{t}_\mu(G^{-t}\omega) \stackrel{*}{<} \mathbf{t}_\mu(\omega)$.

14.1 Thermodynamic Information

Information between a point of the metric space and a binary sequence is introduced as well as the information between two points in metric spaces. The term $\mathbf{H}_\mu(\alpha|t)$ is the fine grained algorithmic entropy of α when the universal Turing machine is relativized to the sequence t.

Definition 37 (Information) *Let $(\mathcal{X}, \mu)$ and $(\mathcal{Y}, \nu)$ be computable measure spaces. For $\alpha \in \mathcal{X}$, $\beta \in \mathcal{Y}$ and $t \in \{0,1\}^* \cup \{0,1\}^\infty$,*

- $\mathbf{I}(\alpha;t) = \mathbf{H}_\mu(\alpha) - \mathbf{H}_\mu(\alpha|t)$.
- $\mathbf{I}(\alpha:\beta) = \mathbf{H}_\mu(\alpha) + \mathbf{H}_\nu(\beta) - \mathbf{H}_{\mu\times\nu}(\alpha,\beta)$.

If dynamics are used to increase or decrease algorithmic thermodynamic entropy by a non trivial amount, then the encoded dynamics shares algorithmic information with the ending or starting state, respectively. Proposition 22 is due to [Gac94], with usage of algorithmic fine grained entropy.

> *If you want to increase the entropy of a state, you need information about its ending state and if you want to decrease the entropy of a state, you need information about its starting state.*

Proposition 22 $-\mathbf{I}(\alpha;t) <^+ \mathbf{H}_\mu(G^t\alpha) - \mathbf{H}_\mu(\alpha) <^+ \mathbf{I}(G^t\alpha;t)$.

Proof. By definition

$$\begin{aligned}\mathbf{I}(\alpha;t) &= \mathbf{H}_\mu(\alpha) - \mathbf{H}_\mu(\alpha|t)\\ \mathbf{I}(G^t\alpha;t) &= \mathbf{H}_\mu(G^t\alpha) - \mathbf{H}_\mu(G^t\alpha|t).\end{aligned}$$

Since the function $2^{-\mathbf{H}_\mu(G^t\alpha)}$ is a μ test,

$$\mathbf{H}_\mu(G^t\alpha) >^+ \mathbf{H}_\mu(\alpha|t) = \mathbf{H}_\mu(\alpha) - \mathbf{I}(\alpha;t)..$$

Which gives us the first inequality. In addition $2^{-\mathbf{H}_\mu(G^t\alpha)}$ is also a μ test, so

$$\mathbf{H}_\mu(G\alpha) >^+ \mathbf{H}_\mu(G^t\alpha|t) = \mathbf{H}_\mu(G^t\alpha) - \mathbf{I}(G^t\alpha;t),$$

which gives the second inequality. □

Proposition 23 (Conservation of Information) $\mathbf{I}(G^t\alpha:\beta) <^+ \mathbf{I}(\alpha:\beta)$.

Proof. We have that G^t and $G^t \times \mathbf{Id}$ are μ and $\mu\times\nu$ preserving so $2^{-\mathbf{H}_\mu(G^{-t}\alpha)}$ is a μ test and $2^{-\mathbf{H}_{\mu\times\nu}((G^t\times\mathbf{Id})(\alpha,\beta))}$ is a $\mu\times\nu$ test. So $\mathbf{H}_\mu(G^{-t}\alpha) >^+ \mathbf{H}(\alpha)$ implies $\mathbf{H}_\mu(\alpha) >^+ \mathbf{H}(G^t\alpha)$. And also $\mathbf{H}_{\mu,\nu}(\alpha,\beta) <^+ \mathbf{H}_{\mu,\nu}(G^t\alpha,\beta)$. So

$$\begin{aligned}\mathbf{I}(G^t\alpha:\beta) &= \mathbf{H}_\mu(G^t\alpha) + \mathbf{H}_n(\beta) - \mathbf{H}_{\mu,\nu}(G^t\alpha,\beta)\\ &<^+ \mathbf{H}_\mu(\alpha) + \mathbf{H}_n(\beta) - \mathbf{H}_{\mu,\nu}(\alpha,\beta).\end{aligned}$$

□

14.2 Entropy Balance

The following section is due to [Gac94]. Lets say there exists two independent systems $(\mathcal{X},\mu)$ and $(\mathcal{Y},\nu)$ represented as computable measure spaces that are put under joint dynamics G. We show that under mild assumptions, an increase of entropy in one subsystem implies a decrease is entropy in another system. Let $(\alpha_t,\beta_t) = G^t(\alpha,\beta)$, and $\Delta\mathbf{H}_\mu(\alpha) = \mathbf{H}_\mu(\alpha^t) - \mathbf{H}_\mu(\alpha)$, and similarly for $\Delta\mathbf{H}_\nu(\beta)$.

Lemma 20 $\Delta\mathbf{H}_\mu(\alpha) + \Delta\mathbf{H}_\nu(\beta) >^+ \mathbf{I}(\alpha_t:\beta_t) - \mathbf{I}(\alpha:\beta) - \mathbf{I}((\alpha,\beta);t)$.

Proof. Due to Proposition 22 applied to (α, β), $\Delta\mathbf{H}_{\mu\times\nu}(\alpha,\beta) >^+ -\mathbf{I}((\alpha,\beta);t)$. So

$$\begin{aligned}\mathbf{H}_\mu(\alpha_t) + \mathbf{H}_\nu(\beta_t) =& \mathbf{H}_{\mu\times\nu}(\alpha_t,\beta_t) + \mathbf{I}((\alpha_t:\beta_t)\\ >^+& \mathbf{H}_{\mu\times\nu}(\alpha,\beta) - \mathbf{I}((\alpha,\beta);t) + \mathbf{I}(\alpha_t:\beta_t)\\ =^+& \mathbf{H}_\mu(\alpha) + \mathbf{H}_\nu(\beta) + \mathbf{I}(\alpha_t:\beta_t) - \mathbf{I}(\alpha:\beta) - \mathbf{I}((\alpha:\beta);t).\end{aligned}$$

□

The last term is almost always negligible. If one wants to lower the thermodynamic entropy of a state, the information of the state must be encoded into the dynamics or an independent environment can be coupled with the system which will absorb the entropy.

14.3 Maxwell's Demon

We revisit Maxwell's demon, providing yet another interpretation. This is done by reworking Lemma 20 to the specific case of binary sequences. For the recording space, we use the set $\{0,1\}^\infty$ of infinite sequences with any computable probability measure λ over $\{0,1\}^\infty$. Thus by Proposition 18, $\mathbf{H}_\lambda(\alpha) =^+ -\mathbf{D}(\alpha|\lambda)$, where $\mathbf{D}$ is the deficiency of randomness. We couple the computable measure space $(\{0,1\}^\infty,\lambda)$ with a typical system $(\mathcal{X},\mu)$, such as where the phase space is the momentum and position of N particles, for large N. We couple a starting state $\alpha \in \mathcal{X}$, with recording state $\beta \in \{0,1\}^\infty$ that has room to record information, for example, where λ is the uniform measure and $\beta = 0^{1000}\kappa$, for some ML random string κ. The states are independent, with $\mathbf{I}(\alpha:\beta) \approx 0$. Joint dynamics are applied to get $(\alpha^t,\beta^t) = G(\alpha,\beta)$. By Lemma 20,

$$\mathbf{H}_\mu(\alpha^t) - \mathbf{H}_\mu(\alpha) >^+ \mathbf{D}(\beta|\lambda) - \mathbf{D}(\beta^t|\lambda) - \mathbf{I}((\alpha,\beta);t).$$

Again, for most times, $\mathbf{I}((\alpha,\beta);t)$ will be negligible. Thus after α decreases in algorithmic fine grain thermodynamic entropy, the contents of the register fills up, with a decrease is its deficiency of randomness $\mathbf{D}$. This shows that one benefit of an algorithmic formulation of thermodynamics is that pure algorithmic information and thermodynamic entropy can be exhanged in the course of joint dynamics. A graphical depiction of this phenomenon can be seen in Figure 14.2.

14.4 Distribution of Algorithmic Fine Grained Entropy

We say that measure ν is absolutely continuous with respect to μ, $\nu \ll \mu$ if $\mu(A) = 0$ implies $\nu(A) = 0$ for all $A \subseteq X$. The Radon–Nikodym theorem states that if $\nu \ll \mu$ there exists a measurable (over the Borel sets of $\mathcal{X}$) function f, uniquely defined up to a μ-nullset, such that for any measurable set $A \subseteq X$,

$$\nu(A) = \int_A f d\mu.$$

The function f can be written as $\frac{d\mu}{d\nu}$ or $\frac{\mu(dx)}{\nu(dx)}$. If $\mu \ll \nu$, then $\frac{\nu(dx)}{\mu(dx)} = \left(\frac{\mu(dx)}{\nu(dx)}\right)^{-1}$. If $\nu \ll \mu \ll$ then $\frac{d\nu}{\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}$. We use the short hand $\mu^x f(x) = \int f d\mu$. We define

$$\mathcal{H}_\nu(\mu) = -\int \log\left(\frac{d\mu}{d\nu}\right) d\mu = -\mu^x \log\frac{\mu(dx)}{\nu(dx)} = -\nu^x f(x)\log f(x).$$

If both ν and μ are probability measures, then $-\mathcal{H}_\nu(\mu) = \mathcal{D}(\mu\|\nu)$, where $\mathcal{D}$ is the Kullback–Leibler divergence. The following proposition shows that $\mathcal{H}_\nu(\mu)$ is non-positive when ν and μ are probability measures.

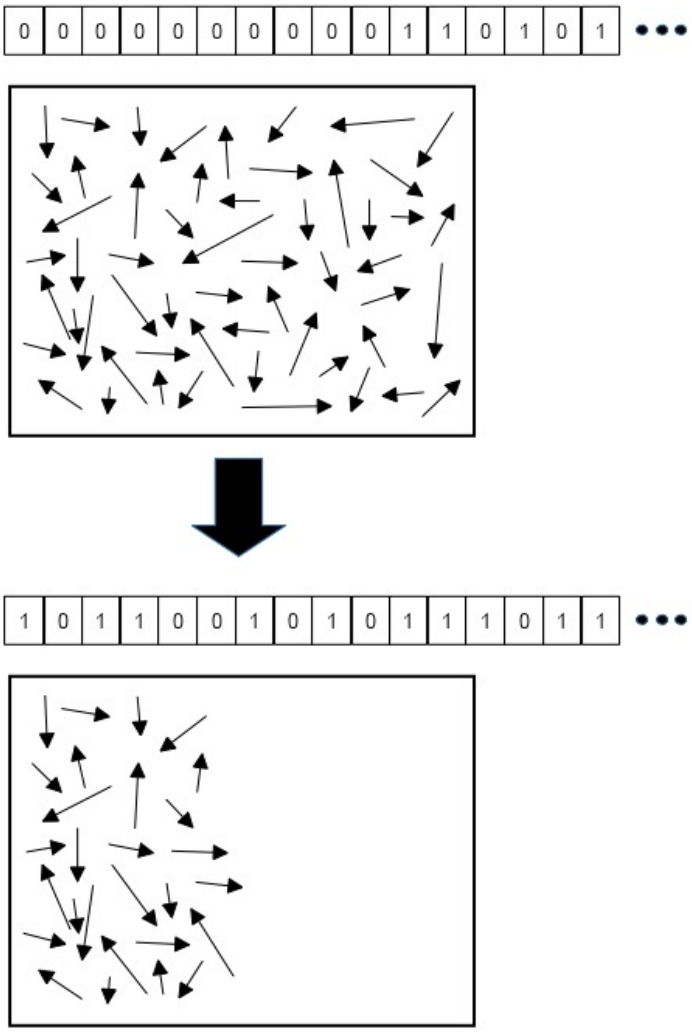


Figure 14.2: A graphical depciation of an algorithmic interpretation of Maxwell's demon. 2D particles in a box are paired with a register of information, represented as an infinite sequence with a high deficiency of randomness. After dynamics occur, the subsystem of the particle box moves to a state of low entropy, in this case with particles in a smaller region. Due to entropy balance Lemm 20, the register will fill up, moving to a sequence with low deficiency of randomness.

Proposition 24 *Over a space X,*

$$\mathcal{H}_\nu(\mu) \leq -\mu(X) \log \frac{\mu(X)}{\nu(X)}.$$

Proof. We use the inequality $-a \ln a \leq -a \ln b + b - a$. Letting $a = f(x)$ and $b = \mu(X)/\nu(X)$ and integrating by ν gives us:

$$\begin{aligned}(\ln 2)\mathcal{H}_\nu(\mu) = -\nu^x f(x) \ln f(x) &\leq -\mu(X) \ln \frac{\nu(X)}{\nu(X)} + \frac{\mu(X)}{\nu(X)}\nu(X) - \mu(X) \\ &= -\mu(X) \ln \frac{\mu(X)}{\nu(X)}.\end{aligned}$$

□

Theorem 72 *For computable metric space $\mathcal{X}$, let μ be measure that that $\mu(X) \geq 1$. Then*

$$\mathcal{H}_\nu(\mu) \leq \mu^x \mathbf{H}_\nu(x).$$

Proof. Let δ be the measure with density $\mathbf{t}_\nu(x)$ with respect ν, with $\mathbf{t}_\nu(x) = \frac{\delta(dx)}{\nu(dx)}$. Since $\nu^x \mathbf{t}_\nu(x) \leq 1$, $\delta(X) \leq 1$. Since $\mathbf{t}$ is a universal uniform test, $\mathbf{t}_n u(x) > 0$. Thus $\delta \ll \nu$, so by properties of the Radon-Nikodym derivative, $\frac{\nu(dx)}{\delta(dx)} = \left(\frac{\delta(dx)}{\nu(dx)}\right)^{-1}$. Using properties of the Radon-Nikodym derivative and Proposition 24,

$$\begin{aligned}\nu(\mu) &= -\mu^x \log \frac{\mu(dx)}{\nu(dx)} \\ -\mu^x \mathbf{H}_\nu(x) = \mu^x \log \frac{\delta(dx)}{\nu(dx)} &= -\mu^x \log \frac{\nu(dx)}{\delta(dx)} \\ \mathcal{H}_\nu(\mu) - \mu^x \mathbf{H}_\nu(x) = -\mu^x \log \frac{\mu(dx)}{\delta(dx)} &\leq -\mu(X) \log \frac{\mu(X)}{\delta(X)} \leq 0.\end{aligned}$$

□

14.5 Addition Equality

In this section, the universal Turing machine U is relativized to measures μ and ν. This means there is a fast Cauchy sequence to the measures in $\mathfrak{M}(\mathcal{X})$ space encoded in two auxilliary tapes of U. By Proposition 14, this means algorithms can lower compute the μ and ν measure of effectively open sets.

Proposition 25 $\mathbf{H}_\mu(x|\nu) <^+ -\log \nu^y 2^{-\mathbf{H}_{\mu,\nu}(x,y)}$.

Proof. Let $f(x) = -\log \nu^y 2^{-\mathbf{H}_{\mu,\nu}(x,y)}$. The function f is upper computable and has $\mu^x 2^{-f(x)} \leq 1$. Due to the universal properties of $\mathbf{t}_\mu$ and thus minimum property of $\mathbf{H}_\mu$, the inequality is proven. □

Proposition 26 *For a computable function $f : N^2 \to \mathbb{N}$,*

$$\mathbf{H}_\mu(x|y) <^+ \mathbf{K}(z) + \mathbf{H}_\mu(x|f(y,z)).$$

Proof. The function

$$g(x,y) = \sum_z 2^{-\mathbf{H}_\mu(x|f(y,z))-\mathbf{K}(z)},$$

is lower computable and $\mu^x g(x,y) \leq \sum_z 2^{-\mathbf{K}(z)} \leq 1$. So $g(x,y) \stackrel{*}{<} 2^{-\mathbf{H}_\mu(x|y)}$. The left hand side is a summation, so the inequality holds for each element of the sum, proving the proposition. □

Proposition 27 *If $i < j$, then*

$$i + \mathbf{H}_\mu(x|i) <^+ j + \mathbf{H}_\mu(x|j).$$

Proof. Using Proposition 26, with $f(i,n) = i+n$, we have

$$\mathbf{H}_\mu(x|i) - \mathbf{H}_\mu(x|j) <^+ \mathbf{K}(j-i) <^+ j - i.$$

The following proposition has a different proof to that of [Ǵ21], leveraging the results in Chapter 13. The Kolmogorov complexity of any upper computable function $f : X \to \mathbb{R} \cup \{-\infty\}$ is $\mathbf{K}(f)$, the length of the shortest program to upper compute f.

Proposition 28 *Let $F : \mathcal{Y} \to \mathbb{Z} \cup \{-\infty\}$ be an upper semicomputable function. By Corollary 21, among F uniform tests $g(x,y)$ with $\nu^y g(x,y) \leq 2^{-F(x)}$ there is a maximal F uniform test f within a multiplicative constant. For all y,*

$$f(x,y) \stackrel{*}{=} 2^{-F(x)} \mathbf{t}_\nu(y|x, F(x)).$$

Proof. To prove the inequality $\stackrel{*}{>}$, let $g(x,y,m) = \max_{i \geq m} 2^{-i} \mathbf{t}_\nu(y|x,i)$. This function is lower computable, and decreasing in m. Let $g(x,y) = g(x,y,F_\nu(x))$ is lower semicomputable since F is upper semi-computable. The multiplicative form of Proposition 27 implies

$$g(x,y,m) \stackrel{*}{=} 2^{-m} \mathbf{t}_\nu(y|x,m)$$
$$g(x,y) \stackrel{*}{=} 2^{-F(x)} \mathbf{t}_\nu(y|x,F(x)).$$

Since $\mathbf{t}_\nu$ is a test:

$$\nu^y 2^{-m} \mathbf{t}_\nu(y|x,m) \leq 2^{-m}$$
$$\nu^y g(x,y) \stackrel{*}{<} 2^{-F(x)},$$

which implies

$$g(x,y) \stackrel{*}{<} f(x,y)/\mathbf{m}(g|x) \stackrel{*}{<} f(x,y) 2^{\mathbf{K}(F)} \stackrel{*}{<} f(x,y)$$

by the optimality of $f(x,y)$. Note that if $F(x) = -\infty$, then $g(x,y) = \infty$ for all y and so $f(x,y) = 2^{-F(x)} \mathbf{t}_\nu(y|x,F(x)) = \infty$. We now consider the upper bound. Given fixed x, $2^{F(x)} f(x,y)$ is a ν-test conditional on x and $F(x)$. So

$$2^{F(x)} f(x,y) \stackrel{*}{<} \mathbf{t}_\nu(y|x,F(x))/\mathbf{m}(f(x,y)|x,F(x))) \stackrel{*}{<} \mathbf{t}_\nu(y|x,F(x)) 2^{\mathbf{K}(F)} \stackrel{*}{<} \mathbf{t}_\nu(y|x,F(x)).$$

□

Theorem 73 *Relativized to measures μ and ν,*

$$\mathbf{H}_{\mu\times\nu}(x,y) =^+ \mathbf{H}_\mu(x) + \mathbf{H}_\nu(y|x, \lceil \mathbf{H}_\mu(x) \rceil).$$

Proof. Let $f(x,y) = 2^{-\mathbf{H}_{\mu,\nu}(x,y)}$. Proposition 25 implies $\nu^y f(x,y) \stackrel{*}{<} 2^{-\mathbf{H}_\mu(x)}$. Let $F(x) = \lceil \mathbf{H}_\mu(x) \rceil + c$ for some proper constant $c \in \mathbb{N}$. Note that if h is a lower computable function such that $\nu^y h(x,y) \stackrel{*}{<} 2^{-\mathbf{H}_\mu(x)}$, then $\mu^x \nu^y h(x,y) \stackrel{*}{<} \mu^x \mathbf{t}_\mu(x) \stackrel{*}{<} 1$, so $h \stackrel{*}{<} f$, so f is a universal F-test. Proposition 28 (noting $\mathbf{K}(\mathbf{H}) = O(1)$) gives

$$\mathbf{H}_{\mu,\nu}(x,y) = -\log f(x,y,\mu) =^+ F(x) + \mathbf{H}_\nu(y|x, F(x)) =^+ \mathbf{H}_{\mu,\nu}(x,y) + \mathbf{H}_\nu(y|x, \lceil \mathbf{H}_\mu(x) \rceil).$$

□

Exercise 19 ([Vov]) *The randomness deficiency over the space $\{0,1\}^\infty \times \{0,1\}^\infty$, is $\mathbf{D}(\alpha,\beta)|\mu,\nu) = \sup_n -\log \mu(\alpha[0..n]) - \log \nu(\beta[0..n]) - \mathbf{K}(\alpha[0..n]\beta[0..n])$. Show that relativized to computable probabilities μ and ν over $\{0,1\}^\infty$,*

$$\mathbf{D}(\alpha,\beta|\mu,\nu) =^+ \mathbf{D}(\alpha|\mu) + \mathbf{D}(\beta|\nu, \lceil \mathbf{D}(\alpha|\mu) \rceil).$$

Exercise 20 *Let $\mathcal{X}$ and $\mathcal{Y}$ be two computable metric spaces and let μ be a computable measure over $\mathcal{X}$. Let κ be a computable probability kernel between $\mathcal{X}$ and $\mathcal{Y}$ (see Definition 50). For the joint distribution $\mu\kappa(\alpha,\beta) = \mu(\alpha)\kappa_\alpha(\beta)$, prove, relativized to μ and κ,*

$$\mathbf{H}_{\mu\kappa}(\alpha,\beta) =^+ \mathbf{H}_\mu(\alpha) + \mathbf{H}_{\kappa_\alpha}(\beta|\alpha, \mathbf{H}_\mu(\alpha)).$$

Chapter 15

Oscillation of Algorithmic Fine Grained Entropy

In this chapter, it is proven that the algorithmic fine grained entropy of states will oscillate in the presence of continuous dynamics. This is over computable dynamics starting at a point that has finite mutual information with the halting sequnce, $\mathcal{H}$. To prove this result, properties abouts sets of finite and infinite sequences and their relationship to $\mathcal{H}$ need to be proven, which is done in Sections 15.1 and 15.2. The main theorem of this chapter matches entropy oscillations proven to occur in the case of discrete ergodic dynamics, seen in Chapter 16.

15.1 On Exotic Sets of Natural Numbers

The following result also implies the existence of outliers in the physical world, which we detail now. An outlier is an observation that is set apart from a population. There are many reasons that such anomalies occur, including measurement error and human error. However recent results have shown that outliers are ingrained into the nature of algorithms and dynamics. In [Eps21b], anomalies were proven to occur in sampling algorithms. In [Eps22a], anomalies were proven to exist in the outputs of probabilistic algorithms. They were also proven to be emergent computable ergodic dynamics on the Cantor space. In [Eps22b] anomalies were shown to emergent in a more general (but still computable) class of dynamics.

But what about measurements of systems that are too complex to be considered algorithmic? One example is the global weather system. One can attest to the fact that there are many strange formations that occur! To show that anomalies occur, one can use the Independence Postulate, [Lev84, Lev13], which is introduced in Chapter **??**. The Independence Postulate is a finitary Church-Turing thesis, postulating that certain finite and infinite sequences cannot be easily be found with a short "physical address".

In this section, its proved that large sets of numbers whose members are typical (i.e. low $\mathbf{d}$ score) with respect to two probabilities will have large information with the halting sequence. Therefore, by the Independence Postulate, such set (i.e. meassurements) cannot be found in the physical world. Thus all large sets of measurements that have small addresses must have large outlier (i.e. $\mathbf{d}$) scores.

Lemma 21 *For computable probabilities p, q over $\mathbb{N}$, $D \subset \mathbb{N}$, $|D| = 2^s$,*
$s < \max_{a \in D} \min\{\mathbf{d}(a|p), \mathbf{d}(a|q)\} + \mathbf{I}(D; \mathcal{H}) + O(\mathbf{K}(\mathbf{I}(D; \mathcal{H}), p, q, s))$.

Proof. We relativize the universal Turing machine to $\langle s,p,q\rangle$. Let Q be a probability measure that realizes $\mathbf{Ks}(D)$, with $d=\max\{\mathbf{d}(D|Q),1\}$. Let $F\subseteq\mathbb{N}$ be a random set where each element $a\in\mathbb{N}$ is selected independently with probability $d2^{-s}$. $\mathbf{E}[p(F)]=\mathbf{E}[q(F)]\leq d2^{-s}$. Furthermore

$$\mathbf{E}[Q(\{G:|G|=2^s,G\cap F=\emptyset\})]\leq\sum_G Q(G)(1-d2^{-s})^{2^s}<e^{-d}.$$

Thus finite $W\subset\mathbb{N}$ can be chosen such that $p(W)\leq 4d2^{-s}$, $q(W)\leq 4d2^{-s}$, and $Q(\{G:|G|=2^s,G\cap W=\emptyset\})\leq e^{2-d}$. $D\cap W\neq\emptyset$, otherwise, using the Q-test, $t(G)=[|G|=2^s,G\cap W=\emptyset)]e^{d-2}$, we have $t(D=e^{d-2}$. So

$$1.44d<^+\log t(x)<^+\mathbf{d}(D|Q,d)<^+d+\mathbf{K}(d),$$

which is a contradiction for large enough d which one can assume without loss of generality. Thus there is an $a\in D\cap W$, where

$$\begin{aligned}\mathbf{K}(a)&<^+\min\{-\log q(a),-\log p(a)\}+\log d-s+\mathbf{K}(d)+\mathbf{K}(Q)\\ s&<^+\min\{\mathbf{d}(a|p),\mathbf{d}(a|q)\}+\mathbf{Ks}(D).\end{aligned}$$

Making the relativization of $\langle s,p,q\rangle$ explicit, and using Lemma 3 results in

$$\begin{aligned}s&<^+\min\{\mathbf{d}(a|p),\mathbf{d}(a|q)\}+\mathbf{Ks}(D)+O(\mathbf{K}(s,p,q))\\ s&<\max_{a\in D}\min\{\mathbf{d}(a|p),\mathbf{d}(a|q)\}+\mathbf{Ks}(D)+O(\mathbf{K}(s,p,q))\\ s&<\max_{a\in D}\min\{\mathbf{d}(a|p),\mathbf{d}(a|q)\}+\mathbf{I}(D;\mathcal{H})+O(\mathbf{K}(\mathbf{I}(D;\mathcal{H}),s,p,q))).\square\end{aligned}$$

□

Exercise 21 *For computable probabilities $\{p_i\}_{i=1}^n$ over $\mathbb{N}$, $D\subset\mathbb{N}$, $|D|=2^s$, show that* $s<\max_{a\in D}\min_i\{\mathbf{d}(a|p_i)\}+\mathbf{I}(D;\mathcal{H})+\log n+O(\mathbf{K}(\mathbf{I}(D;\mathcal{H}),\{p_i\},s))$.

15.2 On Exotic Sets of Reals

The results in Section 15.1 are limited in that each observation is a natural number. However observations in the physical world oftentimes have an infinite degree of precision. To that end, Lemma 21 can be extended to infinite sequences. That is, large sets $Z\subset\{0,1\}^\infty$ with no high randomness deficiency scores will have high mutual information with the halting sequence. To prove this, one needs to manipulate prefixes $Z_n=\{\alpha[0..n]:\alpha\in Z\}$ of the original sets Z and then apply Lemma 21. However the computation of the exact cutoff point n is tricky, because as we will show, n must be simple relative to the halting sequence $\mathcal{H}$, that is n must be in the range of a so-called busy-beaver function, $\mathbf{bb}$.

To this end, let $\Omega=\sum\{2^{-\|p\|}:U(p)\text{ halts}\}$ be Chaitin's Omega, $\Omega_n\in\mathbb{Q}_{\geq 0}$ be be the rational formed from the first n bits of Ω, and $\Omega^t=\sum\{2^{-\|p\|}:U(p)\text{ halts in time }t\}$. For $n\in\mathbb{N}$, let $\mathbf{bb}(n)=\min\{t:\Omega_n<\Omega^t\}$. Note that the busy-beaver function $\mathbf{bb}$ is uncomputable, but computable relative to $\mathcal{H}$. $\mathbf{bb}^{-1}(m)=\arg\min_n\{\mathbf{bb}(n-1)<m\leq\mathbf{bb}(n)\}$. Note that $\mathbf{bb}(\mathbf{bb}^{-1}n)\geq n$. Let $\Omega[n]\in\{0,1\}^*$ be the first n bits of Ω. The following lemma shows that very large numbers m can be used to compute $\mathbf{bb}^{-1}(m)$ bits of Chaitin's omega Ω.

Lemma 22 *For $n=\mathbf{bb}^{-1}(m)$, $\mathbf{K}(\Omega[n]|m,n)=O(1)$.*

Proof. For a string x, let $BB(x) = \inf\{t : \Omega^t > 0.x\}$. Enumerate strings of length n, starting with 0^n, and return the first string x such that $BB(x) \geq m$. This string x is equal to $\Omega[n]$, otherwise let y be the largest common prefix of x and $\Omega[n]$. Thus $BB(y) = \mathbf{bb}(\|y\|) \geq BB(x) \geq m$, which means $\mathbf{bb}^{-1}(m) \leq \|y\| < n$, causing a contradiction. □

The following lemma, while lengthy, is a series of straightforward application of inequalities.

Lemma 23 *For computable probabilities P, Q, over $\{0,1\}^\infty$, $Z \subset \{0,1\}^\infty$, $|Z| = 2^s$, $s < \max_{\alpha \in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{I}(\langle Z\rangle : \mathcal{H}) + O(\mathbf{K}(s, P, Q) + \log \mathbf{I}(\langle Z\rangle; \mathcal{H}))$.*

Proof. We relativize the universal Turing machine to $\langle s, P, Q\rangle$, which can be done due to the precision of the theorem. Let $Z_n = \{\alpha[0..n] : \alpha \in Z\}$ and $m = \arg\min_m |Z_m| = |Z|$. Let $n = \mathbf{bb}^{-1}(m)$ and $k = \mathbf{bb}(n)$, where $k \geq m$. Let p and q be probabilities over $\{0,1\}^*$, where $p(x) = [\|x\| = k]P(x)$ and $\langle p\rangle = \langle k\rangle$ and let $q(x) = [\|x\| = k]Q(x)$ and $\langle q\rangle = \langle k\rangle$. Using $D = Z_k$, Lemma 21, relativized to k, produces $x \in Z_k$, where

$$\begin{aligned}
s &< \min\{\mathbf{d}(x|p), \mathbf{d}(x|q)\} + \mathbf{I}(Z_k; \mathcal{H}|k) + O(\mathbf{K}(\mathbf{I}(Z_k; \mathcal{H}|k), q, p|k)) \\
&< \max_{\alpha \in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{K}(Z_k|k) + \mathbf{K}(k) - \mathbf{K}(Z_k|k, \mathcal{H}) + O(\mathbf{K}(\mathbf{I}(Z_k; \mathcal{H}|k), q, p|k)). \\
&< \max_{\alpha \in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{K}(Z_k|k) + \mathbf{K}(k) - \mathbf{K}(Z_k|k, \mathcal{H}) + O(\mathbf{K}(P, Q) + \log \mathbf{I}(Z_k; \mathcal{H}|k)).
\end{aligned}$$

Since $\mathbf{K}(k) <^+ n + \mathbf{K}(n)$, by the chain rule,

$$\begin{aligned}
&\mathbf{K}(Z_k|k) + \mathbf{K}(k) \\
<^+ &\mathbf{K}(Z_k|k, \mathbf{K}(k)) + \mathbf{K}(\mathbf{K}(k)|k) + \mathbf{K}(k) \\
< &\mathbf{K}(Z_k, k) + O(\log n) \\
< &\mathbf{K}(Z_k) + O(\log n).
\end{aligned}$$

So

$$s < \max_{\alpha \in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{K}(Z_k) - \mathbf{K}(Z_k|k, \mathcal{H}) + O(\log n + \mathbf{K}(P, Q) + \log \mathbf{I}(Z_k; \mathcal{H}|k))).$$

Since $\mathbf{K}(k|n, \mathcal{H}) = O(1)$, $\mathbf{K}(Z_k|\mathcal{H}) <^+ \mathbf{K}(Z_k|k, \mathcal{H}) + \mathbf{K}(n)$,

$$s < \max_{\alpha \in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{I}(Z_k; \mathcal{H}) + O(\log n + \mathbf{K}(P, Q) + \log \mathbf{I}(Z_k; \mathcal{H}|k)).$$

Furthermore since $\mathbf{I}(Z_k; \mathcal{H}|k) + \mathbf{K}(k) < \mathbf{I}(Z_k; \mathcal{H}) + O(\log n)$,

$$s < \max_{\alpha \in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{I}(Z_k; \mathcal{H}) + O(\log n + \mathbf{K}(P, Q)) + O(\log \mathbf{I}(Z_k; \mathcal{H})). \tag{15.1}$$

By Lemma 22, $\mathbf{K}(\Omega[n]|Z_k) <^+ \mathbf{K}(n)$ so by Lemma 1, and by using the common fact that the first n bits of Ω has $n - O(\log n)$ bits of mutual information with $\mathcal{H}$, one gets

$$n <^{\log} \mathbf{I}(\Omega[n]; \mathcal{H}) <^{\log} \mathbf{I}(Z_k; \mathcal{H}) + \mathbf{K}(n) <^{\log} \mathbf{I}(Z_k; \mathcal{H}). \tag{15.2}$$

Combining Equations 15.1 and 15.2, results in

$$s < \max_{\alpha \in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{I}(Z_k; \mathcal{H}) + O(\mathbf{K}(P, Q) + \log \mathbf{I}(Z_k; \mathcal{H})). \tag{15.3}$$

By the definition of mutual information $\mathbf{I}$ between infinite sequences

$$\mathbf{I}(Z_k;\mathcal{H}) <^{+} \mathbf{I}(Z:\mathcal{H}) + \mathbf{K}(Z_k|Z) <^{\log} \mathbf{I}(Z:\mathcal{H}) + \mathbf{K}(k|Z). \tag{15.4}$$

Combining Equations 15.3 and 15.5 results in

$$s < \max_{\alpha\in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{I}(Z:\mathcal{H}) + \mathbf{K}(x|Z) + O(\mathbf{K}(P,Q) + \log \mathbf{I}(Z_k;\mathcal{H})). \tag{15.5}$$

Now m is simple relative to Z and by Lemma 22, $\Omega[n]$ is simple relative to m and n. Furthermore k is simple relative to $\Omega[n]$. Therefore using Equations 15.2 and 15.5

$$\mathbf{K}(k|Z) <^{+} \mathbf{K}(n) < O(\log \mathbf{I}(Z_k;\mathcal{H})) < O(\log(\mathbf{I}(Z:\mathcal{H}) + \mathbf{K}(k|Z))) < O(\log(\mathbf{I}(Z:\mathcal{H}))). \tag{15.6}$$

So combining Equations 15.5 and 15.6, one gets,

$$s < \max_{\alpha\in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{I}(Z:\mathcal{H}) + O(\log n) + O(\mathbf{K}(P,Q) + \log \mathbf{I}(Z;\mathcal{H}))$$
$$s < \max_{\alpha\in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{I}(Z:\mathcal{H}) + O(\mathbf{K}(s,P,Q) + \log \mathbf{I}(Z;\mathcal{H}))).$$

□

Exercise 22 *For computable probabilities $\{P_i\}_{i=1}^n$ over $\{0,1\}^\infty$, $Z \subset \{0,1\}^\infty$, $|Z| = 2^s$, show that $s < \max_{\alpha\in Z} \min_i\{\mathbf{D}(\alpha|P_i)\} + \mathbf{I}(\langle Z\rangle : \mathcal{H}) + \log n + O(\mathbf{K}(s, \{P_i\}_) + \log \mathbf{I}(\langle Z\rangle;\mathcal{H}))$.*

15.3 Oscillations Occur

This section contains one the main results of the manuscript, that algorithmic fine grained entropy will fluctuate during the course of dynamics. A fluction of entropy size n will occur at least approximately 2^{-n} amount of the time. In Section 15.4, it will be shown that such oscillations rarely occur, making for tight lower and upper bounds on their frequency. These bounds only occur if the starting point has finite mutual information with the halting sequence.

By the Independence Postulate (see Chapter **??**), starting points with infinite mutual information with the halting sequence can be seen as exotic on unphysical. Thus the results in this section establish a deep connection between complexity theoretic notions of the halting sequence with thermodynamic theoretic notions of entropy fluctuations. In Chapter 16, analagous results are proven for discrete ergodic processes.

We first introduce a new definition, scoring the level of mutul information a point α in a computable metric space has with the halting sequence $\mathcal{H}$. The idea is to define a set A of all encoded fast Cauchy sequences to α and then define information to be the infimum of mutual information of elements of A with $\mathcal{H}$. Thus this definition is independent of the way a point is coded.

Definition 38 (Mutual Information with the Halting Sequence) *An encoding of a fast Cauchy sequence $\overrightarrow{x}$ is $\langle\overrightarrow{x}\rangle \in \{0,1\}^\infty$, with $\langle\overrightarrow{x}\rangle = \langle x_1\rangle\langle x_2\rangle\ldots$. Each $x_i \in \overrightarrow{x}$ is an ideal point, and i is its order in the enumeration of $\overrightarrow{x}$. Each point $x \in \mathcal{X}$ has a certain mutual information with the halting sequence $\mathbf{I}(x:\mathcal{H}) = \inf\{\mathbf{I}(\langle\overrightarrow{x}\rangle : \mathcal{H}) : \overrightarrow{x}$ is a fast Cauchy sequence for $x\}$.*

The following theorem, stated in [Lev74] and reproved in [Ver21], says that randomized methods cannot increase the average mutual information with the halting sequence.

Theorem 74 ([Ver21, Lev74]) *Let P_ρ, be a family of probability distributions over $\{0,1\}^\infty$, indexed by $\rho \in \{0,1\}^\infty$. Assume that there is a Turing machine T such that for all $\rho \in \{0,1\}^\infty$ computes P_ρ having oracle access to ρ. By "compute" we mean all the measures of the cylinder sets $P_\rho(x\{0,1\}^\infty)$, can be computed, uniformly in $x \in \{0,1\}^*$. Then there is a constant $c_T > 0$ solely dependent on T such that*

$$P_\rho\{\gamma : \mathbf{I}(\langle \gamma, \rho\rangle : \mathcal{H}) > m\} < 2^{\mathbf{I}(\rho:\mathcal{H})-m+c_T}.$$

The following theorem shows oscilliations of entropy occur during dynamics. Note that for any computable transform group G^t, and $a, b \in \mathbb{Q}_{>0}$, $a < b$, there exists another transformation group F^t such that $F^t = G^{a-(a-b)t}$. This means that the entropy is constantly fluctuating at every time interval, no matter how small. Theorem 92 in Chapter 19 is more general and the proof is more straightforward, but the bounds are looser. Note that the measure preserving requirement for the transformation group is not needed for Theorem 75.

Theorem 75 (Oscillation of Thermodynamic Entropy) *Let L be the Lebesgue measure over $\mathbb{R}$, $(\mathcal{X}, \mu)$ be a computable measure space, $\alpha \in \mathcal{X}$, with finite $\mathbf{I}(\alpha : \mathcal{H})$. For transformation group G^t acting on $\mathcal{X}$, there is a constant c with $L\{t \in [0,1] : \mathbf{H}_\mu(G^t\alpha) < \log \mu(X) - n\} > 2^{-n-\mathbf{K}(n)-c}$.*

Proof. We first assume not. There exists $(G^t, \mathcal{X})$ and computable measure space $(\mathcal{X}, \mu)$ and there exists $\alpha \in X$ such that for all $c \in \mathbb{N}$, there exists n, where

$$L(\{t \in [0,1] : \mathbf{H}_\mu(G^t\alpha) < \log \mu(\mathcal{X}) - n\}) < 2^{-n-\mathbf{K}(n)-c}$$
$$L(\{t \in [0,1] : n - \log \mu(\mathcal{X}) < \log \mathbf{t}_\mu(G^t\alpha)\}) < 2^{-n-\mathbf{K}(n)-c}.$$

We sample $2^{n+\mathbf{K}(n)+c-1}$ elements F by choosing a time t uniformly between $[0,1]$. The probability that all samples $\beta \in F$ have $\mathbf{t}_\mu(G^\beta\alpha) \leq n - \log \mu(\mathcal{X})$ is

$$\begin{aligned}
&\prod_{i=1}^{|F|} L\{t \in [0,1] : \log \mathbf{t}_\mu(G^t\alpha) \leq n - \log \mu(\mathcal{X})\} \\
\geq &(1 - |F|2^{-n-\mathbf{K}(n)-c}) \\
\geq &(1 - 2^{n+\mathbf{K}(n)+c-1}2^{-n-\mathbf{K}(n)-c}) \\
\geq &1/2.
\end{aligned}$$

Let $(\{0,1\}^\infty, \Gamma)$ be the Cantor space with the uniform measure. The binary representation (see Theorem 70) creates an isomorphism (ϕ, ϕ^{-1}) of computable probability spaces between the spaces $(\{0,1\}^\infty, \Gamma)$ and $([0,1], L)$. It is the canonical function $\phi(\gamma) = 0.\gamma$. Thus for all Borel sets $A \subseteq [0,1]$, $\Gamma(\phi^{-1}(A)) = L(A)$. Since $\{t \in [0,1] : \log \mathbf{t}_\mu(G^t\alpha) \leq n - \log \mu(\mathcal{X})\}$ is closed,

$$L\{t \in [0,1] : \log \mathbf{t}_\mu(G^t\alpha) \leq n - \log \mu(\mathcal{X})\} = \Gamma\{\gamma \in \{0,1\}^\infty : \log \mathbf{t}_\mu(G^{\phi(\gamma)}\alpha) \leq n - \log \mu(\mathcal{X})\}.$$

So

$$1/2 \leq \prod_{i=1}^{|F|} \Gamma\{\gamma \in \{0,1\}^\infty : \log \mathbf{t}_\mu(G^{\phi(\gamma)}\alpha) \leq n - \log \mu(\mathcal{X})\}.$$

Let (δ, μ_δ) be a binary representation (see Definition 30), for the computable measure space $(\mathcal{X}, \mu)$. Thus μ_δ is a computable (not necessarily probability) measure over $\{0,1\}^\infty$. By Lemma 19, there is a $c' > 0$, where

$$\prod_{i=1}^{|F|} \Gamma\{\gamma \in \{0,1\}^\infty : \log \mathbf{t}_{\mu_\delta}(\delta^{-1}(G^{\phi(\gamma)}\alpha)) \leq n - \log \mu(\mathcal{X}) + c'\} \geq 1/2.$$

Let $f : \{0,1\}^\infty \times \{0,1\}^\infty \to \{0,1\}^\infty$, where $f(\gamma, \langle \overrightarrow{\zeta} \rangle) = \delta^{-1}(G^{\phi(\gamma)}\zeta)$. Note, $f(\gamma, \langle \overrightarrow{\zeta} \rangle)$ can be undefined when $\mathbf{t}_\mu(G^{\phi(\gamma)}\zeta) = \infty$, because the morphism δ^{-1} is only proven to be defined on a constructive G_δ set of full measure which includes random points. Let $\xi = \langle \overrightarrow{\alpha} \rangle$ be an encoding of a fast Cauchy sequence $\overrightarrow{\alpha}$ such that $\mathbf{I}(\xi : \mathcal{H}) < \infty$. The sequence ξ is guaranteed to exist because the assumption of the theorem statement. So

$$\prod_{i=1}^{|F|} \Gamma\{\gamma \in \{0,1\}^\infty : \log \mathbf{t}_{\mu_\delta}(f(\gamma, \xi)) \leq n - \log \mu(\mathcal{X}) + c'\} \geq 1/2.$$

By Proposition 18, (and also updating c')

$$\prod_{i=1}^{|F|} \Gamma\{\gamma \in \{0,1\}^\infty : \mathbf{D}(f(\gamma, \xi)|\mu_\delta) \leq n - \log \mu(\mathcal{X}) + c' + \mathbf{K}(\mu_\delta)\} \geq 1/2.$$

Let $\overline{\mu}_\delta(\alpha) = \mu_\delta(\alpha)/\mu_\delta(\{0,1\}^\infty)$, which is a computable probability measure over $\{0,1\}^\infty$.

$$\prod_{i=1}^{|F|} \Gamma\{\gamma \in \{0,1\}^\infty : \mathbf{D}(f(\gamma, \xi)|\overline{\mu}_\delta) \leq n + c' + \mathbf{K}(\mu_\delta)\} \geq 1/2.$$

Let Γ^{n+c} be a computable distribution over the product of $1 + 2^{n+\mathbf{K}(n)+c-1}$ independent probability measures over $\{0,1\}^\infty$, encoding into a $\{0,1\}^\infty$ in the standard way. The first probability distribution gives measure 1 to ξ and the last $2^{n+\mathbf{K}(n)+c}$ probability measures are the uniform distribution Γ over $\{0,1\}^\infty$. So

$$\Gamma^{n+c}(\text{Encoding of } 1 + 2^{n+\mathbf{K}(n)+c-1} \text{ elements with the first encoded sequence being } \xi$$
$$\text{and the rest of encoded sequences } \beta \text{ has } \mathbf{D}(f(\beta, \xi)|\overline{\mu}_\delta) \leq n + c' + \mathbf{K}(\mu_\delta)) \geq 1/2.$$

Let $n^* = \langle n, \mathbf{K}(n) \rangle$. There is an infinite sequence $\eta = \langle n, \mathbf{K}(n), c \rangle \xi$ and a Turing machine T, such that T computes Γ^{n+c} when given oracle access to η. By Theorem 74, with the universal Turing machine relativized to n^*, and folding the constants together,

$$\begin{aligned}
&\Gamma^{n+c}(\{\gamma : \mathbf{I}(\gamma : \mathcal{H}|n^*) > m\}) \\
<&\Gamma^{n+c}(\{\gamma : \mathbf{I}(\langle \gamma, \eta \rangle : \mathcal{H}|n^*) >^+ m\}) \\
\overset{*}{<}&2^{-m+\mathbf{I}(\eta:\mathcal{H}|n^*)+c_T} \\
\overset{*}{<}&2^{-m+\mathbf{K}(n,\mathbf{K}(n),c|n^*)+\mathbf{I}(\xi:\mathcal{H}|n^*)+c_T} \\
\overset{*}{<}&2^{-m+\mathbf{K}(c)}.
\end{aligned}$$

Therefore,

$$\Gamma^{n,c}(\{\gamma : \mathbf{I}(\gamma : \mathcal{H}|n^*) >^+ \mathbf{K}(c)\}) \leq 1/4.$$

Thus, by probabilistic arguments, there exists $\kappa \in \{0,1\}^\infty$, such that $\kappa = \langle D, \xi \rangle$, where $D \subset \{0,1\}^\infty$ and $|D| = 2^{n+\mathbf{K}(n)+c-1}$ and each $\beta \in D$ has $\mathbf{D}(f(\beta,\xi)|\overline{\mu}_\delta) \leq n+c'+\mathbf{K}(\mu_\delta)$ and $\mathbf{I}(\kappa : \mathcal{H}|n^*) <^+ \mathbf{K}(c)$. Thus since $\mathbf{K}(f(D,\xi)|\kappa, n^*) = O(1)$ we have $\mathbf{I}(f(D,\xi) : \mathcal{H}|n^*) <^+ \mathbf{I}(\kappa : \mathcal{H}|n^*) <^+ \mathbf{K}(c)$. By Lemma 23, relativized to n^*, on the set $D' = f(D,\xi)$ and probability $\overline{\mu}_\delta$, there exists constants $d, f \in \mathbb{N}$ where

$$\begin{aligned} m = \log|D| &< \max_{\beta \in D'} \mathbf{D}(\beta|\overline{\mu}_\delta, n^*) + 2\mathbf{I}(D' : \mathcal{H}|n^*) + d\mathbf{K}(m|v) + f\mathbf{K}(\overline{\mu}_\delta|n^*) \\ m &< \max_{\beta \in D'} \mathbf{D}(\beta|\overline{\mu}_\delta) + \mathbf{K}(n) + 2\mathbf{I}(D' : \mathcal{H}|n^*) + d\mathbf{K}(m|n^*) + f\mathbf{K}(\mu_\delta|n^*) \\ &<^+ \max_{\beta \in D'} \mathbf{D}(\beta|\overline{\mu}_\delta) + \mathbf{K}(n) + 2\mathbf{K}(c) + d\mathbf{K}(m|v) + f\mathbf{K}(\mu_\delta|n^*) \\ &<^+ n + \mathbf{K}(n) + d\mathbf{K}(m|v) + 2\mathbf{K}(c) + (f+1)\mathbf{K}(\mu_\delta). \end{aligned} \tag{15.7}$$

Therefore:

$$\begin{aligned} m &= n + \mathbf{K}(n) + c - 1 \\ \mathbf{K}(m|n^*) &<^+ \mathbf{K}(c). \end{aligned} \tag{15.8}$$

Plugging Equation 15.8 back into Equation 18.1 results in

$$\begin{aligned} n + \mathbf{K}(n) + c &<^+ n + \mathbf{K}(n) + 2\mathbf{K}(c) + d(\mathbf{K}(c) + O(1)) + (f+1)\mathbf{K}(\mu_\delta) \\ c &<^+ (2+d)\mathbf{K}(c) + dO(1) + (f+1)\mathbf{K}(\mu_\delta). \end{aligned}$$

This result is a contradiction for sufficiently large c solely dependent $\mathcal{X}$, G, μ, and the universal Turing machine. □

Corollary 23 (Oscillation of Marginal Thermodynamic Entropy) *Let L be the Lebesgue measure over $\mathbb{R}$, $(\mathcal{X} \times \mathcal{Y}, \mu \times \nu)$ be a computable product measure space, $(\alpha, \beta) \in \mathcal{X} \times \mathcal{Y}$, with finite $\mathbf{I}((\alpha,\beta) : \mathcal{H})$. For transformation group G^t acting on $\mathcal{X} \times \mathcal{Y}$, there is a constant c where if $(\alpha^t, \beta^t) = G^t(\alpha,\beta)$, then $L\{t \in [0,1] : \mathbf{H}_\mu(\alpha^t) < \log \mu(X) - n\} > 2^{-n-\mathbf{K}(n)-c}$.*

15.4 Oscillations are Rare

The following lemma shows that Theorem 75 is tight. It does require the measure preserving condition of G^t. It is the same statement as Theorem 6 in [Gac94], with an updated proof.

Lemma 24 *Let L be the Lebesque measure over $\mathbb{R}$, $(\mathcal{X}, \mu)$ be a computable measure space, and $\alpha \in \mathcal{X}$. For transformation group G^t acting on $\mathcal{X}$, there is a constant c where $L\{t \in [0,1] : \mathbf{H}_\mu(G^t\alpha) < \mathbf{H}_\mu(\alpha) - m\} < 2^{-m+c}$.*

Proof. Since

$$\int_{\mathcal{X}} \int_{[0,1]} 2^{-\mathbf{H}_{\mu\times L}(\alpha,t)} dL(t) d\mu(\alpha) = \int_{\mathcal{X}} \int_{[0,1]} \mathbf{t}_{\mu\times L}(\alpha,t) dL(t) d\mu(\alpha) \leq 1,$$

the function $f(\alpha) = \int_{[0,1]} 2^{-\mathbf{H}_{\mu\times L}(\alpha,t)} dL(t)$ is a μ-test. So

$$\int_{[0,1]} 2^{-\mathbf{H}_{\mu\times L}(\alpha,t)} dt = f(\alpha) \stackrel{*}{<} \mathbf{t}_\mu(\alpha) \stackrel{*}{=} 2^{-\mathbf{H}_\mu(\alpha)}.$$

So

$$\{t \in [0,1] : 2^{-\mathbf{H}_{\mu\times L}(\alpha,t)} > 2^{m-\mathbf{H}_\mu(\alpha)}\} \stackrel{*}{<} 2^{-m}$$
$$\{t \in [0,1] : \mathbf{H}_{\mu\times L}(\alpha,t) < \mathbf{H}_\mu(\alpha) - m\} \stackrel{*}{<} 2^{-m}$$

$\mathbf{H}_{\mu\times L}(\alpha,t) <^+ \mathbf{H}_\mu(G^t\alpha)$ because

$$\begin{aligned}
&\int_{[0,1]}\int_{\mathcal{X}} \mathbf{t}_\mu(G^t\alpha)d\mu(\alpha)dL(t)\\
=&\int_{[0,1]}\int_{\mathcal{X}} \mathbf{t}_\mu(\alpha)d\mu(G^{-t}\alpha)dL(t)\\
=&\int_{[0,1]}\int_{\mathcal{X}} \mathbf{t}_\mu(\alpha)d\mu(\alpha)dL(t),\\
=&\int_{[0,1]} 1dL(t)\\
\leq& 1,
\end{aligned}$$

which means $\mathbf{t}_\mu(G^t\alpha) \stackrel{*}{<} \mathbf{t}_{\mu\times L}(\alpha,t)$ and thus $2^{-\mathbf{H}_\mu(G^t\alpha)} \stackrel{*}{<} 2^{-\mathbf{H}_{\mu\times L}(\alpha,t)}$. Thus

$$\{t \in [0,1] : \mathbf{H}_\mu(G^t\alpha) < \mathbf{H}_\mu(\alpha) - m\} \stackrel{*}{<} 2^{-m}.$$

□

Exercise 23 *Let L be the Lebesque measure over $\mathbb{R}$, $(\mathcal{X},\mu)$ be a computable measure space, and $\alpha \in \mathcal{X}$. For transformation group G^t acting on $\mathcal{X}$, prove that there is a constant c where $L\{t \in [0,T] : \mathbf{H}_\mu(G^t\alpha) < \mathbf{H}_\mu(\alpha) - \mathbf{K}(T) - m\} < 2^{-m+c}T$.*

Combining Theorem 75 and Lemma 24 together, one gets a full characterization of the dynamics of states in the phase space. This corollary was shown in the introduction, and is a central result of the manuscript.

Corollary 24 *Let L be the Lebesgue measure over $\mathbb{R}$, and $(\mathcal{X},\mu)$ be a computable measure space, and $\alpha \in \mathcal{X}$ with finite $\mathbf{I}(\alpha : \mathcal{H})$. For transformation group G^t acting on $\mathcal{X}$, there is a constant c with*

$$2^{-n-\mathbf{K}(n)-c} < L\{t \in [0,1] : \mathbf{H}_\mu(G^t\alpha) < \log\mu(\mathcal{X}) - n\} < 2^{-n+c}.$$

Chapter 16

Discrete Dynamics

In the previous chapter, results were proven about continuous dynamics were proven. In this chapter, results about discrete dynamics are detailed, with an emphasis on ergodic dynamics. Theorem 27 shows that oscillations occur in ergodic dynamics. Furthermore, the longer the dynamics occur, the greater the fluctuation is guaranteed to occur. This closely mirrors Corollary 24 (down to even the error terms), which characterizes continuous dynamics. Results are also proven about ergodic dynamics, showing that lower computable open sets will have a hitting frequency equal to their measure.

Definition 39 (Discrete Dynamics) *Discrete dynamics is modeled by a transform group G^t from Definition 36, but with $t \in \mathbb{Z}$, being an integer. This means G^t is measure preserving. We assume there no $\alpha \in \mathcal{X}$ with a finite orbit.*

Proposition 29 *The sets $\{\alpha : \mathbf{H}_\mu(\alpha) \neq -\infty\}$ and $\{\alpha : \mathbf{H}_\mu(\alpha) = -\infty\}$ are conserved under a discrete transformation group G.*

Proof. If $\mathbf{H}_\mu(\alpha) = -\infty$, then α is in an effective null set $\bigcap_n U_n$. Thus $G^t\alpha$ is in the null set $\bigcap_n G^t U_n$ and thus $\mathbf{H}_\mu(G^t\alpha) = -\infty$. The same reasoning for $-t$ proves that $\{\alpha : \mathbf{H}_\mu(\alpha) = -\infty\}$ is closed under G^t, proving the case when $\mathbf{H}_\mu(\alpha) \neq -\infty$.

16.1 Synchronized Oscillations

This section continues the discussion of Section 15.1 that outliers occur in algorithms, dynamic, and the physical world. The result in this section generalize the results from infinite sequences of numbers or reals to computable metric spaces, where the outlier score is the uniform universal test $\mathbf{t}_\mu$. Results in this section prove discrete dynamics will visit states with ever increasing $\mathbf{t}_\mu$ and $\mathbf{t}_\nu$ score. Otherwise the setup is exotic, with high mutual information with the halting sequence. By the independence postulate, such constructs are non physical.

Definition 40 (Information of a Set of Points with $\mathcal{H}$) *Given a finite set $D \subset \mathcal{X}$, with $D = \{\alpha_i\}_{i=1}^n$, its mutual information with the halting sequence is defined by*

$$\mathbf{I}(D : \mathcal{H}) = \inf_{\overrightarrow{\alpha_1},\ldots,\overrightarrow{\alpha_n}} \mathbf{I}(\langle \overrightarrow{\alpha_1}, \ldots, \overrightarrow{\alpha_n} \rangle : \mathcal{H}),$$

which is the infimum over all encoded fast Cauchy sequences to members of D. This is a similar construction to Definition 42.

Lemma 25 *Given dual computable measure space $(\mathcal{X}, \mu, \nu)$ with $U = \mu(\mathcal{X}) = \nu(\mathcal{X})$, there is a constant $c_{\mathcal{X},\mu,\nu}$, with universal uniform tests $\mathbf{t}_\mu$ and $\mathbf{t}_\nu$, for a finite set $Z \subset \mathcal{X}$ with $n = \lceil \log |Z| \rceil$,*

$$n < \log \max_{\alpha \in Z} \min\{\mathbf{t}_\mu(\alpha), \mathbf{t}_\nu(\alpha)\} + \log U + \mathbf{I}(\langle Z \rangle : \mathcal{H}) + O(\log \mathbf{I}(\langle Z \rangle : \mathcal{H}) + \mathbf{K}(n)) + c_{\mathcal{X},\mu,\nu}.$$

Proof. Since μ and ν are computable, U is computable. Let $(\{0,1\}^\infty, \mu_\delta, \nu_\delta)$ be a dual binary represention that are isomorphic to computable measure spaces $(\mathcal{X}, \mu)$ and $(\mathcal{X}, \nu)$, with $\delta : (\{0,1\}^\infty, \mu_\delta) \to (\mathcal{X}, \mu)$ and $\delta : (\{0,1\}^\infty, \nu_\delta) \to (\mathcal{X}, \nu)$. If $\max_{\alpha \in Z} \min\{\mathbf{t}_\mu(\alpha), \mathbf{t}_\nu(\alpha)\} = \infty$, then the lemma is proven. Thus for all $\alpha \in Z$, either $\mathbf{t}_\mu(\alpha) < \infty$ or $\mathbf{t}_\nu(\alpha) < \infty$, so by Lemma 19, $\delta^{-1}(\alpha)$ is defined and unique. Let $\overline{\mu_\delta} = \mu_\delta / U$ and $\overline{\nu_\delta} = \nu_\delta / U$ be computable probability measures over $\{0,1\}^\infty$. Let $W = \delta^{-1}(Z) \subset \{0,1\}^\infty$. By Theorem 23 applied to W, $\overline{\mu_\delta}$, and $\overline{\nu_\delta}$ with $s = n - O(1)$, gives

$$s < \max_{\alpha \in W} \min\{\mathbf{D}(\alpha | \overline{\mu_\delta}), \mathbf{D}(\alpha | \overline{\nu_\delta})\} + \mathbf{I}(W : \mathcal{H}) + O(\log \mathbf{I}(W : \mathcal{H}) + \mathbf{K}(s)) + c_{\mathcal{X},\mu,\nu}.$$

Due to Proposition 18,

$$n < \max_{\alpha \in W} \min\{\log \mathbf{t}_{\overline{\mu_\delta}}(\alpha), \log \mathbf{t}_{\overline{\nu_\delta}}(\alpha)\} + \mathbf{I}(W : \mathcal{H}) + O(\log \mathbf{I}(W : \mathcal{H}) + \mathbf{K}(n)) + c_{\mathcal{X},\mu)},$$
$$n < \max_{\alpha \in W} \min\{\log \mathbf{t}_{\mu_\delta}(\alpha), \log \mathbf{t}_{\nu_\delta}(\alpha)\} + \log U + \mathbf{I}(W : \mathcal{H}) + O(\log \mathbf{I}(W : \mathcal{H}) + \mathbf{K}(n)) + c_{\mathcal{X},\mu}.$$

Since $(\{0,1\}^\infty, \mu_\delta)$ is isomorphic to $(\mathcal{X}, \mu)$ and $(\{0,1\}^\infty, \nu_\delta)$ is isomorphic to $(\mathcal{X}, \nu)$, due to Lemma 19,

$$n < \max_{\alpha \in Z} \min\{\log \mathbf{t}_\mu(\alpha), \log \mathbf{t}_\nu(\alpha)\} + \log U + \mathbf{I}(W : \mathcal{H}) + O(\log \mathbf{I}(W : \mathcal{H}) + \mathbf{K}(n) + c_{\mathcal{X},\mu}).$$

Given any encoding of the fast Cauchy sequences of the members of Z, one can compute W with δ^{-1}, thus $\mathbf{K}(W|Z) = O(1)$, so

$$n < \max_{\alpha \in Z} \min\{\log \mathbf{t}_\mu(\alpha), \log \mathbf{t}_\nu(\alpha)\} + \log U + \mathbf{I}(Z : \mathcal{H}) + O(\log \mathbf{I}(Z : \mathcal{H}) + \mathbf{K}(n) + c_{\mathcal{X},\mu}).$$

□

Thus as a consequence to Lemma 25 is that discrete time dynamics will hit ever lower entropy scores $\mathbf{H}_\mu$ as time approaches infinity. A stronger statement can be proven that states the dynamics will repeatedly hit intermediate low $\mathbf{H}_\mu$ scores.

Theorem 76 *Let $(\mathcal{X}, \mu, \nu)$ be a dual computable measure space, with $U = \mu(\mathcal{X}) = \nu(\mathcal{X})$ and $\alpha \in \mathcal{X}$, with finite $\mathbf{I}(\alpha : \mathcal{H})$. For discrete time dynamics G^t, there is a c such that*

$$\max_{\gamma \in G^{\{1,\ldots,2^n\}}\alpha} \max\{\mathbf{H}_\mu(\gamma), \mathbf{H}_\nu(\gamma)\} < \log U - n + O(\mathbf{K}(n)) + c.$$

Proof. Let $Z_n = G^{\{1,\ldots,2^n\}}\alpha$. Lemma 25, applied to $(\mathcal{X}, \mu, \nu)$ and Z_n, results in $\gamma \in Z_n$ such that

$$n < \min\{\log \mathbf{t}_\mu(\gamma), \log \mathbf{t}_\nu(\gamma)\} + \log U + \mathbf{I}(Z_n : \mathcal{H}) + O(\log \mathbf{I}(Z_n : \mathcal{H}) + \mathbf{K}(n)) + c_{\mathcal{X},\mu,\nu,\alpha}.$$

So by using Definition 42, $\mathbf{I}(Z_n : \mathcal{H}) <^+ \mathbf{I}(\alpha : \mathcal{H}) + \mathbf{K}(n)$, and one gets

$$n < \min\{\log \mathbf{t}_\mu(\gamma), \log \mathbf{t}_\nu(\gamma)\} + \log U + \mathbf{I}(\alpha : \mathcal{H}) + O(\log \mathbf{I}(\alpha : \mathcal{H}) + \mathbf{K}(n)) + c_{\mathcal{X},\mu,\nu,\alpha,G}.$$

The theorem is proven by noting $\mathbf{I}(\alpha : \mathcal{H}) < \infty$. □

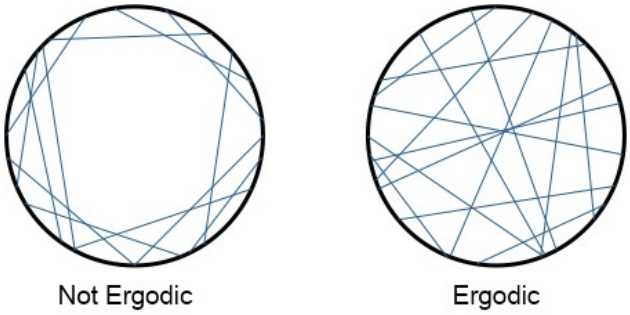


Figure 16.1: The question of ergodicity in a perfectly collisionless ideal gas with specular reflections.

16.2 Ergodic Dynamics

This section deals in the case where the dynamics are ergodic. For a measure space $(\mathcal{X}, \mu)$ a function $T : X \to X$ is ergodic if it is measure preserving and all the invariant sets have measure 0 or $\mu(X)$, as seen in Figure 16.1. Ergodic dynamics are prevalent in many areas of mathematics and physics. For example, particles in box (as seen in Example 5) adhere to ergodic dynamics.

16.2.1 Single Points

The following theorem adapts Theorem 6 of [BDH+12] to computable measure spaces using the recommendations of the proof sketch of Theorem 12. Comparable results can be found in [FMN12]. The main difference between this proof and that of Theorem 6 is that overlapping open balls are used instead of cylinders.

Theorem 77 *Let $(\mathcal{X}, \mu)$ be a computable measure space. Let $T : X \to X$ be an computable ergodic function. Let A be an effectively open subset of X, where $\mu(A) < \mu(X)$. Let A^* be the set of points $x \in X$ such that $T^i(x) \in A$ for all $i \geq 0$. Then $\mathbf{H}_\mu(x) = -\infty$ for all $x \in A^*$.*

It is sufficient to prove A^* is an effectively null set, introduced in Definition 32. We recall that from Corollary 20, there is an enumeration $\{B_i\}$ of the basis of "almost decidable" of open balls such that their borders have null μ-measure. Let $\nu(x) = \mu(x)/\mu(X)$ be a computable probability measure over $\mathcal{X}$, due to Claim 1. Let r be a real number such that $\nu(A) < r < 1$. Given an enumerated ball B_j, we want to find an n such that $\nu(B_j \cap \bigcap_{i \leq n} T^{-i}(A)) \leq r\nu(B)$. Note that it could be that $B_j \cap B_k \neq \emptyset$ for $j \neq k$. This gives an effective open cover of $A^* \cap B_j$ having measure at most $r\nu(B_j)$. For each j you iterate the process until you get the an effectively open cover of $B_j \cap A^*$ with measure $< r2^{-j}\nu(B_j)$. Thus the union of all efffectively open covers of A^* has measure less than r. This process is repeated without end to get an ν effectively null set.

To estimate $\nu(B \cap \bigcap_{i \leq n} T^{-i}(A))$, we note that it does not exceed $\min_{i \leq n} \nu(B \cap T^{-i}(A))$ which does not exceed $\frac{1}{n+1} \sum_{i \leq n} \nu(B \cap T^{-i}(A))$. This average,

$$\frac{1}{n+1} \left[\nu(B \cap A) + \nu(B \cap T^{-1}(A) + \cdots + \nu(B \cap T^{-n}(A))\right] \tag{16.1}$$

is equal to

$$\frac{1}{n+1} \left[\nu(T^{-n}(B) \cap T^{-n}(A)) + \nu(T^{-(n-1))}(B) \cap T^{-n}(A) + \cdots + \nu(B \cap T^{-n}(A))\right],$$

because G is measure preserving. The latter expression is the scalar product of the indicator function of $T^{-n}(A)$ and the average $a_n = (\mathbf{1}_0 + \cdots + \mathbf{1}_n)/(n+1)$, where $\mathbf{1}_i$ is the indicator of $T^{-i}(B)$.

As $n \to \infty$, the average a_n converges in L_2 to the constant function $\nu(B)$ due to von Neumann's mean ergodic theorem. By Cauchy-Schwarz inequality this means the the scalar product converges to $\nu(A)\nu(B)$, so it does not exceed $r\nu(B)$ for n large enough.

It remains to find an effective value for n for which the L_2-distance between a_n and the constant function $\nu(B)$ is small. Note that for all i the set $T^{-i}(B)$ is an effectively open set of measure $\nu(B)$, and, since T is measure preserving, $\nu(B)$ is computable. There for any i and $\epsilon > 0$, one can uniformly approximate $T^{-i}(B)$ by its ad-set subset U (see Definition 31) where $\nu(T^{-i}(B) \setminus U) < \epsilon$ can be computed, due to Proposition 17. This means that the L_2-distance between a_n and the constant function $\nu(B)$ can be computed effectively, so one can continue computing this value until it finds and n such that the average (Equation 16.1) is less than $r\nu(B)$. We then have $\nu(B \cap \bigcap_{i \le n} T^{-i}(A)) < r\nu(B)$. □

The above theorem has implications for algorithmic coarse grain entropy and in particular Theorem 85 which says that if a state travels through enough partitions (effective open sets) then oscillations will occcur. Theorem 77 says that a state $x \in X$, with $\mathbf{H}_\mu(x) \neq -\infty$, under ergodic dynamics will travel through all the partitions if there are finitely many of them or an ever increasing number of partitions if there are infinite many of them.

16.2.2 Indicator Functions

The following theorem adapts Theorem 8 from [BDH+12] to computable measure spaces using the proof sketch in Theorem 12.

Theorem 78 *Let $(\mathcal{X}, \mu)$ be a computable measure space. Let $T : X \to X$ be a computable ergodic function. Let U be an effectively open set. If $\mathbf{H}_\mu(\omega) \neq -\infty$ then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_U(T^k(\omega)) = \mu(U)/\mu(X).$$

This also applies to effectively closed sets.

Proof. Let $\nu(x) = \mu(x)/\mu(X)$ be a computable probability measure over $\mathcal{X}$ due to Claim 1. Let $g_n(\omega) = \frac{1}{n}\sum_{k=0}^{n-1} \mathbf{1}_U(T^k(\omega))$ be the frequency of U elements among the first n iterations of ω. We first prove $\limsup g_n(\omega) \le \nu(U)$. We then prove $\liminf g_n(\omega) \ge \nu(U)$.

Let $r > \nu(U)$ be some rational number and $G_N = \{\omega : (\exists n \ge N) g_n(\omega) > r\}$ be the set of points where some far enough frequency exceeds r. The set G_N is an effectively open set; the functions g_n are lower computable uniformly in n; the condition $g_n(\omega) > r$ is enumerable. The set G_N is decreasing in N. By the classical Birkhoff's pointwise ergodic theorem that $\nu(\bigcap_N G_N) = 0$ as the sequence of functions g_n converges to $\nu(U) < r$ ν-almost everywhere. So there exists N, where $\nu(G_N) < 1$. We can then apply Theorem 77 to effectively open set G_n we get that for $\omega \in U$ with $\mathbf{H}_\mu(\omega) \neq -\infty$ and k such that $T^k(\omega) \notin G_N$. So $\limsup_n g_n(T^k(\omega)) \le r$ Since finite number of iterations does not change lim sup, we have $g_n(\omega) \le r$. Since r was an arbitray rational number greater than $\nu(U)$ so $\limsup g_n(\omega) \le \nu(U)$.

(2) We now prove that $\liminf g_n(\omega) \geq \nu(U)$. Since U is open it is a countable union of almost decidable balls. Taking ad-set $D \subset U$, we can apply the previous statement to $X \setminus \overline{\overline{D}}$. It says the orbit of a point ω with $\mathbf{H}_\mu(\omega) \neq -\infty$ will be in $\overline{\overline{D}}$ with frequence at least $\nu(\overline{\overline{D}}) = \nu(D)$. Since $\nu(D)$ can be arbitrarily close to $\nu(U)$, we have that $\liminf g_n(\omega) \geq \nu(U)$. □

Proposition 30 *For computable non-atomic measure space $(\mathcal{X}, \mu)$,*

(1) Every measurable set E with $\mu(E) > 0$ contains measurable sets of arbitrarily small positive measure.

(2) For any $\delta \in [0, \mu(X)]$, there exists an measurable set A where $\frac{1}{2}\delta \leq \mu(A) \leq \delta$.

Proof.

(1) Let $B_1 \subset E$ be a set such that $0 < \mu(B_1) < \mu(E)$. Then either $\mu(B_1) \leq \mu(E)/2$ and we set $A_1 = B_1$ or $\mu(X \setminus B_1) \leq \mu(E)/2$, and we set $A_1 = X \setminus B_1$. Now repeat the process with A_1 instead of E, obtaining a measurable subset A_2 of A_1, with $0 < \mu(A_2) < \mu(E)/4$. Continuing in this way we see that X contains subsets with arbitrary small measure.

(2) We prove the existence of a measurable set with the desired property and use the fact that μ is regular to imply this set can be open. Let $\mathcal{C}$ be the collection of measurable subsets A of X for which $\mu(A) < \frac{1}{2}\delta$. If $\mathcal{C}$ is not closed under unions then the lemma is proved. For example, if $A, B \in C$ but $A \cup B \notin C$ then $\frac{1}{2}\delta \leq \mu(A \cup B) \leq \delta$. Therefore $\mathcal{C}$ is closed under binary unions. Taking limits, this implies $\mathcal{C}$ is closed under countable unions.

Let $\beta = \sup_{C \in \mathcal{C}} \mu(C)$. There exists a sequence of sets $\{B_i\}$ for which $\mu(B_n) \nearrow \beta$. Let $B = \bigcup B_n$, this implies $\mu(B) = \beta$ and since $B \in \mathcal{C}$, we have $\beta < \frac{1}{2}\delta$. But then by (2) we can find a subset $E \subseteq X \setminus B$ whose measure is less than $\frac{1}{2}\delta - \beta$, which would imply $B \cup E \in \mathcal{C}$ contradicting the fact that B attains $\sup_{C \in \mathcal{C}} \mu(C)$ □

Corollary 25 *For computable non-atomic measure space $(\mathcal{X}, \mu)$, for any $\delta \in [0, \mu(X)]$, there exists an open set A where $\frac{1}{2}\delta \leq \mu(A) \leq \delta$.*

Proof. This follows from Proposition 30 and the fact that every finite Borel measure on a metric space is regular.

Proposition 31 *Given non-atomic computable measurable space $(\mathcal{X}, \mu)$, there is a $c \in \mathbb{N}$, where for all n, $\mu(X)2^{-n-\mathbf{K}(n)-c} < \mu(\{x : \mathbf{H}_\mu(x) < \log \mu(X) - n\}) < \mu(X)2^{-n}$.*

Proof. By Corollary 25, for every $\delta \in [0, \mu(X)]]$, there exists an open set A, with $\frac{1}{2}\delta \leq \mu(A) \leq \delta$. Thus one can uniformly, in $n \in \mathbb{W}$, enumerate an effectively open sets $\{D_n\}$ such that $\mathbf{m}(n)\mu(X)2^{-n-1} < \mu(D_n) < \mathbf{m}(n)\mu(X)2^{-n}$ such that $D_n \cap D_m = \emptyset$ if $n \neq m$, where $\mu(X)$ is computable because μ is computable. The reasoning is as follows.

Let $\{\widehat{D}_n\}$ be current ad-sets all originally $\emptyset$ such that $\widehat{\mathbf{m}}(n)2^{-n-1}\mu(X) < \mu(\hat{D}_n) < \widehat{\mathbf{m}}(n)\mu(X)2^{-n}$, where $\widehat{\mathbf{m}}$ is a lower approximation of $\mathbf{m}$. One can lower compute the interval

$$[\mathbf{m}(n)\mu(X)2^{-n-1}, \mathbf{m}(n)\mu(X)2^{-n}]$$

for all n, and if the interval shifts by some rational amount, by Corollary 25, one can add an ad-set $D \subseteq X \setminus \overline{\overline{\bigcup_{i=1}^{\infty} \widehat{D}_i}}$, such that $\widehat{\mathbf{m}}(n)\mu(X)2^{-n-1} < D \cup \hat{D}_n < \widehat{\mathbf{m}}(n)\mu(X)2^{-n-1}$, and then set $\widehat{D}_n = D \cup \widehat{D}_n$, and continue with the enumeration.

Let the μ-test $t(\alpha) = \sup_{n:\alpha\in D_n} 2^{n-\log\mu(X)}$. Thus since t is lower computable and $\int_X t d\mu \leq \sum_n \mu(D_n)2^n/\mu(X) = \sum_n \mathbf{m}(n) < 1$, we have that $t \stackrel{*}{<} \mathbf{t}_\mu$. Since $\mu\{x : \mathbf{t}_\mu(x) > 2^n/\mu(X)\} < \mu(X)2^{-n}$, we get that there exists $c \in \mathbb{N}$, with $\mu(X)2^{-n-\mathbf{K}(n)-c} < \mu(\{x : \mathbf{H}_\mu(x) < \log\mu(X) - n\}) < \mu(X)2^{-n}$. □

Corollary 26 *Given non-atomic computable measurable space $(\mathcal{X}, \mu)$, there is a $c \in \mathbb{N}$, where for all n, $\mu(X)2^{-n-\mathbf{K}(n)-c} < \mu(\{x : \mathbf{t}_\mu(x) > 2^n/\mu(X)\}) < \mu(X)2^{-n}$.*

The following shows that during the course of ergodic dynamics, the state will be guaranteed to oscillate in its algorithmic fine grained thermodynamic entropy. Small oscillations are frequent, and larger flucuations are more rare. This theorem parallels Theorem 75 in its inequalities.

Theorem 79 (Discrete Oscillations) *Let $(\mathcal{X}, \mu)$ be a non-atomic computable measure space. There is a $c \in \mathbb{N}$ with the following properties. Let $T : X \to X$ be a computable ergodic function, and $U_n = \{x : \mathbf{H}_\mu(x) < \log\mu(X) - n\}$. If $\omega \in X$ has $\mathbf{H}_\mu(\omega) \neq -\infty$,*

$$2^{-n-\mathbf{K}(n)-c} < \lim_{n\to\infty} \frac{1}{n}\sum_{t=0}^{n-1} \mathbf{1}_{U_n}(T^t(\omega)) < 2^{-n}.$$

Proof. By Proposition 31, there is a c where $\mu(X)2^{-n-\mathbf{K}(n)-c} < \mu(U_n) < \mu(X)2^{-n}$. By Theorem 78, $\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbf{1}_{U_n}(T^k(\omega)) = \mu(U_n)/\mu(X)$. So

$$2^{-n-\mathbf{K}(n)-c} < \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \mathbf{1}_{U_n}(T^k(\omega)) < 2^{-n}.$$

16.2.3 Lower Computable Functions

The following theorem adapts Theorem 9 in [BDH+12] to computable measure spaces.

Theorem 80 *Let $(\mathcal{X}, \mu)$ be a computable measure space. Let $T : X \to X$ be a computable ergodic function. Let $f : X \to \mathbb{R}^+ \cup \infty$ be lower computable. If $\mathbf{H}_\mu(\omega) \neq -\infty$,*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n=1} f(T^t(\omega)) = \frac{1}{\mu(X)}\int f d\mu.$$

Proof. Let $\nu = \mu/\mu(\mathcal{X})$ be a computable measure over $\mathcal{X}$, due to Claim 1. Let f be a lower computable function with a finite integral. Let $f_n = \frac{1}{n}(f + \cdots + f\cdot T^{n-1})$. Let $r > \int f d\nu$ be a rational number and $G_N = \{\omega : (\exists n \geq N) f_n(\omega) > r\}$. The set G_N is effectively open and $\nu(\bigcap_N G_N) = 0$ as $f_n(\omega) = \int f d\nu < r$ for ν-almost every ω by the classical version of Birkoff's ergodic theorem. As a result, there exists N where $\nu(G_N) < 1$. By Theorem 77, if $\mathbf{H}_\nu(\omega) \neq -\infty$, then there exists k such that $T^k(\omega) \neq G_n$. So $\limsup f_n(T^k(\omega)) \leq r$ and $\limsup f_n(\omega) = \limsup f_n(T^k(\omega)) \leq r$. Since $r > \int f d\nu$ can be arbitrarily close to the integral, we have that $\limsup f_n(\omega) <^+ \int f d\nu = \frac{1}{\mu(X)}\int f d\mu$.

It remains to prove that $\liminf f_n(\omega) \geq \int d\nu$. This is true for any lower semicontinuous f. Consider some lower bound for f that is of the form $\hat{f}(\omega) = \sum_{i=1}^n c_n \mathbf{1}_{B_n}(\omega)$, where each B_n is an almost decidable ball. For these basic functions the statement of the theorem is true using the reasoning of Theorem 78, and their integrals can be arbitrarily close to $\int f\nu = \frac{1}{\mu(X)}\int f d\mu$. □

Exercise 24 *Let $(\mathcal{X}, \mu)$ be a computable measure space and $T : X \to X$ be a computable ergodic function. Prove that for $\omega \in X$ with $\mathbf{H}_\mu(\omega) \neq -\infty$, then*

$$\mu(X) \lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} 2^{-\mathbf{H}_\mu(T^t\omega)} < 1.$$

Chapter 17

Dynamics on Product Spaces

Theorem 75 proves the existence of oscillations in thermodynamic entropy. However the situation becomes more complicated if one were to examine product spaces. If a product state is typical of the space then its marginal entropies cannot oscillate in synch. Furthermore if a joint system is typical and non-exotic, then the independent systems will evolve in time such that their entropies are out of sync. For example take N particles in a box. If one were to split the box down the middle and let the $N/2$ particles of each subsystem evolve separately, then almost surely the two subsystems will have times where their entropies are different by an increasing amount. This is one of the central results of the manuscript. In this chapter, we prove this fact for two cases. The first case is using continuous dynamics and the second case is using discrete ergodic dynamics. In both such cases, marginal entropies cannot be synchronized.

17.1 Continuous Dynamics

The first result deals with continuous dynamics according to a transformation group (see Definition 36).

17.1.1 Exotic Intervals

Definition 41 *An r-interval $v \subseteq (0,1)$ is a finite collection of open intervals with rational endpoints.*

Lemma 26 *For r-interval v, Lebesgue measure L, there exists a rational number $r \in v$ such that $\mathbf{K}(r) <^{\log} -\log L(v) + \mathbf{I}(\langle r \rangle; \mathcal{H})$.*

Proof. Let $n = \lceil -\log L(v) \rceil$. We condition the universal Turing machine on n, which can be done given the precision of the theorem. Let Q be an elementary probability measure that realizes the stochasticity of v (see Chapter 1), and $d = \max\{\mathbf{d}(r|Q), 1\}$ and $\mathbf{Ks}(x) = \mathbf{K}(Q) + 3\log d$. Without loss of generality, the support of Q can be assumed to consist entirely of r-intervals s with $n = \lceil -\log L(s) \rceil$. We sample $d2^n$ real numbers R in the interval $(0,1)$ using L. For r-interval s, $R \subset \mathbb{R}$, we define the indicator function $\mathbf{i}(s,R) = [R \cap s = \emptyset]$.

$$\mathbf{E}_{R \sim L(0,1)^{d2^n}} \mathbf{E}_{s \sim Q} \mathbf{i}(s,R) = \mathbf{E}_{s \sim Q} \mathbf{E}_{R \sim L(0,1)^{d2^n}} \mathbf{i}(s,R) = (1-s)^{d2^n} \leq (1-2^{-n})^{d2^n} < e^{-d}.$$

Thus given Q and d, one can determine a set of $d2^n$ rationals $W \subset \mathbb{Q}$, such that $\mathbf{E}_{s\sim Q}\mathbf{i}(s, W) < e^{-d}$. Thus $t(s) = \mathbf{i}(s, W)e^d$ is a Q-test, with $\mathbf{E}_{s\sim Q}[t(s)] < 1$. It must be that $t(v) = 0$, otherwise

$$1.44d \leq \log t(v) <^+ \mathbf{d}(v|Q, d) <^+ d + \mathbf{K}(d),$$

which is a contradiction for large d which one can assume without loss of generality. Thus there exists a rational $r \in W \cap v$ such that, using Lemma 3,

$$\begin{aligned}\mathbf{K}(r) &<^+ \log |W| + \mathbf{K}(W)\\ &<^+ n + \log d + \mathbf{K}(d, Q)\\ &<^+ n + 3\log d + \mathbf{K}(Q)\\ &<^+ n + \mathbf{Ks}(v)\\ &<^{\log} n + \mathbf{I}(\langle v\rangle; \mathcal{H}).\end{aligned}$$

□

Exercise 25 *A b-interval is a finite collection of rectangles with rational coordinates in* $[0, 1]\times[0, 1]$. *Show that given b-interval* w, *there exists two rationals* $x, y \in [0, 1]$ *such that* $(x, y) \in w$ *and* $\mathbf{K}(x, y) <^{\log} -\log L_2(w) + \mathbf{I}(w; \mathcal{H})$, *where* L_2 *is the 2D Lebesgue measure.*

17.1.2 No Synchronization Under Continuous Dynamics

The following theorem states that, for typical systems, the marginal entropies will be increasing out of sync. The theorem uses the following definition of the mutual information of two points in metric spaces with the halting sequence.

Definition 42 (Mutual Information with the Halting Sequence) *An encoding of a fast Cauchy sequence* $\overrightarrow{x}$ *is* $\langle\overrightarrow{x}\rangle \in \{0, 1\}^\infty$, *with* $\langle\overrightarrow{x}\rangle = \langle x_1\rangle\langle x_2\rangle\ldots$. *Each* $x_i \in \overrightarrow{x}$ *is an ideal point, and* i *is its order in the enumeration of* $\overrightarrow{x}$. *A pair points* $x, y \in \mathcal{X}$ *has a certain mutual information with the halting sequence* $\mathbf{I}((x, y) : \mathcal{H}) = \inf\{\mathbf{I}(\langle\overrightarrow{x}\rangle\langle\overrightarrow{y}\rangle : \mathcal{H}) : \overrightarrow{x}, \overrightarrow{y}$ *are fast Cauchy sequences for* x *and* $y\}$.

Definition 43

Theorem 81 *Let* $(\mathcal{X}\times\mathcal{Y}, \mu\times\nu)$ *be a computable product measure space. Let* G^t *be a transformation group with* $(\alpha^t, \beta^t) = G^t(\alpha, \beta)$. *If* $\mathbf{H}_\mu(\alpha, \beta) > -\infty$ *and* $\mathbf{I}((\alpha, \beta) : \mathcal{H}) < \infty$ *then* $\sup_{t\in[0,1]} \left|\mathbf{H}_\mu(\alpha^t) - \mathbf{H}_\nu(\beta^t)\right| = \infty$.

Proof. Let $\overline{\mu} = \mu/\mu(X)$ and $\overline{\nu} = \nu/\nu(Y)$ be computable probability measures due to Claim 1. Assume not and $d = \lceil\sup_{t\in[0,1]} \left|\mathbf{H}_{\overline{\mu}}(\alpha^t) - \mathbf{H}_{\overline{\nu}}(\beta^t)\right|\rceil < \infty$. Let $U_n \subset [0, 1]$ be an open set where $U_n = \{t : \mathbf{H}_{\overline{\mu}}(\alpha^t) < -n\}$. By Corollary 23, there is a $c \in \mathbb{N}$ where for Lebesgue measure L, $2^{-n-2\log n-c} < L(U_n)$.

Given (α, β) and n, one can enumerate an increasing r-interval $v \subseteq U_n$ and stop when $L(v) > L(2^{-n-2\log n-c-1})$. By Lemma 26, there exists a rational $r \in v$, with

$$\mathbf{K}(r) <^{\log} n + \mathbf{I}(r; \mathcal{H}) <^{\log} n + \mathbf{I}((\alpha, \beta) : \mathcal{H}) + \mathbf{K}(n, c) <^{\log} n + \mathbf{I}((\alpha, \beta) : \mathcal{H}),$$

where $\mathbf{K}(c)$ is folded into the additive constants. Let $A_n = \{\gamma : \mathbf{t}_{\overline{\mu}}(\gamma) > 2^n\}$ and $B_n = \{\gamma : \mathbf{t}_{\overline{\nu}}(\gamma) > 2^n\}$. By the definition of tests, $\mu\times\nu(A_n \times B_n) < 2^{-2n}$. This enables us to create the following

$\overline{\mu} \times \overline{\nu}$, test

$$\begin{aligned} t_n(\gamma, \lambda) &= [(\gamma, \lambda) \in G^{-r}(A_n \times B_{n-d})] 2^{2n-d} \\ \mathbf{K}(t_n) &<^{+} \mathbf{K}(r, n) <^{\log} n + \mathbf{I}((\alpha, \beta) : \mathcal{H}) \\ &<^{\log} n, \end{aligned} \tag{17.1}$$

where Equation 17.1 is due to the assumption that $\mathbf{I}((\alpha, \beta) : \mathcal{H}) < \infty$. It must be that $t_n(\alpha, \beta) \neq 0$ because $r \in U_n$. Thus

$$\begin{aligned} \mathbf{t}_{\overline{\mu} \times \overline{\nu}}(\alpha, \beta) &\stackrel{*}{>} \sum_n \mathbf{m}(t_n) t_n(\alpha, \beta) \\ &\stackrel{*}{=} \sum_n 2^{-n - O(\log n)} 2^{2n-d} \\ &= \infty. \end{aligned}$$

Thus $\mathbf{H}_{\overline{\mu} \times \overline{\nu}}(\alpha, \beta) = -\infty$, which means $\mathbf{H}_{\mu \times \nu}(\alpha, \beta) = -\infty$, causing a contradiction. □

17.2 Ergodic Dynamics

17.2.1 Effective Convergence Time of Ergodic Dynamics

To prove properties of ergodic dynamics, an effective means of determining the convergence of Birkoff averages was needed, and thus the results from [GHR10] were leveraged. So the contents in this subsection reviews this referenced work.

Definition 44
Let $(\mathcal{X}, \mu)$ be a computable measure space.

- *Random variables f_n effectively converge in probability to f if for each $\epsilon > 0$, $\mu\{x : |f_n(x) - f(x)| < \epsilon\}$ converges effectively to 1, uniformly in ϵ. Thus there is a computable function $n(\epsilon, \delta)$ where for all $n \geq n(\epsilon, \delta)$, $\mu\{|f_n - f| \geq \epsilon\} < \delta$.*
- *Random variables f_n effectively converge almost surely to f if $f'_n = \sup_{k \geq n} |f_k - f|$ effectively converge in probability to 0.*
- *A simple function $f : X \to \mathbb{Q}$ is of the form $f(x) = \max_{i=1}^{n} q_i \mathbf{1}_{B_i}(x)$, where each q_i is a rational number an B_n is an almost decidable ball.*
- *The quotient space $L^1(X, \mu)$ can be made into computable metric space. Let $d_1(f, g) = \int_X |f - g| d\mu$ be the distance metric, with the relation $f \sim g$ if $d_1(f, g) = 0$. The ideal points are simple functions. An integral function $f : X \to \mathbb{R}_{>0}$ is $L^1(X, \mu)$ computable if it is a computable point of the computable metric space $L^1(X, \mu)$. Basic operations $L^1(X, \mu)$ such as addition, multiplication by a scalar, min, max, are computable. If f is computable, then so is $\int f d\mu$.*

Proposition 32 *If f is a computable element of computable metric space $L^1(X, \mu)$ and if $T : X \to X$ is computable and μ-measure preserving, then $f \circ T$ is computable.*

Proof. Since f is computable, there is a computable fast Cauchy sequences of simple functions $(f_1, f_2, \dots)$ converging to f. Given a simple function $g(x) = q\mathbf{1}_B(x)$, we will show how to compute a simple function h such that $d_1(h, g \circ T)$ can be arbitrarily small. The process can be easily generalized to arbitrary simple functions and thus to each f_i. Since B is almost-decidable $\mu(B)$ can be computed, and thus the effectively open set $T^{-1}(B)$ can be enumerated by ad-sets A such that $\mu(A)$ is arbitrarily close to $\mu(B)$. Thus one can get the simple function $\ell(x) = q\mathbf{1}_A(x)$ such that $d_1(\ell, g)$ is arbtitrarily close. □

Proposition 33 *Let $(\mathcal{X}, \mu)$ be a computable metric space and $T : X \to X$ be a computable ergodic function. Let f be a computable element of $L^1(X, \mu)$. The L^1 convergence of Birkhoff averages of f is effective.*

Proof. Replacing f with $f - \int f d\mu$, we can assume $\int f d\mu = 0$. Let $A_n = (f + f \circ T + \cdots + f \circ T^{n-1})/n$. The sequence $\|A_n\|$ is computable and converges to 0 by Birkoff's Ergodic Theorem. Given $p \in \mathbb{N}$ we have $m \in \mathbb{N}$ where $m = np + k$, with $0 \leq K < p$. So

$$\begin{aligned} A_{np+k} &= \frac{1}{np_k}\left(\sum_{i=0}^{n-1} pA_p \circ T^{pi} + kA_k \circ T^{pn}\right) \\ \|A_{np+k}\| &= \frac{1}{np_k}\left(np\|A_p\| + k\|A_k\|\right) \\ &\leq \|A_p\| + \frac{\|A_k\|}{n} \\ &\leq \|A_p\| + \frac{\|f\|}{n} \end{aligned}$$

Let $\epsilon >)$. We can copmute some $p = p(\epsilon)$ such that $\|A_p\| < \epsilon/2$. Then we can compute some $n(\epsilon) \geq \frac{2}{\epsilon}\|f\|$. The function $m(\epsilon) = n(\epsilon)p(\epsilon)$ is computable and for all $m \geq m(\epsilon)$, $\|A_m\| \leq \epsilon$. □

Lemma 27 (Maximal Ergodic Theorem) *For $f \in L^1(X, \mu)$ and $\delta > 0$, $\mu\{x : \sup_n |A_n^f(x) > \delta\} \leq \frac{1}{\delta}\|f\|_1$.*

Theorem 82 *Let $(\mathcal{X}, \mu)$ be a computable measure space and $T : X \to X$ be a computable ergodic function. If f is $L^1(X, \mu)$- computable, then the Birkoff average effectively converge almost surely.*

Proof. Let $\epsilon, \delta > 0$. Compute p such that $\|A_p^f\| \leq \delta\epsilon/2$. Applying the maximal ergodic theorem with $g = A_p^t$ has

$$\mu\{x : \sup_n |A_n^g(x)| > \delta/2\} \leq \epsilon. \tag{17.2}$$

One has that

$$A_n^g = A_n^f + \frac{u \circ T^n - u}{np},$$

where $u = (p-1)f + (p-2)f \circ T + \cdots + f \circ T^{p-2}$. $\|u\|_\infty \leq \frac{p(p-1)}{2}\|f\|_\infty$ so if $n \geq n_0 \geq 4(p-1)\|f\|_\infty/\delta$ then $\|A_n^g - A_n^f\|_\infty \leq \delta/2$. So if $|A_n^f(x)| > \delta$ for some $n \geq n_0$, then $|A_n^g(x)| > \delta/2$. Using Equation 17.2, we get

$$\mu\{x : \sup_{n \geq n_0} |A_n^f(x)| > \delta\} \leq \epsilon.$$

As n_0 can be computed from δ and ϵ, we prove the theorem. □

17.2.2 Busy Beaver Functions

We review the material on busy beaver functions, detailed in 15.2. Let $\Omega = \sum\{2^{-\|p\|} : U(p) \text{ halts}\}$ be Chaitin's Omega, $\Omega_n \in \mathbb{Q}_{\geq 0}$ be be the rational formed from the first n bits of Ω, and $\Omega^t = \sum\{2^{-\|p\|} : U(p) \text{ halts in time } t\}$. For $n \in \mathbb{N}$, let $\mathbf{bb}(n) = \min\{t : \Omega_n < \Omega^t\}$. $\mathbf{bb}^{-1}(m) = \arg\min_n\{\mathbf{bb}(n-1) < m \leq \mathbf{bb}(n)\}$. Let $\Omega[n] \in \{0,1\}^*$ be the first n bits of Ω.

Lemma. *For $n = \mathbf{bb}^{-1}(m)$, $\mathbf{K}(\Omega[n]|m,n) = O(1)$.*

17.2.3 Marginal Entropies of Ergodic Dynamics

Theorem 83 *Let $(\mathcal{X} \times \mathcal{Y}, \mu \times \nu)$ be a non-atomic computable product measure space. Let G^t be an ergodic transformation group. Let $(\alpha, \beta) \in X \times Y$, with $(\alpha^t, \beta^t) = G^t(\alpha, \beta)$. If $\mathbf{H}_{\mu\times\nu}(\alpha,\beta) > -\infty$ and $\mathbf{I}((\alpha,\beta) : \mathcal{H}) < \infty$ then $\sup_{t\in\mathbb{N}} |\mathbf{H}_\mu(\alpha^t) - \mathbf{H}_\nu(\beta^t)| = \infty$.*

Proof. Let $\overline{\mu} = \mu/\mu(X)$ and $\overline{\nu} = \nu/\nu(Y)$ be computable probability measures due to Claim 1. Assume not. Then there exists $c \in \mathbb{N}$, $c > \lceil \max_t |\mathbf{H}_{\overline{\mu}}(\alpha^t) - \mathbf{H}_{\overline{\nu}}(\beta^t)| \rceil$. Fix $n \in \mathbb{N}$. Let $U_n = \{(\gamma,\lambda) : \mathbf{H}_{\overline{\mu}}(\gamma) < -n\}$. By Proposition 31, there exists $d \in \mathbb{N}$ where $\overline{\mu}(U_n) > 2^{-n-2\log n - d}$. Given n, one can compute an ad-set $V_n \subset U_n$ with computable $p_n = -\log \mu(V_n)$ and $n + 2\log n + d < p_n < n + 2\log n + d + 1$. Let $B^n_m = (\mathbf{1}_{V_n} + \mathbf{1}_{V_n} \circ T + \cdots + \mathbf{1}_{V_n} \circ T^{m-1})/m$, which is computable, as T is measure preserving and due to Proposition 32. By Theorem 82, given $\delta, \epsilon > 0$, there is a computable $m(\delta,\epsilon,n)$ such that $\mu\{(\gamma,\lambda) : \sup_{m>m(\delta,\epsilon,n)} |B^n_m(\gamma,\lambda) - 2^{-p_n}| > \delta\} < \epsilon$. Let $m_n = m(2^{-p_n} - 2^{-1.5p_n}, 2^{-n}, n)$. Let $W_n = \{(\gamma,\lambda) : \sup_{m>m_n} |B^n_m(\gamma,\lambda) - 2^{-p_n}| > 2^{-p_n} - 2^{-1.5p_n}\}$. Either (1) there is an infinite number of n where $(\alpha,\beta) \in W_n$, or (2) there is an infinite number of n where $(\alpha,\beta) \notin W_n$.

Case (1). Each W_n is an effectively open set, computable uniformly in n. Furthermore, $\mu(W_n) < 2^{-n}$. Thus $t(\gamma,\lambda) = \sup_n[(\gamma,\lambda) \in W_n]\mathbf{m}(n)2^n$ is a $\overline{\mu}\times\overline{\nu}$ test. So $\infty = t(\alpha,\beta) \stackrel{*}{<} \mathbf{t}_{\overline{\mu}\times\overline{\nu}}(\alpha,\beta)$, which implies $\mathbf{H}_{\mu\times\nu}(\alpha,\beta) = -\infty$, causing a contradiction.

Case (2). Fix one such $n \in \mathbb{N}$, where $(\alpha,\beta) \notin W_n$. Thus $\sup_{m>m_n} |2^{-p_n} - B^n_m(\alpha,\beta)| \leq 2^{-p_n} - 2^{-1.5p_n}$ implies $\sup_{m>m_n} B^n_m(\alpha,\beta) \geq 2^{-1.5p_n}$. Each $T^{-\ell}V_n$ is an effectively open set, uniformly in k and ℓ. So for all $m > m_n$, there are at least $2^{-1.5p_n}m$ indices ℓ, where $(\alpha,\beta) \in T^{-\ell}V_n$. Let $b_n = \mathbf{bb}^{-1}(m_n+1)$ and N be the smallest power of 2 not less than $\mathbf{bb}(b_n)$. Thus, due to Lemma 22, $\mathbf{K}(N|(\alpha,\beta)) <^+ \mathbf{K}(n,b_n)$. Thus there are at least $2^{-1.5(n+2\log n+d+1)}N$ indices $\ell \in [1,\ldots,N]$ where $(\alpha,\beta) \in T^{-\ell}V_n$. Let $D \subseteq \{0,1\}^{\log N}$, where if $x \in D$ then $(\alpha,\beta) \in T^{-\mathrm{Num}(x)}V_n$ and $|D| \geq 2^{-1.5(n+2\log n+d+2)}N$. The function $\mathrm{Num} : \{0,1\}^{\log N} \to \{1,2,\ldots,N\}$ converts strings to numbers in the natural way. Thus $\mathbf{K}(D|(\alpha,\beta)) <^+ \mathbf{K}(n,b_n)$. This is because $T^{-\ell}V_n$ are effectively open sets, uniformly in ℓ and it is guaranteed that at least $2^{-1.5(n+2\log n+d+1)}N$ indices ℓ have $(\alpha,\beta) \in T^{\ell}V_n$. So after $2^{-1.5(n+2\log n+d+2)}N$ indices have been found, they can be collected into a set D. Let $\mathrm{Uniform}(N)$ be the uniform measure over $\{0,1\}^{\log N}$. By the EL Theorem (Corollary 47), applied to $\mathrm{Uniform(N)} \stackrel{*}{<} \mathbf{m}/\mathbf{m}(N)$, and the definition of $\mathbf{I}$, there exists $x_n \in D$, with

$$\begin{aligned}
\mathbf{K}(x_n) &<^{\log} \mathbf{K}(\mathrm{Uniform}(N)) - \log|D| + \mathbf{I}(D;\mathcal{H}) \\
&<^{\log} \mathbf{K}(N) + 1.5n + 3\log n + \mathbf{I}((\alpha,\beta) : \mathcal{H}) + \mathbf{K}(n,b_n) \\
&<^{\log} \mathbf{K}(\Omega[b_n]) + 1.5n + \mathbf{I}((\alpha,\beta) : \mathcal{H}) + \mathbf{K}(b_n). \qquad (17.3)
\end{aligned}$$

Due to Lemma 22, $\mathbf{K}(\Omega[b_n]|(\alpha,\beta),n,b_n) = O(1)$. Furthermore, it is well known that for bits of Chaitin's Omega, $\mathbf{K}(\Omega[b_n]|\mathcal{H}) <^+ \mathbf{K}(b_n)$ and that $b_n <^+ \mathbf{K}(\Omega[b_n])$. So

$$b_n <^+ \mathbf{K}(\Omega[b_n]) <^{\log} \mathbf{I}(\Omega[b_n];\mathcal{H}) <^{\log} \mathbf{I}((\alpha,\beta):\mathcal{H}) + \mathbf{K}(b_n,n) <^{\log} \mathbf{I}((\alpha,\beta):\mathcal{H}) + \mathbf{K}(n) \tag{17.4}$$

Combining Equations 17.3 and 17.4 together, we get

$$\mathbf{K}(x_n) <^{\log} 1.5n + 2\mathbf{I}((\alpha,\beta):\mathcal{H}).$$

We define the test

$$\begin{aligned}
&t_{n,y}(\gamma,\lambda) = \left[\mathbf{H}_{\overline{\mu}}(T^{\mathrm{Num}(y)}(\gamma,\lambda)_1) < -n \text{ and } \mathbf{H}_{\overline{\nu}}(T^{\mathrm{Num}(y)}(\gamma,\lambda)_2) < -n+c\right] 2^{2n-c},\\
&\mathbf{t}_{\overline{\mu}\times\overline{\nu}}(\gamma,\lambda) \stackrel{*}{>} \sum_n \mathbf{m}(t_{n,x_n}) t_{n,x_n}(\gamma,\lambda),\\
&\stackrel{*}{>} \sum_n \left[\mathbf{H}_{\overline{\mu}}(T^{\mathrm{Num}(x_n)}(\gamma,\lambda)_1) < -n \text{ and } \mathbf{H}_{\overline{\nu}}(T^{\mathrm{Num}(x_n)}(\gamma,\lambda)_2) < -n+c\right] \frac{2^{.5n-2\mathbf{I}((\alpha,\beta):\mathcal{H})}}{(n+\mathbf{I}((\alpha,\beta):\mathcal{H}))^{O(1)}}.
\end{aligned}$$

In recap, since $(\alpha,\beta) \notin W_n$, $|B^n_N(\alpha,\beta) - 2^{-p_n}| > 2^{-p_n} - 2^{-1.5p_n}$, so $B^n_N > 2^{-1.5p_n} > 2^{-1.5(n+2\log n+d)}$. Thus one can create a large enough set $D \subset \{0,1\}^N$, and find a simple enough $x_n \in D$ such that $(\alpha,\beta) \in T^{-\mathrm{Num}(x_n)}V_n$. By the assumptions of the theorem

$$\mathbf{H}_{\overline{\mu}}(T^{\mathrm{Num}(x_n)}(\alpha,\beta)_1) < -n \text{ and } \mathbf{H}_{\overline{\nu}}(T^{\mathrm{Num}(x_n)}(\alpha,\beta)_2) < -n+c.$$

Thus $\mathbf{m}(t_n,x_n)t_{n,x_n}(\alpha,\beta) = \frac{2^{.5n-2\mathbf{I}((\alpha,\beta):\mathcal{H}}}{(n+\mathbf{I}((\alpha,\beta):\mathcal{H}))^{O(1)}}$. Furthermore, since $\mathbf{I}(\alpha,\beta) < \infty$ and there is an infinite number of n where $(\alpha,\beta) \notin W_n$, $\mathbf{t}_{\overline{\mu}\times\overline{\nu}}(\alpha,\beta) = \infty$, so $\mathbf{H}_{\mu\times\nu}(\alpha,\beta) = -\infty$, causing a contradiction. □

Corollary 27 (Independent Systems) *Let $(\mathcal{X}\times\mathcal{Y},\mu\times\nu)$ be a non-atomic computable product measure space. Let G^t_X and G^t_Y be ergodic transformation groups over $\mathcal{X}$ and $\mathcal{Y}$ respectively. Let $(\alpha,\beta) \in X\times Y$. If $\mathbf{H}_{\mu\times\nu}(\alpha,\beta) > -\infty$ and $\mathbf{I}((\alpha,\beta):\mathcal{H}) < \infty$ then $\sup_{t\in\mathbb{N}} |\mathbf{H}_\mu(G^t_X(\alpha)) - \mathbf{H}_\nu(G^t_Y(\beta))| = \infty$.*

The above theorem has immediate applications to the Cantor space. Let σ be the shift operator and $(\alpha,\beta) = \alpha[1]\beta[1]\alpha[2]\beta[2]\ldots$ Let λ be the uniform distribution over $\{0,1\}^\infty$.

Exercise 26 *Show that if $(\alpha,\beta) \in \{0,1\}^\infty \times \{0,1\}^\infty$ is ML Random and $\mathbf{I}((\alpha,\beta):\mathcal{H}) < \infty$ then $\sup_n |\mathbf{D}(\sigma^{(n)}\alpha|\lambda) - \mathbf{D}(\sigma^{(n)}\beta|\lambda)| = \infty$.*

Chapter 18

Algorithmic Coarse Grained Entropy

18.1 Fundamentals of Coarse Grained Entropy

Coarse grained entropy was introduced in [Gac94] as an update to Boltzmann entropy. The goal was a parameter independent formulation of entropy. It was defined using cells. In this section we define coarse grained entropy with respect to open sets, leveraging Chapter 13. Let $\Pi(\cdot)$ be a set of disjoint uniformly enumerable open sets in the computable metric space $\mathcal{X}$, as shown in Figure 18.1.

Definition 45 (Algorithmic Coarse Grained Entropy) $\mathbf{H}_\mu(\Pi_i) = \mathbf{K}(i|\mu) + \log \mu(\Pi_i)$.

Remark 6 (Paradox of Typicality) *Discrete dynamics are subject to a paradox of typicality, which is due to a remark by Vladimir Vovk and detailed in [Gac94]. This remark generalizes Example 6. Lets say P is a partition and state ω is a typical state in partition Γ, which has low entropy, thus $\mathbf{H}_\mu(\Gamma)$ is low. Thus $\mathbf{H}_\mu(\alpha) \approx \mathbf{H}(\Gamma)$. Say ω is subject to discrete time dynamics, for some simple $t \in \mathbb{N}$. This new state $\omega' = G^t\omega$ is in a partition Γ' of much greater entropy. Since t is simple, by Proposition 21, $\mathbf{H}_\mu(\omega')$ is still low. However the coarse grain entropy of ω' is $\mathbf{H}_\mu(\Gamma')$ which is high. So the locally typical state ω turns into locally non-typical state ω'. Assuming only the macroscopic variablesof ω' are detectable, its low algorithmic fine grain entropy and history is inaccessible to observers.*

Note that this violation of typicality is in effect for a single, simple transform. For continuous transform groups or the averages ergodic discrete ones, this phenomenon is not guaranteed to occur.

Coarse grained entropy is an excellent approximation of fine grained entropy, as shown by Proposition 34 and Lemma 29.

Proposition 34 *Let $(\mathcal{X}, \mu)$ be a computable measure space. If $\mu(\Pi_i)$ is uniformly computable and $\alpha \in \Pi_i$ then $\mathbf{H}_\mu(\alpha) <^+ \mathbf{H}_\mu(\Pi_i) + \mathbf{K}(\Pi)$.*

Proof. Let $t(\alpha) = \sum_i [\alpha \in \Pi_i]\mathbf{m}(\Pi_i)/\mu(\Pi_i)$. t is lower semi-computable and $\int_{\mathcal{X}} t(\alpha)d\mu(\alpha) = \sum_i \int_{\Pi_i} (\mathbf{m}(\Pi_i)/\mu(\Pi_i)d\mu(\alpha) = \sum_i \mathbf{m}(\Pi_i) \leq 1$. Thus $\mathbf{t}_\mu(\alpha) \stackrel{*}{>} t(\alpha)$.

Example 6 *We detail changes in fine and coarse grained entropy during the course of dynamics, adapted from Example 6.7 from [Gac94], and described in Figure 18.2. Let $(\{0,1\}^{\mathbb{Z}}, \mu)$ be a computable measure space consisting of all bi-infinite sequences. The distance metric between*

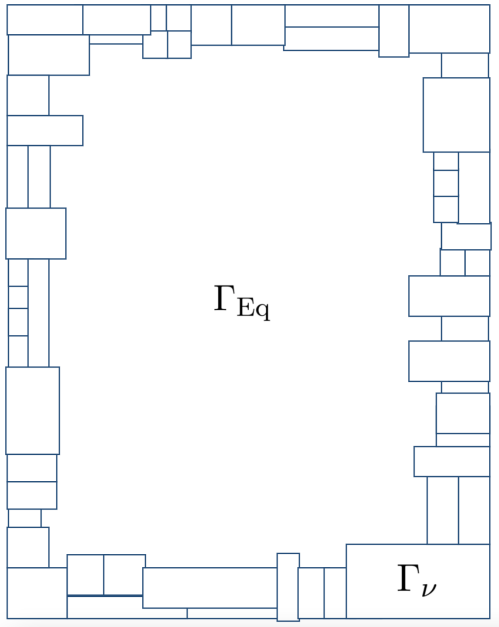


Figure 18.1: Partition of phase space into macro sets Γ_ν , with the thermal equilibrium set Γ_{Eq} taking up most of the volume.

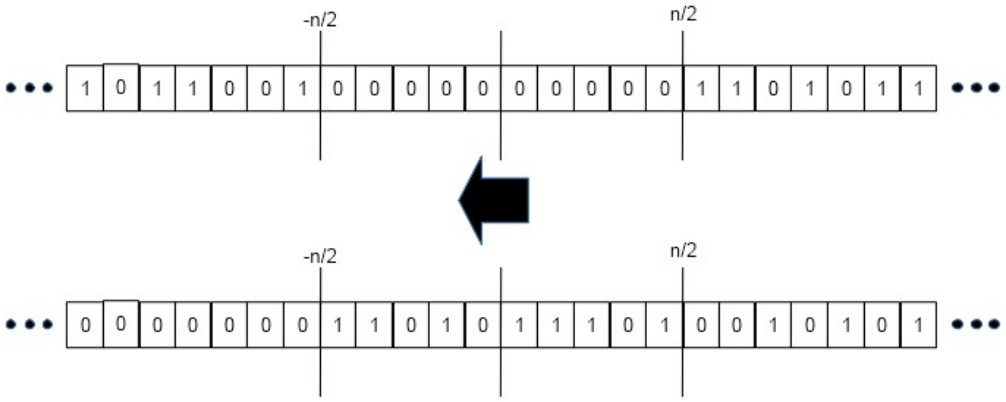


Figure 18.2: The following figure shows an example of changes in coarse grained entropy and fine grained entropy. One the first bi-infinite sequence, there are zeros between the position $-n/2$ and $n/2$ are 0, and beyond that are random. Thus both the coare grained entropy using cells of length n and the fine grained entropy are zero. After dynamics causing the sequence to shift by n positions, the coarse grained entropy is high but the fine grained entropy is still low.

two points $\alpha, \beta \in \{0,1\}^{\mathbb{Z}}$, is $d(\alpha,\beta) = \exp(-\arg\max_n \alpha[-n/2, n/2] = \beta[-n/2, n/2])$. We use the uniform distribution, where the measure of a cylinder $C = \{0,1\}^{\mathbb{N}} x \{0,1\}^{\mathbb{N}}$ is $\mu(C) = 2^{-\|x\|}$. The dynamics is the shift operation, a.k.a. Baker's map. We use the partitions Γ_x, over $x \in \{0,1\}^n$, where $\omega \in \Gamma_x$ if $\omega[-n/2, n/2] = x$. Thus for all $x \in \{0,1\}^n$, $\mu(\Gamma_x) = 2^{-n}$. Let $\omega \in \{0,1\}^{\mathbb{Z}}$, where $\omega[-n/2, n/2] = 0^n$, and all other bits are ML random. Thus the coarse grained entropy of ω is equal to

$$\mathbf{H}_\mu(\Gamma_{0^n}) =^+ \mathbf{K}(0^n) + \log \mu(\Gamma_{0^n}) =^+ \mathbf{K}(n) - n.$$

Since $\mathbf{K}(\mu(\Gamma_x)) =^+ \mathbf{K}(n)$, due to Proposition 34,

$$\mathbf{H}_\mu(\omega) <^+ 2\mathbf{K}(n) - n.$$

Now suppose we subject ω to dynamics, by applying the shift operator n times to produce ω'. Thus $\omega' \in \Gamma_x$, for some random string $x \in \{0,1\}^n$. Thus ω' has algorithmic coarse grained entropy,

$$\mathbf{H}(\Gamma_x) =^+ \mathbf{K}(x) + \log \mu(\Gamma_x) =^+ 0.$$

However, while the coarse entropy has changed dramatically for ω', its fine grained entropy has only gradually shifted. Due to Proposition 21,

$$\mathbf{H}_\mu(\omega') <^+ \mathbf{K}(n) + \mathbf{H}_\mu(\omega) <^+ 3\mathbf{K}(n) - n.$$

Lemma 28 *For computable measure space $(\mathcal{X}, \mu)$, for lower computable function f, and enumerable open set U, $\int_U f d\mu$ is lower computable.*

Proof. For a finite union of balls $V = \bigcup_{j=1}^n B_{i_j}$ and an enumerable open set $W = \bigcup_{j=1}^\infty B_{k_j}$ and a computable measure μ, the term $\mu(V \cap W)$ is lower computable. Due to Proposition 14, the term $\mu(\bigcup\{B : \exists_{s,t} \text{ such that } B \subseteq B_{i_s} \text{ and } B \subseteq B_{k_t}\}) = \mu(V \cap W)$ is lower computable.

The integral of a finite supremum of step functions over U is lower computable by induction. For the base case $\int_U f_{i,j} d\mu = q_j \mu(B_i \cap U)$ is lower computable by the above reasoning. For the inductive step

$$\int_U \sup\{f_{i_1,j_1}, \ldots f_{i_k,j_k}\} d\mu = q_{j_m} \mu((B_{i_1} \cup \cdots \cup B_{i_k}) \cap U) + \int_U \sup\{f_{i_1,j'_1}, \ldots f_{i_k,j'_k}\} d\mu,$$

where q_{j_m} is minimal among $\{q_{j_1}, \ldots, q_{j_k}\}$ and $q_{j'_1} = q_{j_1} - q_{i_m}, \ldots, q_{j'_k} = q_{j_k} - q_{i_k}$. The first term on the right is lower-computable and by the induction assumption, the last term on the right is lower-computable. □

The following lemma is an update to the Stability Theorem 5 in [Gac94], using open sets instead of cells.

Lemma 29 *For computable measure space $(\mathcal{X}, \mu)$, $\mu\{\alpha \in \Pi_i : \mathbf{H}_\mu(\alpha) < \mathbf{H}_\mu(\Pi_i) - \mathbf{K}(\Pi) - m\} \stackrel{*}{<} 2^{-m}\mu(\Pi_i)$.*

Proof. Let $f(i) = \int_{\Pi_i} \mathbf{t}_\mu(\alpha) d\mu(\alpha)$. By Lemma 28, the function $f(i)$ is lower computable, and $\sum_i f(i) \leq 1$. Thus $f(i) \stackrel{*}{<} \mathbf{m}(i|\mu)/\mathbf{m}(\Pi)$. So

$$\mu(\Pi_i)^{-1} \int_{\Pi_i} 2^{-\mathbf{H}_\mu(\alpha)} d\mu(\alpha) \stackrel{*}{<} 2^{-\mathbf{H}_\mu(\Pi_i) + \mathbf{K}(\Pi)}.$$

By the Markov inequality,

$$\mu\{\alpha \in \Pi_i : \mathbf{H}_\mu(\alpha) < \mathbf{H}_\mu(\Pi_i) - \mathbf{K}(\Pi) - m\} \stackrel{*}{<} 2^{-m}\mu(\Pi_i).$$

□

Corollary 28 *For computable measure space* $(\mathcal{X}, \mu)$, $\mu\{\alpha : \mathbf{H}_\mu(\alpha) < \log \mu(\mathcal{X}) - m\} \stackrel{*}{<} 2^{-m}\mu(\mathcal{X})$.

Theorem 84 *For non-atomic computable mesure space* $(\mathcal{X}, \mu)$*, with uniformly computable* $\mu(\Pi_i)$, $\mu\{\alpha \in \Pi_i : \mathbf{H}_\mu(\alpha) < \mathbf{H}_\mu(\Pi_i) + \mathbf{K}(\Pi) - m\} \stackrel{*}{>} \mu(\Pi_i)2^{-m}$.

Proof. Similar to that of the proof of Proposition 31, one can compute, uniformly in m, an ad-set $D_m \subset \Pi_i$ such that $\mu(\Pi_i)2^{-m-2} < \mu(D_m) < \mu(\Pi_i)2^{-m-1}$ and $D_n \cap D_m = \emptyset$ for $n \neq m$. Suppose this has been done for D_{m-1}. The open set $\Pi_i \setminus \overline{\overline{D}}_{m-1}$ can be enumerated and $\mu(\Pi_i \setminus \overline{\overline{D}}_{m-1})$ can be lower computed. The enumeration of $\Pi_i \setminus \overline{\overline{D}}_{m-1}$ stops when $\mu(\Pi_i \setminus \overline{\overline{D}}_{m-1}) > \mu(\Pi_i)2^{-m-1}$. Using Proposition 30, one can find an ad-set (see Definition 31) $D_m \subset \mu(\Pi_i \setminus \overline{\overline{D}}_{m-1})$, such that $\mu(\Pi_i)2^{-m-2} < \mu(D_m) < \mu(\Pi_i)2^{-m-1}$. We define μ-test t to be $t(\alpha) = \sum_m [\alpha \in D_m]\mathbf{m}(m)2^m/\mu(\Pi_i)$. Since $\mathbf{m}(i)\mathbf{m}(\Pi)t \stackrel{*}{<} \mathbf{t}_\mu$, for universal lower computable test $\mathbf{t}_\mu$ and $\mathbf{H}_\mu = -\log \mathbf{t}_\mu$,

$$\mu(\Pi_i)2^{-m} < \mu\{\alpha : \mathbf{H}_\mu(\alpha) < \mathbf{K}(i) + \mathbf{K}(\Pi) + \log \mu(\Pi_i) - m\}$$
$$\mu(\Pi_i)2^{-m} < \mu\{\alpha : \mathbf{H}_\mu(\alpha) < \mathbf{H}_\mu(\Pi_i) + \mathbf{K}(\Pi) - m\}.$$

□

The following Theorem is the coarse grained entropy to the oscillation Theorem 75. As a state travels through different cells Π_i and Π_j, the coarse grained entropy will oscillate, in that its max value $\mathbf{H}_\mu(\Pi_{i_{\max}})$ will become increasingly larger than its min value, $\mathbf{H}_\mu(\Pi_{i_{\min}})$. If the dynamics are ergodic, then by Theorem 77, the state is guaranteed to hit every cell if there are a finite number of them, and if there are an infinite number of cells, the state is guaranteed to hit an unbounded number of cells.

Theorem 85 *Let* $(\mathcal{X}, \mu)$ *be a computable measure space,* G^t *be a tranformation group, and* $\{\Pi_i\}$ *a partition of* $\mathcal{X}$*. If* $i \mapsto \mu(\Pi_i)$ *is uniformly computable and if a state* $\alpha \in \mathcal{X}$*, travels through at least* 2^n *partitions* $\{\Pi_i\}_{i=1}^{2^n}$ *over* $t \in [0, 1]$*, then, relativized to* μ*,*

$$\min_{i \in \{1,\ldots,2^n\}} \mathbf{H}_\mu(\Pi_i) <^{\log} \max_{i \in \{1,\ldots,2^n\}} \mathbf{H}_\mu(\Pi_i) - n + \mathbf{I}(\alpha : \mathcal{H}).$$

Proof. Let $f(i) = \lceil \log \mu(\Pi_i) \rceil$. Let $D \subset \mathbb{N}$, $|D| = 2^n$ be a set of partitions that α travels through in time $t \in [0, 1]$, so $\mathbf{K}(D|\alpha) <^+ \mathbf{K}(n)$. Theorem 114, on $f : D \to \mathbb{N}$, produces $x \in D$, where

$$f(x) + \mathbf{K}(x) <^{\log} -\log \sum_{a \in D} \mathbf{m}(a)2^{-f(a)} + \mathbf{I}((f, D); \mathcal{H})$$
$$f(x) + \mathbf{K}(x) + n <^{\log} \max_{a \in D} f(a) + \mathbf{K}(a) + \mathbf{I}((f, D); \mathcal{H})$$
$$\mathbf{H}_\mu(x) + n <^{\log} \max_{a \in D} \mathbf{H}_\mu(a) + \mathbf{I}(\alpha : \mathcal{H}).$$

□

18.2 Entropy Variance in Cells

In this section, we will prove a lower bound of states with low algorithmic fine grained entropy with respect to the uniform distribution over a particular cell. Thus, regardless of what measures are used to define entropy, low entropy states are bound to occur in cells. A result is first proved about infinite sequences, and measures over the Cantor Space, relying on Lemma 15.2.

Theorem 86 *For computable probability measures μ, ρ and nonatomic λ over $\{0,1\}^\infty$ and $n \in \mathbb{N}$, $\lambda\{\alpha : \mathbf{D}(\alpha|\mu) > n \text{ and } \mathbf{D}(\alpha|\rho) > n\} > 2^{-n-\mathbf{K}(n,\mu,\rho,\lambda)-O(1)}$.*

Proof. We first assume not. For all $c \in \mathbb{N}$, there exist computable nonatomic measures μ, ρ λ, and there exists n, where $\lambda\{\alpha : \mathbf{D}(\alpha|\mu) > n \text{ and } \mathbf{D}(\alpha|\rho) > n\} \leq 2^{-n-\mathbf{K}(n,\mu,\lambda)-c}$. Sample $2^{n+\mathbf{K}(n,\mu,\rho,\lambda)+c-1}$ elements $D \subset \{0,1\}^\infty$ according to λ. The probability that all samples $\beta \in D$ has $\mathbf{D}(\beta|\mu) \leq n$ or $\mathbf{D}(\beta|\rho) \leq n$ is

$$\prod_{\beta \in D} \lambda\{\mathbf{D}(\beta|\mu) \leq n \text{ or } \mathbf{D}(\beta|\rho) \leq n\} \geq$$
$$(1 - |D|2^{-n-\mathbf{K}(n,\mu,\lambda,\rho)-c}) \geq$$
$$(1 - 2^{n+\mathbf{K}(n,\mu,\lambda,\rho)+c-1}2^{-n-\mathbf{K}(n,\mu,\lambda,\rho)-c}) \geq 1/2.$$

Let $\lambda^{n,c}$ be the probability of an encoding of $2^{n+\mathbf{K}(n,\mu,\lambda,\rho)+c-1}$ elements each distributed according to λ. Thus

$$\lambda^{n,c}(\text{Encoding of } 2^{n+\mathbf{K}(n,\mu,\lambda,\rho)+c-1} \text{ elements } \beta, \text{ each having } \mathbf{D}(\beta|\mu) \leq n \text{ or } \mathbf{D}(\beta|\rho) \leq n) \geq 1/2.$$

Let v be a shortest program to compute $\langle n, \mu, \rho, \lambda\rangle$. By Theorem 74, with the universal Turing machine relativized to v,

$$\lambda^{n,c}(\{\gamma : \mathbf{I}(\gamma : \mathcal{H}|v) > m\}) \stackrel{*}{<} 2^{-m+\mathbf{K}(\lambda^{n,c}|v)} \stackrel{*}{<} 2^{-m+\mathbf{K}(n,\mathbf{K}(n,\mu,\lambda,\rho),c,\lambda|v)} \stackrel{*}{<} 2^{-m+\mathbf{K}(c)}.$$

Therefore,

$$\lambda^{n,c}(\{\gamma : \mathbf{I}(\gamma : \mathcal{H}|v) > \mathbf{K}(c) + O(1)\}) \leq 1/4.$$

Thus, by probabilistic arguments, there exists $\alpha \in \{0,1\}^\infty$, such that $\alpha = \langle D\rangle$ is an encoding of $2^{n+\mathbf{K}(n,\mu,\rho,\lambda)+c-1}$ elements $\beta \in D \subset \{0,1\}^\infty$, where each β has $\mathbf{D}(\beta|\mu) \leq n$ or $\mathbf{D}(\beta|\rho) \leq n$ and $\mathbf{I}(\alpha : \mathcal{H}|v) <^+ \mathbf{K}(c)$. By Lemma 23, relativized to v, there are constants $d, f, g \in \mathbb{N}$ where

$$\begin{aligned} m = \log|D| &< \max_{\beta \in D} \min\{\mathbf{D}(\beta|\mu, v), \mathbf{D}(\beta|\rho, v)\} + 2\mathbf{I}(D : \mathcal{H}|v) + d\mathbf{K}(m|v) + f\mathbf{K}(\mu|v) + g\mathbf{K}(\rho|v) \\ m &< \max_{\beta \in D} \min\{\mathbf{D}(\beta|\mu), \mathbf{D}(\beta|\rho)\} + \mathbf{K}(v) + 2\mathbf{I}(D : \mathcal{H}|v) + d\mathbf{K}(m|v) + f\mathbf{K}(\mu|v) + g\mathbf{K}(\rho|v) \\ &<^+ n + \mathbf{K}(n,\mu,\lambda,\rho) + d\mathbf{K}(m|v) + 2\mathbf{K}(c) + (f+g)O(1). \end{aligned} \tag{18.1}$$

Therefore:

$$m = n + \mathbf{K}(n,\mu,\rho,\lambda) + c - 1$$
$$\mathbf{K}(m|v) <^+ \mathbf{K}(c).$$

Plugging the inequality for $\mathbf{K}(m|v)$ back into Equation 18.1 results in

$$n + \mathbf{K}(n, \mu, \lambda, \rho) + c <^+ n + \mathbf{K}(n, \mu, \lambda, \rho) + 2\mathbf{K}(c) + d\mathbf{K}(c) + (f + g)O(1)$$
$$c <^+ (2 + d)\mathbf{K}(c) + (f + g)O(1).$$

This result is a contradiction for sufficiently large c solely dependent on the universal Turing machine. □

Theorem 87 *Given computable measures μ, ρ, and computable non-atomic probability measure λ, over a computable metric space $\mathcal{X}$, there is a $c \in \mathbb{N}$, where for all n,* $\lambda(\{\alpha : \mathbf{H}_\mu(\alpha) < \log \mu(X) - n \text{ and } \mathbf{H}_\rho(\alpha) < \log \rho(X) - n\}) > 2^{-n-\mathbf{K}(n)-c}$.

Proof. We fix the algorithmic descriptions of λ, μ, ρ, and $\mathcal{X}$. By Theorem 70, easily generalized to three measures, fix a multi representation $(\delta, \lambda_\delta, \mu_\delta, \rho_\delta)$ for computable multi measure space $(\mathcal{X}, \lambda, \mu, \rho)$. Note that δ is a measure-preserving transform, where $\lambda(A) = \lambda_\delta(\delta^{-1}(A))$ for all Borel sets A. Due to Lemma 19, there is a $c \in \mathbb{N}$,

$$\begin{aligned}
\eta =& \lambda(\{\beta : \mathbf{H}_\mu(\beta) > \log \mu(X) - n \text{ or } \mathbf{H}_\rho(\beta) \geq \log \rho(X) - n\}) \\
=& \lambda(\{\beta : \mathbf{t}_\mu(\beta) \leq 2^n/\mu(X) \text{ or } \mathbf{t}_\rho(\beta) \leq 2^n/\rho(X)\}) \qquad (18.2) \\
=& \lambda_\delta(\delta^{-1}(\{\beta : \mathbf{t}_\mu(\beta) \leq 2^n/\mu(X) \text{ or } \mathbf{t}_\rho(\beta) \leq 2^n/\rho(X)\})) \\
<& \lambda_\delta(\delta^{-1}(\{\beta : \mathbf{t}_{\mu_\delta}(\delta^{-1}(\beta)) < 2^{n+c}/\mu(X) \text{ or } \mathbf{t}_{\mu_\rho}(\delta^{-1}(\beta)) < 2^{n+c}/\rho(X)\})) \\
=& \lambda_\delta(\delta^{-1}(\{\beta \in \delta(\{\alpha : \mathbf{t}_{\mu_\delta}(\alpha) < 2^{n+c}/\mu(X) \text{ or } \mathbf{t}_{\rho_\delta}(\alpha) < 2^{n+c}/\rho(X)\})\})) \\
=& \lambda_\delta(\delta^{-1}(\delta(\{\alpha : \mathbf{t}_{\mu_\delta}(\alpha) < 2^{n+c}/\mu(X) \text{ or } \mathbf{t}_{\rho_\delta}(\alpha) < 2^{n+c}/\rho(X)\})) \\
<& \lambda_\delta(\{\alpha : \mathbf{t}_{\mu_\delta}(\alpha) < 2^{n+c}/\mu(X) \text{ or } \mathbf{t}_{\rho_\delta}(\alpha) < 2^{n+c}/\rho(X)\}). \qquad (18.3)
\end{aligned}$$

Let $\overline{\mu}_\delta = \mu_\delta/\mu(X)$ and $\overline{\rho}_\delta = \rho_\delta/\rho(X)$ be two computable probability measures over $\{0,1\}^\infty$. From Equation 18.3, and Proposition 18, we get,

$$\begin{aligned}
\eta <& \lambda_\delta\{\alpha : \mathbf{D}(\alpha|\mu_\delta) <^+ n + c - \log \mu(X) \text{ or } \mathbf{D}(\alpha|\rho_\delta) <^+ n + c - \log \rho(X)\} \\
=& \lambda_\delta\{\alpha : \mathbf{D}(\alpha|\overline{\mu}_\delta) <^+ n + c \text{ or } \mathbf{D}(\alpha|\overline{\rho}_\delta) <^+ n + c\}. \qquad (18.4)
\end{aligned}$$

From Theorem 86, we get

$$\begin{aligned}
\eta <& \lambda(\{\beta : \mathbf{H}_\mu(\beta) > \log \mu(X) - n \text{ or } \mathbf{H}_\rho(\beta) \geq \log \rho(X) - n\}) \\
<& \lambda_\delta\{\alpha : \mathbf{D}(\alpha|\overline{\mu}_\delta) <^+ n + c \text{ or } \mathbf{D}(\alpha|\overline{\rho}_\delta) <^+ n + c\} \\
<& 1 - 2^{-n-c}.
\end{aligned}$$

□

Corollary 29 *Given computable measure μ and computable non-atomic probability measure λ, over a computable metric space $\mathcal{X}$, there is a $c \in \mathbb{N}$, where for all n, $\lambda(\{\alpha : \mathbf{t}_\mu(\alpha) > 2^n/\mu(X)\}) > 2^{-n-\mathbf{K}(n)-c}$.*

The following corollary states, given a uniform measure over a cell, there is a lower bound over the measure of the states with low algorithmic entropy with respect to two measures. Note that the corollary can be easily generalized to any finite number of measures.

Corollary 30 *For dual computable measure space $(\mathcal{X}, \mu, \nu)$, and with partition Π, let λ be a computable uniform probability measure over cell Π_i. There is a constant $c \in \mathbb{N}$ where $\lambda(\{\alpha : \mathbf{H}_\mu(\alpha) < \log \mu(X) - n \text{ and } \mathbf{H}_\nu(\alpha) < \log \nu(X) - n\}) > 2^{-n-\mathbf{K}(n)-c}$.*

Proof. This follows directly from Theorem 87. □

Corollary 31 *Let $(\mathcal{X}, \nu, \mu)$ be a dual measure space with non-atomic probability ν. There is a $c \in \mathbb{N}$ with the following properties. Let $T : X \to X$ be a computable ν-ergodic function, and $U_n = \{x : \mathbf{H}_\mu(x) < \log \mu(X) - n\}$. If $\omega \in X$ has $\mathbf{H}_\mu(\omega) \neq -\infty$,*

$$2^{-n-\mathbf{K}(n)-c} < \lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} \mathbf{1}_{U_n}(T^t(\omega)).$$

Chapter 19

Computability of Algorithmic Entropy

In this chapter, computability properties of algorithmic entropy are studied. It is easy to see that algorithmic fine grained entropy is uncomputable. In Section 19.1, it is shown that there is no effective method for chooosing times where the entropy of a typical state is continually decreasing. In Section 19.2, it is shown if a function can approximate algorithmic fine grained entropy for an infinite amount of typical states, then it has infinite mutual information with the halting sequence.

19.1 Uncomputability of Decreasing Entropy

If a physicist were given the state of a system, (say particles in a box), she could compute times of decreasing entropy (just wait until the particles are in smaller and smaller spaces). However without access to the state, there are no computable method to predict the times in which the state continually decreases.

Theorem 88 *Let $(\mathcal{X}, \mu)$ be a computable measure space and $T : X \to X$ be a computable ergodic function. If $\alpha \in X$, and $\mathbf{H}_\mu(\alpha) > -\infty$, then there does not exist a computable function $f : \mathbb{N} \to \mathbb{N}$, such that $\mathbf{H}_\mu\left(T^{f(n)}\alpha\right) < -n$.*

Proof. Assume not, and such a computable function f exists. Let $U_n = \{\beta : \mathbf{H}_\mu(\beta) < -n\}$, which is an effectively open set, uniformly in n. One can construct a μ-test, where $t(\beta) = \sup_n[\beta \in T^{-f(n)}U_n]\mathbf{m}(n)2^n$. Thus $\infty = t(\alpha) \stackrel{*}{<} \mathbf{t}_\mu(\alpha)$, and thus $\mathbf{H}_\mu(\alpha) = -\infty$, causing a contradiction. □

Corollary 32 *Let $(\mathcal{X}, \mu)$ be a computable measure space and G^t and transformation group. If $\alpha \in X$, and $\mathbf{H}_\mu(\alpha) > -\infty$, then there does not exist a computable function $f : \mathbb{N} \to \mathbb{R}_{>0}$, such that $\mathbf{H}_\mu(G^{f(n)}\alpha) < -n$.*

Proof. The proof follows analogously to that of Theorem 88. □

19.2 Non-Approximability of Algorithmic Entropy

19.2.1 Kolmogorov Complexity is Exotic

We review the material on busy beaver functions, detailed in 15.2. Let $\Omega = \sum\{2^{-\|p\|} : U(p) \text{ halts}\}$ be Chaitin's Omega, $\Omega_n \in \mathbb{Q}_{\geq 0}$ be be the rational formed from the first n bits of Ω, and $\Omega^t = \sum\{2^{-\|p\|} : U(p) \text{ halts in time } t\}$. For $n \in \mathbb{N}$, let $\mathbf{bb}(n) = \min\{t : \Omega_n < \Omega^t\}$. $\mathbf{bb}^{-1}(m) = \arg\min_n\{\mathbf{bb}(n-1) < m \leq \mathbf{bb}(n)\}$. Let $\Omega[n] \in \{0,1\}^*$ be the first n bits of Ω. For $t \in \mathbb{N}$

define the function $\mathbf{m}^t(x) = \sum\{2^{-\|p\|} : U(p) = x \text{ in } t \text{ steps}\}$ and for $n \in \mathbb{N}$, we have $\mathbf{m}_n(x) = \sum\{2^{-\|p\|} : U(p) = x \text{ in } \mathbf{bb}(n) \text{ steps}\}$.

The following lemma can be strengthened to just $\mathbf{I}(X;\mathcal{H})$, but with an increase to the complexity of the proof.

Lemma 30 *A relation $X = \{(x_i, c_i)\}_{i=1}^{2^n} \subset \{0,1\}^* \times \mathbb{N}$, $|\mathbf{K}(x_i) - c_i| \leq s$, has $n <^{\log} 2s + 2\mathbf{I}(X;\mathcal{H})$.*

Proof. We relativize the universal Turing machine to (n, s), which can be done due to the precision of the theorem. Let $T = \min\{t : |\lceil -\log \mathbf{m}_t(x_i)\rceil - c_i| < s+1\}$. Let $N = \mathbf{bb}^{-1}(T)$ and $M = \mathbf{bb}(N)$. So for all x_i, $-\log \mathbf{m}_M(x_i) - \mathbf{K}(x_i) <^+ 2s$. Let Q be an elementary probability measure that realizes $\mathbf{Ks}(X)$ and $d = \max\{\mathbf{d}(X|Q), 1\}$. Without loss of generality, the support of Q is restricted to binary relations $B \subset \{0,1\}^* \times \mathbb{N}$ of size 2^n. Let $B_1 = \bigcup\{y : (y, c) \in B\}$. Let $S = \bigcup\{B_1 : B \in \text{Support}(Q)\}$. We randomly select each string in S to be in a set R independently with probability $d2^{-n}$. Thus $\mathbf{E}[\mathbf{m}_M(R)] \leq d2^{-n}$. For $B \in \text{Support}(Q)$,

$$\mathbf{E}_R \mathbf{E}_{B\sim Q}[[R \cap B_1 = \emptyset]] = \mathbf{E}_{B\sim Q} \Pr(R \cap B_1 = \emptyset) = (1 - d2^{-n})^{2^n} < e^{-d}.$$

Thus there exists a set $R \subseteq S$ such that $\mathbf{m}_M(R) \leq 2{\cdot}2^{-n}$ and $\mathbf{E}_{B\sim Q}[[R \cap B_1 = \emptyset]] < 2e^{-d}$. Let $t(B) = .5[R \cap B_1 = \emptyset]2^d$. t is a Q-test, with $\mathbf{E}_{B\sim Q}[t(B)] \leq 1$. It must be that $t(X) \neq 0$, otherwise,

$$1.44d - 1 < \log t(X) <^+ \mathbf{d}(X|Q, d) <^+ d + \mathbf{K}(d),$$

which is a contradiction for large enough d, which one can assume without loss of generality. Thus $t(X) \neq 0$ and $R \cap X_1 \neq \emptyset$. Furthermore, if $y \in R$, $\mathbf{K}(y) <^+ -\log \mathbf{m}_M(x) - n + \log d + \mathbf{K}(d, M, R)$. So for $x \in R \cap X_1$, .

$$\begin{aligned}
\mathbf{K}(x) &<^+ -\log \mathbf{m}_M(x) - n + \log d + \mathbf{K}(d, M, R) \\
\mathbf{K}(x) &<^+ \mathbf{K}(x) + 2s - n + \log d + \mathbf{K}(M) + \mathbf{K}(R, d) \\
n &<^+ 2s + \mathbf{K}(M) + \log d + \mathbf{K}(Q, d) \\
n &<^+ 2s + \mathbf{K}(\Omega[N]) + \mathbf{Ks}(X) \\
n &<^+ 2s + \mathbf{K}(\Omega[N]) + \mathbf{I}(X;\mathcal{H}) \qquad (19.1)
\end{aligned}$$

From Lemma 22, $\mathbf{K}(\Omega[N]|T, N) =^+ \mathbf{K}(\Omega[N]|X, N) = O(1)$. Furthermore it is well known for the bits of Chaitin's Omega, $N <^+ \mathbf{K}(\Omega[N])$ and $\mathbf{K}(\Omega[N]|\mathcal{H}) <^+ \mathbf{K}(N)$. So, using Lemma 1,

$$N <^+ \mathbf{K}(\Omega[N]) <^{\log} \mathbf{I}(\Omega[N];\mathcal{H}) <^{\log} \mathbf{I}(X;\mathcal{H}) + \mathbf{K}(N) <^{\log} \mathbf{K}(X;\mathcal{H}). \qquad (19.2)$$

So combining Equations 19.1 and 19.2, one gets

$$n <^{\log} 2s + 2\mathbf{I}(X;\mathcal{H}).$$

□

Exercise 27 (Hard) *Show that if relation $X = \{(x_i, c_i)\}_{i=1}^{2^n} \subset \{0,1\}^* \times \mathbb{N}$ has $0 \leq c_i - \mathbf{K}(x_i) \leq s$ then $n <^{\log} s + \mathbf{I}(X;\mathcal{H})$.*

Lemma 31 *Given a computable probability measure μ, over $\{0,1\}^\infty$, for an infinite set of unique infinite sequences $\{\alpha_i\}_{i=1}^\infty$ where $\mathbf{D}(\alpha_i|\mu) < \infty$ for all i, and an infinite set of numbers $\{c_i\}_{i=1}^\infty$, if $\sup_i |\mathbf{D}(\alpha_i|\mu) - c_i| < \infty$, then $\mathbf{I}(\{(\alpha_i, c_i\} : \mathcal{H}) = \infty$.*

Proof. Suppose $c = \max_i \mathbf{D}(\alpha_i|\mu) < \infty$. Then let $D_n = \{\alpha_i\}_{i=1}^{2^n}$. So $\mathbf{K}(D_n|\{\alpha_i\}) <^+ \mathbf{K}(n)$. By Lemma 23, $n <^{\log} c + \mathbf{I}(D_n : \mathcal{H}) <^{\log} \mathbf{I}(\{\alpha_i\} : \mathcal{H}) + \mathbf{K}(n)$. Since this holds for all n,

$$\infty = \mathbf{I}(\{\alpha_i\} : \mathcal{H}) = \mathbf{I}(\{(\alpha_i, c_i)\} : \mathcal{H}).$$

Suppose $\sup_i \mathbf{D}(\alpha_i|\mu) = \infty$ and $r = \max_i |\mathbf{D}(\alpha_i|\mu) - c_i| < \infty$. Given n and $\{(\alpha_i, c_i)\}$, one can find 2^n unique infinite sequences α_i, such that $|c_i - c_j| > 5s$ for $i \neq j$, $i, j \in \{1, \ldots, 2^n\}$, and using computable μ, due to the definition of $\mathbf{D}$, one can compute prefixes $x_i \sqsubset \alpha_i$ where $|-\log \mu(x_i) - \mathbf{K}(x_i) - c_i| <^+ 2r$. Setting $X_n = \{(x_i, \lceil -\log \mu(x_i) - c_i \rceil)\}$, and $s = 2r + 1$, invoking Lemma 30 gives

$$n <^{\log} 4r + 2\mathbf{I}(X_n; \mathcal{H}) <^{\log} 4r + \mathbf{I}(\{(\alpha_i, c_i)\} : \mathcal{H}) + \mathbf{K}(n),$$

Thus since this hold for arbitrary n, $\mathbf{I}(\{(\alpha_i, c_i)\} : \mathcal{H}) = \infty$. □

19.2.2 Algorithmic Fine Grain Entropy Cannot be Approximated

It is easy to see that algorithmic fine grained entropy is uncomputable. In this section. we go one step further and show that information about algorithmic fine grained entropy coincides with information about the halting sequence.

Definition 46 (Mutual Information of Infinite Set of Points with $\mathcal{H}$) *Given an infinite set of infinite sequences $\{\alpha_i\}_{i=1}^{\infty}$, we encode them into a single infinite sequence $\langle\{\alpha_i\}\rangle$ in the standard way, in that α_1 is encoded into every other bit, α_2 is encoded into every other free space, and so on. Given computable metric space $\mathcal{X}$, the set of encodings for an infinite set of points in X, $\{\alpha_i\}$, $\alpha_i \in X$, $[\{\alpha_i\}]$, is the set of all encoded $\langle\{\beta_i\}\rangle$, where β_i is an encoding of a fast Cauchy sequence for α_i. The mutual information that an infinite set $\{\alpha_i\}$ of points has with the halting sequence is $\mathbf{I}(\{\alpha_i\} : \mathcal{H}) = \inf_{\beta \in [\{\alpha_i\}]} \mathbf{I}(\beta : \mathcal{H})$. In a standard way, a number can be appended to the start of each fast Cauchy sequence.*

Theorem 89 *Let $(\mathcal{X}, \mu)$ be a computable measure space and $\{\alpha_i\}_{i=1}^{\infty}$ be an infinite set of unique points in $\mathcal{X}$ where $\mathbf{H}_\mu(\alpha_i) < \infty$, for each α_i. For the infinite set of numbers, $\{c_i\}_{i=1}^{\infty}$, if $\sup_i |-\mathbf{H}_\mu(\alpha_i) - c_i| < \infty$, then $\mathbf{I}(\{(\alpha_i, c_i)\} : \mathcal{H}) = \infty$.*

Proof. Let $\overline{\mu} = \mu/\mu(X)$ be a computable probability measure due to 1. Let $(\delta, \overline{\mu}_\delta)$ be a binary representation for $(\mathcal{X}, \overline{\mu})$. Since (δ, δ^{-1}) is an isomorphism, by Corollary 22 and Proposition 18, and since $\mathbf{H}_\mu(\alpha_i) < \infty$, there is a constant c, where, for all $\alpha \in \{\alpha_i\}$,

$$\begin{aligned} c >& |-\log \mathbf{t}_{\overline{\mu}_\delta}(\delta^{-1}(\alpha_i)|\overline{\mu}_\delta) - \mathbf{H}_{\overline{\mu}}(\alpha_i)| \\ c >& |-\mathbf{D}(\delta^{-1}(\alpha_i)|\overline{\mu}_\delta) - \mathbf{H}_{\overline{\mu}}(\alpha_i)|. \end{aligned}$$

Let β minimize $\lceil \mathbf{I}(\{(\alpha_i, c_i)\} : \mathcal{H}) \rceil$. So by applying δ^{-1} to each of the fast Cauchy sequences of α_i encoded in β, with $\gamma_i = \delta^{-1}(\alpha_i)$, one can construct an infinite set $\{(\gamma_i, c_i)\}_{i=1}^{\infty}$ of infinite sequences where $\sup_i |\mathbf{D}(\gamma_i|\overline{\mu}_\delta) - c_i| < \infty$. So, using Lemma 31,

$$\infty = \mathbf{I}(\{(c_i, \gamma_i)\}_{i=1}^{\infty} : \mathcal{H}) = \mathbf{I}(\beta : \mathcal{H}) = \mathbf{I}(\{(\alpha_i, c_i)\} : \mathcal{H}).$$

□

19.3 Computability of Measures

In this section, we provide sufficient conditions for a measure to be uncomputable.

Theorem 90 *Let $\mathcal{X}$ be a computable metric space and μ be a non-atomic measure over $\mathcal{X}$. If $\mu(\{\alpha : \mathbf{H}_\mu(\alpha) < -n\}) < 2^{-n-\mathbf{K}(n)-\omega(1)}$ then μ is uncomputable.*

Proof. Assume μ is computable. Let $\overline{\mu} = \mu/\mu(X)$ be a computable probability measure due to Claim 1. Let $(\delta, \overline{\mu}_\delta)$ be a binary representation for $(\mathcal{X}, \overline{\mu})$. Thus $\overline{\mu}_\delta$ is computable. Fix n. Let $U_n = \{\alpha \in X : \mathbf{H}_{\overline{\mu}}(\alpha) < -n\}$ and $\delta^{-1}(U_n) = V_n$. Then there is some $c \in \mathbb{R}_{\geq 0}$, independent of n, such that $V_n \supseteq \{\alpha \in \{0,1\}^\infty : \mathbf{D}(\alpha|\overline{\mu}_\delta) > n - c\}$. By Theorem 86, there is a d, independent of n, where $\overline{\mu}_\delta(V_n) > 2^{-n-\mathbf{K}(n)-d}$. Since δ is a measure-preserving morphism, $\overline{\mu}(U_n) > 2^{-n-\mathbf{K}(n)-d}$ for all n, causing a contradiction.

Definition 47 (Mutual Information of Measures with $\mathcal{H}$) *Let $\mathcal{X}$ be a computable metric space $\mu \in \mathfrak{M}(X)$ be a measure over this space. Its mutual information with the halting sequence is $\mathbf{I}(\mu : \mathcal{H})$, where μ is treated as a point in the computable metric space $\mathfrak{M}(X)$, and Definition 42 is used.*

Definition 48 (Neutral Measures) *Let $\mathcal{X}$ be a computable measure space. A measure μ is weakly neutral if $\forall_{\alpha \in X} \mathbf{H}_\mu(\alpha) > -\infty$. A measure μ is neutral if $\inf_{\alpha \in X} \mathbf{H}_\mu(\alpha) > -\infty$,*

Lemma 32 (Sperner's Lemma) *Let $p_1, \ldots, p_k$ be points of some finite-dimensional space $\mathbb{R}^n$. Suppose that there are closed sets $F_1, \ldots, F_k$ with the property that for every subset $1 \leq i_k < \cdots < i_j < k$, the simplex $S(p_{i_1}, \ldots, p_{i_j})$ spanned by $p_{i_1}, \ldots, p_{i_j}$ is covered by the union $F_{i_1} \cup \cdots \cup F_{i_j}$. Then the intersection $\cap_i F_i$ is not empty.*

Proposition 35 *Let $\mathcal{X}$ be a computable measure space. For every closed set $A \subset X$ and probability measure μ, if $\mu(A) = 1$ then there exists $a \in X$, $\mathbf{t}_\mu(x) \leq 1$.*

Proof. $\int_X \mathbf{t}_\mu d\mu = \mu^x \mathbf{1}_A(x) \mathbf{t}_\mu(x) \leq 1$. □

Part (1) of the following theorem is due to [Lev76], and conveyed in [Gác21].

Theorem 91 *Let $\mathcal{X}$ be a computable metric space.*

1. *If $\mathcal{X}$ is compact then it has a neutral measure.*
2. *If μ is weakly neutral then $\mathbf{I}(\mu : \mathcal{H}) = \infty$.*

Proof. (1) For every $x \in X$, let F_x be the set of measures for which $\mathbf{t}_\mu(x) \leq 1$. Since $\mathcal{X}$ is compact, the space of Borel probability measures $\mathcal{M}(X)$ over $\mathcal{X}$ is compact. Therefore, due to compactness, if every finite subset of $\{F_x : x \in X\}$ of closed sets has a nonempty intersection, then $\cap_x F_x \neq \emptyset$. Let $S(x_1, \ldots, x_k)$ be the simplex of probability measures concentrated on $x_1, \ldots, x_k$. Proposition 35 implies each such measure belongs to one of the sets F_{x_i}.Thus $S(x_1, \ldots, x_k) \subset F_{x_1} \cup \cdots \cup F_{x_k}$ and this is true for any subset of the indices $\{1, \ldots, k\}$. Lemma 32 implies $F_{x_1} \cap \cdots \cap F_{x_k} \neq \emptyset$.

(2) Let $\overline{\mu} = \mu/\mu(X)$, where given a fast Cauchy sequence for μ, one can easily compute a fast Cauchy sequence for $\overline{\mu}$ (just normalize the ideal points). Let $(\delta, \overline{\mu}_\delta)$ be a binary representation of $(\mathcal{X}, \overline{\mu})$. Due to Lemma 19, $\mathbf{t}_{\overline{\mu}_\delta}(\alpha) < \infty$, for all $\alpha \in \{0,1\}^\infty$, so $\overline{\mu}_\delta$ is a weakly neutral measure. Due

to [DJ13], some PA degree is computable from any encoded Cauchy sequence $\langle\overrightarrow{\overline{\mu}_\delta}\rangle \in \{0,1\}^\infty$ of $\overline{\mu}_\delta$ in $\mathcal{M}(\{0,1\}^\infty)$ space. Due to [Lev13], $\mathbf{I}(\langle\overrightarrow{\overline{\mu}_\delta}\rangle : \mathcal{H}) = \infty$. So by

$$\infty = \inf_{\overrightarrow{\overline{\mu}_\delta}} \mathbf{I}(\langle\overrightarrow{\overline{\mu}_\delta}\rangle : \mathcal{H}) = \mathbf{I}(\overline{\mu}_\delta : \mathcal{H}) <^+ \mathbf{I}(\overline{\mu} : \mathcal{H}) <^+ \mathbf{I}(\mu : \mathcal{H}).$$

□

19.4 Exotic Dynamics

Definition 49 (Mutual Information of Transformation Group with $\mathcal{H}$) *Let $\mathcal{X}$ be a computable metric space. Let $\alpha \in X$ be a point in this space and G^t a (potentially uncomputable) transformation group over $\mathcal{X}$. For $\beta \in \{0,1\}^\infty$, let $\mathrm{Num}(\beta) \in [0,1]$ convert the infinite sequence β into a real in $[0,1]$ in the standard way. Let $[(G^t,\alpha)]$ consist of all infinite sequences γ such that $U_\beta(\gamma)$ outputs an encoding of a fast Cauchy sequence to $G^{\mathrm{Num}(\beta)}\alpha$. The amount of information that (G^t,α) has with the halting sequence is $\inf_{\gamma\in[(G^t,\alpha)]} \mathbf{I}(\gamma{:}\mathcal{H})$.*

The following theorem states that if entropy does not oscillate enough during the course of dynamics, then the dynamics and starting point are exotic, containing infinite mutual information with the halting sequence. This theorem is a generalization of Theorem 75 in Chapter 15, though the bounds are looser.

Theorem 92 *Let L be the Lebesgue measure. Let $(\mathcal{X},\mu)$ be a computable measure space. There is a $c \in \mathbb{N}$ such that for (potentially uncomputable) transformation group G^t and point $\alpha \in X$ if $L\{t : \mathbf{H}_\mu(G^t\alpha) < \log\mu(X) - n\} < 2^{-n-c\mathbf{K}(n)}$, then $\mathbf{I}((G^t,\alpha) : \mathcal{H}) = \infty$.*

Proof. Fix n. Let $(\{0,1\}^\infty,\Gamma)$ be the Cantor space with the uniform measure. The binary representation (see Theorem 70) creates an isomorphism (ϕ,ϕ^{-1}) of computable probability spaces between the spaces $(\{0,1\}^\infty,\Gamma)$ and $([0,1],L)$. It is the canonical function $\phi(\gamma) = 0.\gamma$. Thus for all Borel sets $A \subseteq [0,1]$, $\Gamma(\phi^{-1}(A)) = L(A)$. So

$$\begin{aligned}
&L\{t : \mathbf{H}_\mu(G^t\alpha) < \log\mu(X) - n\}\\
=&\Gamma\{\beta : \mathbf{H}_\mu(G^{\phi(\beta)}\alpha) < \log\mu(X) - n\}\\
=&\Gamma\{\beta : \log\mathbf{t}_\mu(G^{\phi(\beta)}\alpha) > 2^n/\mu(X)\}\\
<&2^{-n-c\mathbf{K}(n)}.
\end{aligned}$$

Let (δ,μ_δ) be a binary representation for $(\mathcal{X},\mu)$. Let $\overline{\mu}_\delta = \mu_\delta/\mu_\delta(\{0,1\}^\infty)$ be a computable probability measure over $\{0,1\}^\infty$. Thus, due to Lemma 19 and Proposition 18, there is a $d \in \mathbb{N}$ with $\psi(\beta) = \delta^{-1}(G^\beta\alpha)$ and

$$\begin{aligned}
2^{-n-c\mathbf{K}(n)} >&\Gamma\{\beta : \log\mathbf{t}_\mu(G^{\phi(\beta)}\alpha) > 2^n/\mu(X)\}\\
>&\Gamma\{\beta : \log\mathbf{t}_{\mu_\delta}(\psi(\beta)) > 2^{n-d}/\mu(X)\}\\
>&\Gamma\{\beta : \mathbf{D}(\psi(\beta)|\mu_\delta) > n - \log\mu(X) + d\}\\
>&\Gamma\{\beta : \mathbf{D}(\psi(\beta)|\overline{\mu}_\delta) > n + d\}.
\end{aligned}$$

Let $W \in \{0,1\}^{\infty 2^{n+c\mathbf{K}(n)-1}}$ be a set of $2^{n+c\mathbf{K}(n)-1}$ infinite sequences with each sequence chosen independently the uniform distribution over $\{0,1\}^\infty$. The probability that all $\beta \in W$ has $\mathbf{D}(\psi(\beta)|\overline{\mu}_\delta) \leq n + d$ is

$$(1 - 2^{-n-c\mathbf{K}(n)})^{2^{n+c\mathbf{K}(n)-1}} \geq (1 - 2^{n+c\mathbf{K}(n)-1}2^{-n-c\mathbf{K}(n)}) \geq 1/2.$$

Let $\mathcal{U}$ be a distribution over $\{0,1\}^\infty$ that is the uniform measure applied independently to $2^{n+c\mathbf{K}(n)-1}$ encoded sequences. Let $\gamma \in [(G^t,\alpha)]$ that minimizes $\lceil \mathbf{I}((G^t,\alpha):\mathcal{H})\rceil + 1$ and $\lambda = \langle p\rangle\gamma$, where p is a program to compute $\mathcal{U}$. By Theorem 74,

$$\Pr_{\beta\sim\mathcal{U}}[\mathbf{I}(\langle\beta,\lambda\rangle:\mathcal{H}) > m] \stackrel{*}{<} 2^{-m+\mathbf{I}(\lambda:\mathcal{H})}.$$

Therefore by probabilistic arguments, there exist a set $W \in \{0,1\}^{\infty 2^{n+c\mathbf{K}(n)}}$ such that for all $\beta \in W$,

$$\mathbf{D}(\psi(\beta)|\overline{\mu}_\delta) \leq n+d \text{ and } \mathbf{I}(\psi(W):\mathcal{H}) <^+ \mathbf{I}(W:\mathcal{H}) <^+ \mathbf{I}(\langle W,\lambda\rangle:\mathcal{H}) <^+ \mathbf{I}(\lambda:\mathcal{H}) =^+ \mathbf{I}((G^t,\alpha):\mathcal{H}).$$

Thus Lemma 23 applied to $\psi(W)$ and $\overline{\mu}_\delta$, results in

$$\begin{aligned}
\log|W| &< \max_{\beta\in\psi(W)} \mathbf{D}(\beta|\overline{\mu}_\delta) + 2\mathbf{I}((\psi(W):\mathcal{H}) + O(\mathbf{K}(|W|)) \\
&< \max_{\beta\in\psi(W)} \mathbf{D}(\beta|\overline{\mu}_\delta) + 2\mathbf{I}((G^t,\alpha):\mathcal{H}) + O(\mathbf{K}(|W|)) \\
n + c\mathbf{K}(n) &< n + d + 2\mathbf{I}((G^t,\alpha):\mathcal{H}) + O(\mathbf{K}(n)) \\
c\mathbf{K}(n) &< d + 2\mathbf{I}((G^t,\alpha):\mathcal{H}) + O(\mathbf{K}(n)).
\end{aligned}$$

Thus for proper choice of c, $\mathbf{I}((G^t,\alpha):\mathcal{H}) = \infty$. □

Corollary 33 *Let L be the Lebesgue measure. Let $(\mathcal{X},\mu)$ be a computable measure space. There is a $c \in \mathbb{N}$ such that for (potentially uncomputable) transformation group G^t and point $\alpha \in X$ if $L\{t : \mathbf{t}_\mu(G^t\alpha) > 2^n/\mu(X)\} < 2^{-n-c\mathbf{K}(n)}$, then $\mathbf{I}((G^t,\alpha):\mathcal{H}) = \infty$.*

Chapter 20

Stochastic Thermodynamics

In Stochastic Thermodynamics [Sei08], the evolution operator can randomized. In our manuscript, the evolution operator is the probability kernel, which is defined as follows.

Definition 50 (Probability Kernel) *Given two measurable spaces $\mathcal{A}$ and $\mathcal{B}$, a probability kernel is a function $\kappa : \mathcal{A} \times \mathcal{B} \to \mathbb{R}_{\geq 0}$ such that for all $a \in \mathcal{A}$, $\kappa_a^{\cdot}$ is a probability measure over $\mathcal{B}$, and for all measurable sets $B \subseteq \mathcal{B}$, $\kappa_{\cdot}^{B}$ is a measurable function.*

Given a probability kernel κ, to each measure μ over $\mathcal{A}$ corresponds to a measure over $\mathcal{A} \times \mathcal{B}$. Its marginal over $\mathcal{B}$ is $\kappa^* \mu$. For every measurable function g over $\mathcal{B}$, we define (using Einstein notation)

$$f(x) = \kappa_x g = \kappa_x^y g(y).$$

The operator κ is linear and monotone. Let $\{g_i\}$ be the hat functions used in the proof of Lemma 18. Let ϕ be a surjective computable function from $\mathbb{N}$ to finite sets of $\mathbb{N}$. We say probability kernel κ is computable if

$$f_i(x) = \kappa_x^y \max_{j \in \phi(i)} g_j(y),$$

is uniformly computable in i.

The following theorem is from [Gác21], with a change to the end of the proof.

Theorem 93 (Conservation of Randomness) *Let $\mathcal{X}, \mathcal{Y}$ be a computable metric spaces and let $(\mathcal{X}, \mu)$ be computable probability measure space. Given computable probability kernel κ,*

$$\kappa_x^y \mathbf{t}_{\kappa^* \mu}(y) \overset{*}{<} \mathbf{t}_\mu(x). \tag{20.1}$$

Proof. Let $\mathbf{t}_\mu(x)$ be the universal test over $\mathcal{X}$. The left hand side of Equation 20.1 can be written as $u_\mu = \kappa \mathbf{t}_{\kappa^* \mu}$. Thus $\mu u_\mu = (\kappa^* \mu) \mathbf{t}_{\kappa^* \mu} \leq 1$ since $\mathbf{t}$ is a uniform test. We now show that $u_\mu(x)$ is lower computable. By its construction in the proof of Lemma 18, $\mathbf{t}_\mu(x)$ can be effectively constructed as the supremum of hat functions. Thus for hat functions $\{g_i\}$, there is an enumerable set $N \subseteq \mathbb{N}$ such that $u_\mu = \kappa \sup_{i \in N} g_i$. Since κ is computable, u_μ is lower computable. □

Part IV

Newtonian Gravity

Chapter 21

Introduction

The universal law of gravitation states that the force between two objects is

$$F = G\frac{m_1 m_2}{r^2},$$

where G is the gravitational constant, m_1 is the mass of the first object, m_2 is the mas sof the second object, and r is the distance between the two objects. The gravitational field is a vector field describing the force of several (usually enormous) objects on a receiving object. The force vector applied to an object at position x by n objects at positions $\{x_i\}_{i=1}^n$ with masses $\{m_i\}_{i=1}^n$ is

$$\mathbf{G}_\kappa(x) = -G\sum_{i=1}^{n} m_i \frac{x - x_i}{\|x - x_i\|^3}.$$

The question is how to determine *typical* points in this vector field. In this chapter we proceed with the following approach to tackle this issue. We treat the magnitude of the force vector as a measure κ and define tests t to be lower computable functions over $\mathbb{R}^3$ such that $\kappa(t) < 1$. We prove the existence of a universal lower computable test $\mathbf{T}_\kappa$ and use that a a score of *atypicality* of points in the vector field. This is in line with the construction of algorithmic entropy, introduced in [Gác21].

There are several differences in this approach than that of computable measure spaces in Chapter 13. First off, κ is infinite, so the space $\mathfrak{M}$ of finite measures introduced in Chapter 13 cannot be leveraged. However, because the measure can be defined as a finite number of real numbers, there is a standardized method to encode them for the output of programs or as oracles.

Somewhat surprisingly, we show that $\mathbf{T}_\kappa$ acts like a traditional test over compact spaces with suitable computability properties. Another difference is the assumption of $\mathbb{R}^3$ space instead of a complete, separable space, which enables theorems to leverage this property. In particular, interesting properties we proven about circular orbits of mass points.

The universal test $\mathbf{T}_\kappa$ obeys conservation inequalities and can be defined similarly to the randomnes deficiency term described in Definition 2. It is infinite over the mass points and decreasing in average as a point travels away from the mass point.

Another area covered in this chapter is the orbits of points around mass points. It is proved that typicality $\mathbf{T}_\kappa$ of a point will oscillate as it orbits around a mass point. In addition, under mild conditions, it is proved that two points orbiting around two different mass points cannot have synchronized typicality scores, $\mathbf{T}_\kappa$. On the surface of a sphere, a lower bound is proved on the uniform measure of atypical points that can occur. The chapters in this part of the manuscript are as follows.

- **Chapter 22:** In this chapter, an atypicality score $\mathbf{T}_\kappa$ is defined and some of its fundamental characteristics are detailed.
- **Chapter 23** It is shown that $\mathbf{T}_\kappa$ agree with the universal uniform test $\mathbf{t}_\mu$ up to a multiplicative constant in compact spaces. These results are then applied to orbits and surfaces, including showing that two points orbiting two different paths must be out of sync with respect to their typicality measure $\mathbf{T}_\kappa$.
- **Chapter 24:** In this chapter a chain rule similar to that in Chapter 14 is proven.

Chapter 22

Algorithmic Typicality and Newtonian Gravity

This chapter introduces the Newtonian typicality measure $\mathbf{T}_\kappa$ and describes some fundamental properties about the function. The function $\mathbf{T}_\kappa$ shows a number of interesting properties, including that $\mathbf{T}_\kappa(x) = \infty$ is x is located at a mass point. As $x \in \mathbb{R}^3$ moves away from the mass points, then $\mathbf{T}_\kappa(x)$ approaches 0, in general. $\mathbf{T}_\kappa$ can seen as supremum of terms similar to how the randomness deficiency over infinite sequences, $\mathbf{D}$, is defined, except it uses open boxs instead of cylinders $\alpha[0..n]$. Conservation of randomness is proven when the system contains a single mass point. It is still an open question whether conservation occurs for systems with multiple mass points.

22.1 Preliminaries

Definition 51 *L is the Lebesgue measure over $\mathbb{R}^3$. L_1 is the Lebesgue measure over $\mathbb{R}$.*

Definition 52 (Rational boxs) *The set of all open boxs $s \subset \mathbb{R}^3$ with rational coefficients is S. Elements of S are also referred to as rational open boxs.*

This chapter will leverage computable metric spaces and computable measure spaces, detailed in Chapter 13. It will be particularly useful to define computable measure spaces over closed rational boxs located in $\mathbb{R}^3$. After such a space is defined then results of the previous chapters can be leveraged.

Definition 53 (Computable Measures) *When we say that a measure is R^3-computable if it is a computable measure of the computable metric space $(R^3, \|\cdot\|, \mathbb{Q}^3)$. Equivalently, due to Propositions 14 and 15, measure μ is $\mathbb{R}^3$-computable if $\mu(\mathbb{R}^3)$ is computable and for any finite set of open rational boxs $V \subset S$, $\mu(\cup V)$ is lower computable.*

Definition 54 (Gravitational Field) *The gravity field κ caused by n point masses $\{x_i\}$ is*

$$\mathbf{G}_\kappa(x) = -G \sum_{i=1}^{n} M_i \frac{(x - x_i)}{\|x - x_i\|^3}.$$

Each M_i is the mass of mass point x_i and G is the gravitational constant. The magnitude of the gravitational field is $\|\mathbf{G}_\kappa(x)\|$. We denote κ also as a Borel measure, where for a lower semi-continuous function f over $\mathbb{R}^3$,

$$\kappa(f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y, z) \|\mathbf{G}_\kappa(x, y, z)\| dx dy dz.$$

For Borel set S, $\kappa(S) = \kappa(\mathbf{1}_S)$. The system κ is computable if $\{(m_i, x_i)\}$ is computable and thus the magnitude of the gravitational field at every point can be computed with a program. If κ is given as an oracle, then $\{(m_i, x_i)\}$ is provided in a standardized fashion and thus for all simple functions f, $\kappa(f)$ can be computed.

Remark 7 (Standard Representation) *A representation $\gamma \in \{0,1\}^\infty$ of a system κ is an infinite sequence such that $U(\gamma)$ outputs, with increasing precision, the position and mass of each mass point of κ. A representation λ of κ is standard if for every representation γ of κ there is a program that on input γ outputs λ. In this chapter, when κ is given as an oracle or as an output of a program, it is given as a standard representation. Given a representation $\alpha \in \mathrm{Rep}(\kappa)$, its corresponding system is $\kappa = \mathrm{System}(\alpha)$.*

Definition 55 *A set $W \subset \mathbb{R}^3$ is away from a system κ with mass points $\{x_i\}$ if $\inf_{i, x \in W} \|x - x_i\| > 0$.*

Simple functions are defined as follows. They are used as building blocks for lower computable functions and also tests.

Definition 56 (Simple Function) *A simple function is of the form $f(x,y,z) = [(x,y,z) \in s]q$, where $q \in \mathbb{Q}_{>0}$ and s is a rational box in S. Simple functions can be enumerated as $\{f_n\}$. Given a represenetaion of κ, $\kappa(f_n)$ is uniformly computable in n.*

Definition 57 (Tests) *A lower computable function T is of the form $T(x,y,z) = \sup_{n \in N} f_n(x,y,z)$, where N is an enumerable subset of $\mathbb{N}$. The set of of all lower computable κ-tests is $[\kappa]$, where lower computable functions T is in $[\kappa]$ if*

$$\kappa(T) \leq 1.$$

Among all κ-tests there exist universal tests which dominant every member of $[\kappa]$. This universal test is the key metric for determining the level of atypicality of points in a gravitational field. The greater the score, the greater the level of strangeness of the point. Note that equivalently, $\mathbf{T}_\kappa$ can be defined using lower computable tests over open spheres with rational centers and radii.

Definition 58 (Universal Test) *The universal test is $\mathbf{T}_\kappa(x) = \sum_{T_i \in [\kappa]} \mathbf{m}(i|\kappa) T_i(x)$. $\mathbf{D}_r(x) = \log \mathbf{T}_r(x)$.*

Exercise 28 *Show that $\mathbf{T}_\kappa$ can be equivalently defined using lower computable functions that are built using rational open balls in $\mathbb{R}^3$.*

22.2 Properties of Universal Tests

Despite the fact that for a system κ with mass point x_i, $\lim_{x \to x_i} \mathbf{G}_\kappa(x) = \infty$, $\kappa(B)$ is proportional the radius of balls centered on x_i. Thus, as shown below, as points approach the positions of mass points, they will become more atypical with respct to $\mathbf{T}_\kappa$.

Theorem 94 *Let κ be a system with mass points $\{x_i\}$. Given i, there exists $c, d \in \mathbb{R}^+$ such that for all open spheres G_r centered on x_i with radius $r < c$, $\kappa(G_r) < d \cdot r$.*

Proof. Choose c such that $x_j \notin \overline{G_c}$ for all $j \neq i$. Let κ_j be the system containing just mass point x_j. $\kappa(G_r) \leq \sum_{j=1}^{n} \kappa_j(G_r)$. It is easy to see that since $x_j \notin \overline{G}_c$, for $j \neq i$, we have that $\sum_{j \neq i} \kappa(G_r) < fr^3$ for all $r < c$ and some $f \in \mathbb{R}_{>0}$. $\kappa_i(G_r) = \int_0^r \frac{m_i}{r^2} 4\pi r^2 dr = 4m_i \pi r = gr$, for some $g \in \mathbb{R}_{>0}$. Thus for all $r < c$, $\kappa(G_r) < fr^3 + gr < d \cdot r$. □

Corollary 34 *System κ is a locally finite Borel measure.*

Corollary 35 *For each mass point x_i of a system κ, for each $c > 0$ there is an $r > 0$ such that $\forall_{x:\|x_i - x\| < r} \mathbf{T}_\kappa(x) > c$.*

Proof. Let G_r be an open ball centered on x_i with radius r. By Theorem 94, there is an r_c where $\kappa(G_{r_c}) < 2^{-n}$. We define the test $t_c(x) = [x \in G_{r_c}]2^n$. Thus for all $x \in G_{r_c}$, $\mathbf{T}_\kappa(x) \overset{*}{>} 2^{n - 2\log n}$. □

Exercise 29 *Show that system κ, $\mathbf{T}_\kappa$ is positive over all $\mathbb{R}^3$.*

Thus each center of mass will have infinite atypicality score.

Corollary 36 *For system κ with mass point x, $\mathbf{T}_\kappa(x) = \infty$.*

The following theorem says that the $\mathbf{T}_\kappa$ score will be high near the mass points of the system κ, and then drop off as one moves away from the mass points, as shown in Figure 22.1.

Theorem 95 *Let κ be a system and let $w = \max\{\|x_i\| : x_i \text{ is a mass point of } \kappa\}$. Let U_r be the uniform distribution over the subspace $\{x : r < \|x\| < 2r\}$. Then there is a c where if $r > 2w$ then $U_r(\mathbf{T}_\kappa) < c/r$.*

Proof. Let $\mathbf{T}^r_\kappa(x) = [r < \|x\| < 2r]\mathbf{T}_\kappa(x)$. Let $\theta(z) = \tan^{-1}(w/z)$. Then there is a $c \in \mathbb{N}$ where

$$\begin{aligned}
1 > \kappa(\mathbf{T}^r_\kappa) &= \int_{x,y,z:r<\|(x,y,z)\|<2r} \mathbf{T}_\kappa(x,y,z) \left\| \sum_{i=1}^{n} m_i \frac{\widehat{(x,y,z) - (x_i)}}{\|(x,y,z) - (x_i)\|^3} \right\| dxdydz \\
&\geq \int_{x,y,z:r<\|(x,y,z)\|<2r} \mathbf{T}_\kappa(x,y,z) \left\| \sum_{i=1}^{n} m_i \frac{\cos\theta(r)}{4r^2} \right\| dxdydz \\
&\geq \frac{c}{r^2} \int_{x,y,z:r<\|(x,y,z)\|<2r} \mathbf{T}_\kappa(x,y,z) dxdydz \\
&= \frac{c}{r^2} U_r(\mathbf{T}_\kappa) \frac{4}{3}\pi \left((2r)^3 - r^3\right) \\
U_r(\mathbf{T}_\kappa) &< c/r.
\end{aligned}$$

□

Corollary 37 *Let κ be a system with a single mass point at the origin. Let U_r be the uniform distribution over the subspace $\{x : r < \|x\| < 2r\}$. Then there is a c where if $1 \leq r$ then $U_r(\mathbf{T}_\kappa) < c/r$.*

Theorem 96 *If point $x \in \mathbb{R}^3$ is a computable distance r from a computable point y, then $\mathbf{T}_\kappa(x) = \infty$.*

Figure 22.1: The $\mathbf{T}_\kappa$ scores will be high close to the mass points of system κ and then drop off as one moves away from their positions.

Proof. For positive rational $q \in \mathbb{Q}_{>0}$, let $V_q = \{z : r - q < \|z - y\| < r + q\}$. Let W_q consist of all unions Z of a finite number of open boxs in S such that $V_q \subset Z$. We define the test t_n as follows: enumerate all open sets of W_q for all $q \in \mathbb{Q} \cap (0, 1]$ and stop when an open set $Z \in \cup_{q \in \mathbb{Q} \cap (0,1]} W_q$ is found such that $\kappa(Z) < 2^{-n}$. Then we define the following test, with $t_n(x) = [x \in Z]2^n$. Since $x \in Z$, $t_n(x) = 2^n$. Thus $\mathbf{T}_\kappa(x) \stackrel{*}{>} \sum_n \mathbf{m}(n) t_n(x) \stackrel{*}{>} \sum_n 2^{n - 2 \log n} = \infty$. □

Corollary 38 *If a point $x \in \mathbb{R}^3$ is a computable distance from the mass point of a system κ then $\mathbf{T}_\kappa(x) = \infty$.*

Proof. Though the mass points may not be computable, κ is given as an oracle to the universal test $\mathbf{T}_\kappa$, and thus the mass points are effectively computable. □

Theorem 97 *Given system κ, every point x on a complutable line ℓ in $\mathbb{R}^3$ has $\mathbf{T}_\kappa(x) = \infty$.*

Proof. Let x_1 and x_2 be two computable points on ℓ such that x is on the line segment $[x_1, x_2]$. Let C_n be the open cylinder between x_1 and x_2 with radius 2^{-n}. Volume$(C_n) = \pi 2^{-2n} \|x_1 - x_2\|$. Let test $t_n \in [\kappa]$ be defined by $t_n(y) = [y \in C_n]/\text{Volume}(C_n)$. Since for all $n \in \mathbb{N}$, $t_n(x) \neq 0$, $\mathbf{T}_\kappa(x) \stackrel{*}{>} \sum_n \mathbf{m}(t_n) t_n(x) \stackrel{*}{>} \sum_n \mathbf{m}(n, x_1, x_2) 2^{2n} / (\pi \|x_1 - x_2\|) = \infty$. □

Exercise 30 *Show that every point x on a computable plane has $\mathbf{T}_\kappa(x) = \infty$, for any system κ.*

22.3 Conservation of Randomness

A key property universally shared between notions of typicality and randomness deficiency is conservation of randomness. The author cannot think of a single example of a robust randomness deficiency property that doesn't have this property. Conservation of randomness says that if the measure and the point are transformed by the same means (usually a total computable function) then its randomness deficiency is constant (or usually changed up to an order of the complexity of the transformation function). For example Theorem 43 is a conservation of quantum typicality, over quantum operations. In this section we show that random deficiency with respect to the typicality function $\mathbf{T}_\kappa$ is conserved over non-singular linear transformsations and systems with a single mass point. It is an open question whether there is conservation of randomness over systems with multiple mass points.

For system κ and invertible matrix A, $A\kappa$ is the system where all the mass points have the matrix A applied to them. For invertible matrix A and lower computable functin f, $Af(x) = f(A^{-1}x)$. So $A\mathbf{T}_\kappa(x) = \sum_{T_i \in [\kappa]} \mathbf{m}(i|\kappa) T_i(A^{-1}x)$. Similarly, for 3-vector k, $\kappa + k$ displaces all mass points by k and $(\mathbf{T} + k)_\kappa(x) = \sum_{T_i \in [\kappa]} \mathbf{m}(i|\kappa) T_i(x - k)$. $(A\mathbf{D} + k)(x) = \log(A\mathbf{T} + k)(x)$.

Theorem 98 (Conservation of Randomness) *For system κ with a single mass point (x_i, y_i, z_i), for computable invertible 3×3 matrix A and vector k, We define the transformation of a randomness deficiency function by $(A\mathbf{D} + k)_{A\kappa + k}(x) <^+ \mathbf{D}_\kappa(x)$.*

Proof. Let $\{\lambda_1, \lambda_2, \lambda_3\}$ be the eigenvalues of A, $m = \min |\lambda_i|$ and $M = \max |\lambda_i|$. We use the fact that $m\|x\| \leq \|Ax\| \leq M\|x\|$. We restrict our attention to simple functions, and the generalization to lower computable functions is straightforward. Let simple function f be defined by $f(x, y, z) =$

$[(x,y,z) \in ((x_1,x_2),(y_1,y_2),(z_1,z_2)]v$, with $v \in \mathbb{Q}_{>0}$.

$$\begin{aligned}
A\kappa(f) =& \int\int\int s(x,y,z) \left\| m_i \frac{A(x_i,y_i,z_i)-(x,y,z)}{\|A(x_i,y_i,z_i)-(x,y,z)\|^3} \right\| dxdydz \\
=& v \int_{x_1}^{x_2}\int_{y_1}^{y_2}\int_{z_1}^{z_3} \left\| m_i \frac{A(x_i,y_i,z_i)-(x,y,z)}{\|A(x_i,y_i,z_i)-(x,y,z)\|^3} \right\| dxdydz \\
A\kappa(Af) =& v \int_{x_1}^{x_2}\int_{y_1}^{y_2}\int_{z_1}^{z_2} \left\| m_i \frac{A(x_i,y_i,z_i)-A(x,y,z)}{\|A(x_i,y_i,z_i)-A(x,y,z)\|^3} \right\| dxdydz \\
\leq& \frac{v}{m^3} \int_{x_1}^{x_2}\int_{y_1}^{y_2}\int_{z_1}^{z_2} \left\| m_i \frac{A(x_i,y_i,z_i)-A(x,y,z)}{\|(x_i,y_i,z_i)-(x,y,z)\|^3} \right\| dxdydz \\
\leq& \frac{vM}{m^3} \int_{x_1}^{x_2}\int_{y_1}^{y_2}\int_{z_1}^{z_2} \left\| m_i \frac{(x_i,y_i,z_i)-(x,y,z)}{\|(x_i,y_i,z_i)-(x,y,z)\|^3} \right\| dxdydz. \\
\stackrel{*}{<}& \kappa(f).
\end{aligned}$$

Furthermore, since A is computable, $\mathbf{m}(Af|\kappa) \stackrel{*}{<} \mathbf{m}(f|\kappa)/\mathbf{m}(A)$. The k offset argument is self evident. □

Corollary 39 *For computable matrix A with eigenvalues that have identical absolute values, computable vector k, and system κ. $(A\mathbf{D}+k)_{A\kappa+k}(x) <^+ \mathbf{D}_\kappa(x)$*

Proof. This corollary follows from the reasoning in the proof of Theorem 98. □

22.4 Comparable Definitions of Typicality

A point $x \in \mathbb{R}^3$ is test random with respect to system κ iff $\mathbf{T}_\kappa(x) < \infty$. A κ-ML test is a series of uniformly effectively open sets $\{V_n\}$, relative to oracle κ, such that

$$\kappa(V_n) < 2^{-n}.$$

There is a universal ML test U_n such that if $x \in \cap_n V_n$, then $x \in \cap_n U_n$. We say that a point x is ML random if $x \notin \cap_n U_n$.

Proposition 36 *For system κ, $\mathbf{T}_\kappa(x) \neq \infty$ iff x is ML random.*

Proof. We define the ML test $V_n = \{x : \mathbf{T}_\kappa(x) > 2^n\}$. Clearly $\kappa(V_n) < 2^{-n}$, since $\mathbf{T}_\kappa$ is a test. Thus if $\mathbf{T}_\kappa(x) = \infty$, then $x \in \cap_n V_n \subseteq \cap_n U_n$. Given a ML test V_n, we define the test $T(x) = \sup_n \mathbf{m}(n)[x \in V_n]2^n$. So if $x \in \cap_n V_n$, then $T(x) = \infty$ and thus $\mathbf{T}_\kappa(x) = \infty$. □

Definition 59 (Randomness Deficiency) *The randomness deficiency of a point $x \in \mathbb{R}^3$ with respect to a system κ is $\mathbf{d}(x|\kappa) = \sup_{x \in s \in S} -\log \kappa(s) - \mathbf{K}(s|\kappa)$, where S ranges over all open boxs.*

Definition 60 (Mixed boxs) *Let $\hat{S}$ consist of all "mixed" boxs with rational boundaries, and each boundary can be open or closed. Thus there are 2^6 types of such boxs. One example box is $\{(x,y,z) : x \in [x_1,x_2), y \in (y_1,y_2], z \in [z_1,z_2]\}$. Let $\hat{\mathbf{d}}(x|\kappa) = \sup_{x \in \hat{s} \in \hat{S}} -\log \kappa(\hat{s}) - \mathbf{K}(\hat{s}|\kappa)$.*

Proposition 37 *Given a system κ and mixed box $\hat{s} \in \hat{S}$ there exists a rational open box $s \in S$ such that $\hat{s} \subseteq s$, $\kappa(s) < 2\kappa(\hat{s})$, and $\mathbf{K}(s|\kappa) <^+ \mathbf{K}(\hat{s}|\kappa)$.*

Proof. Because every locally finite Borel measure on a separable complete metric space is regular, κ is regular. So there exists an open cover $W \subset \mathbb{R}^3$ such that $\hat{s} \subseteq W$ and $\kappa(W) < 2\kappa(\hat{s})$. So there exists an open box $s \in S$, $s \subseteq W$ that contains $\hat{s}$. This box can be found using brute force search given $\hat{s}$ and κ. □

Note that the following terms hold over the range of $\mathbb{R} \cup \{\infty\}$.

Theorem 99 *For system κ, $\mathbf{d}(x|\kappa) =^+ \mathbf{D}_\kappa(x)$.*

Proof.

(1) $\mathbf{d}(x|\kappa) <^+ \mathbf{D}_\kappa(x)$. For each $s \in S$, one can define a test $T(x) = \sum_{s \in S} [x \in s]\mathbf{m}(s|\kappa)/\kappa(s)$. It is easy to see that $\kappa(T) \leq \sum_{s \in S} \mathbf{m}(s|\kappa)(\kappa(s)/\kappa(s)) < 1$, and that T is lower computable. So $\mathbf{d}(x|\kappa) <^+ \log T(x) <^+ \mathbf{D}_\kappa(x)$.

(2) If $\mathbf{D}_\kappa(x) \in \mathbb{R}$, then $\mathbf{D}_\kappa(x) <^+ \mathbf{d}(x|\kappa)$. One can assume, without loss of generality, the range of $\mathbf{T}_\kappa$ are 0 and powers of 2. For $n \in \mathbb{Z}$, Let $W_n = \{x : T_r(x) = 2^n\}$. Thus since $\mathbf{T}_r$ is a test, $r(W_n) \leq p(n)2^{-n}$, for some probability p over $\mathbb{Z}$. Let $V_n \subset S$ be the (uniformly in n) enumerable set of disjoint boxs $\hat{s} \in \hat{S}$ such that $\cup_{\hat{s} \in V_n} \hat{s} = W_n$. Each $\hat{s} \in V_n$ can be identified by a code of size $\mathbf{K}(\hat{s}|\kappa) <^+ -\log \kappa(\hat{s}) - n$. Thus if $n = \mathbf{D}_\kappa(x) <^+ -\log \kappa(\hat{s}) - \mathbf{K}(\hat{s}|\kappa) <^+ \hat{\mathbf{d}}(x|\kappa)$.

Let $\hat{s} \in \hat{S}$ realize $\mathbf{D}_\kappa(x) + O(1)$, that is the subset of W_n that x is a member of. Thus by Proposition 37 there is an open box $s \in S$ such that $s \supset \hat{s}$, $\kappa(\hat{s}) < \kappa(s) < 2\kappa(\hat{s})$ and $\mathbf{K}(s|\kappa) <^+ \mathbf{K}(\hat{s}|\kappa)$. Thus $\hat{\mathbf{d}}(x|\kappa) <^+ -\log \kappa(\hat{s}) - \mathbf{K}(\hat{s}|\kappa) <^+ -\log \kappa(s) - \mathbf{K}(s|\kappa) <^+ \mathbf{d}(x|\kappa)$.

(3) If $\mathbf{D}_\kappa(x) = \infty$ then $\mathbf{d}(x|\kappa) = \infty$. Thus there is a ML test $\{V_n\}$ such that $x \in \cap_n V_n$. Fix n. Thus there is a finite or infinite set $W_n \subset \hat{S}$ of disjoint mixed boxs such that $\cup_{\hat{s} \in W_n} \hat{s} = V_n$. Since $\kappa(V_n) < 2^{-n}$, each $\hat{s} \in W_n$ can be identified by a code of size $\mathbf{K}(\hat{s}|\kappa) <^+ -\log \kappa(s) - n + \mathbf{K}(n|\kappa)$. Thus there is some $\hat{s} \in W_n$ such that $x \in \hat{s}$. So $-\log \kappa(\hat{s}) - \mathbf{K}(\hat{s}|\kappa) >^{\log} n$. Since this occurs for each n, $\hat{\mathbf{d}}(x|\kappa) = \sup_{x \in \hat{s} \in \hat{S}} -\log \kappa(s) - \mathbf{K}(\hat{s}|\kappa) = \infty$. Using the same reasoning as in (2), one can see that $\mathbf{d}(x|\kappa) = \infty$. □

Corollary 40 *For computable invertible 3×3 matrix A and vector k, $\mathbf{d}(Ax+k|A\kappa+k) <^+ \mathbf{d}(x|\kappa)$.*

22.5 Systems with Single Mass Points

In this section, we investigate the special case of systems with a single mass point. The mass point is assumed to be at the origin, but the results still hold if it is at any location.

Theorem 100 *Let κ be a computable system with a single mass point z at the origin. For all computable $c \in \mathbb{R}_{>0}$, $\mathbf{T}_\kappa(x) \stackrel{*}{=} \mathbf{T}_{\kappa(cx)}$.*

Proof. The case for $x = (0,0,0)$ is trivial, so we assume $\|x\| > 0$. Assume $t \in [\kappa]$. We construct a new test t' such that $ct'(cx) = t(x)$. Select a simple function $f(x) = v[x \in s]$ of t, where $s \in S$ is an open rational box and enumerate all open rational balls $B \in s$. For each ball B, create a new simple function $f'(x) = [x \in cB]v/c$ to be aggregated into the new test t'. Obviously $ct'(cx) = t(x)$. Furthermore $t' \in [\kappa]$ because of the following reasoning.

$$\kappa(cB) = \int_{(x,y,z) \in cB} \kappa(x,y,z)dxdydz.$$

Substituting $x' = x/c$, $y' = y/c$, $z' = z/c$, we get

$$\begin{aligned}\kappa(cB) =& \int_{(x',y',z')\in B} \kappa(cx', cy', cz')c^3 dx'dy'dz'.\\ =& c\int_{(x',y',z')\in B} \kappa(x', y', z')dx'dy'dz'.\\ =& c\kappa(B).\end{aligned}$$

Thus each ball cB has c more κ measure than B, but this is offset by the fact that t' values are c-times less than that of t. Thus $\mathbf{T}_\kappa(cx) \stackrel{*}{>} \sum_n \mathbf{m}(n|\kappa)t'_n(cx) \stackrel{*}{=} \sum_n \mathbf{m}(n|\kappa)t_n(x)/c \stackrel{*}{=} \mathbf{T}_\kappa(x)$. Since this holds for c and $1/c$, the theorem is proved. □

Theorem 101 *For computable rotation matrix A, and computable system κ with a single mass point at the origin, $\mathbf{T}_\kappa(x) \stackrel{*}{=} \mathbf{T}_\kappa(Ax)$.*

Proof. Let R^t be the computable transformation group associated with A. Let $t \in [\kappa]$. We construct a $t' \in [\kappa]$ such that $t'(A^{-1}x) = t(x)$. Let $f(x) = v[x \in s \in S]$ be a simple function enumerated by the algorithm to compute t. The algorithm to computable t' enumerates open balls $B \subset S$ with center y and radius r and applies R^t to y. By Claim 2 this produces a fast Cauchy sequence $\{y_i\}$ that converges to y such that $\|y - y_i\| < 2^{-i}$. As this sequence is enumeratd, t' takes the supremum of the function $f'(x) = \sup_i v[x \in B_i]$, where $B_i = \{z : \|z - y_i\| < r - 2^{-i}\}$. Since $\kappa(B) = \kappa(R^t B)$, $t' \in [\kappa]$. Furthermore $\mathbf{K}(t') =^+ \mathbf{K}(t)$, up to an additive constant dependent on A. Thus $\mathbf{T}_\kappa(x) \stackrel{*}{=} \mathbf{T}_\kappa(Ax)$. □

Corollary 41 *For system κ with a single mass point at the origin and orbit $O = ((0,0,0), r, \hat{\mathbf{x}}, \hat{\mathbf{y}})$, with computable $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$, and rotational transformation group R^t (from Definition 64), for any computable $\ell \in [0,1], x \in O$, $\sup_{t\in[0,1]} |\mathbf{T}_\kappa(R^t x) - \mathbf{T}_\kappa(R^{t+\ell}x)| < \infty$.*

Definition 61 (Unit Sphere Computable Metric Space) *Let $C = \{x : \|x\| = 1\}$ be the surface of the unit sphere and $D = \{(x,y,z) : \|(x,y,z)\| = 1, (x,y,z) \in \mathbb{Q}\times\mathbb{Q}\times\mathbb{Q}\}$ be its ideal points. Let $w = (0,0,0)$. We define the following computable metric space $\mathcal{W} = (C, d, D)$, where the distance metric d is the radian angle between point in the sphere, with $d(y,v) = \cos^{-1}(\overrightarrow{wy}\cdot\overrightarrow{wv})$. The basis open sets are $B(v,\epsilon) = \{y : d(y,v) > \epsilon, \|\overrightarrow{wy}\| = \|\overrightarrow{wv}\| = 1\}$, where $\epsilon \in \mathbb{Q}_{\geq 0}$ and $(x,y,z) \in \mathbb{Q}\times\mathbb{Q}\times\mathbb{Q}$. Let L_C be the uniform metric over C. Thus $(\mathcal{W}, L_C)$ is a computable measure space.*

The folllowing theorem says that the points along a ray starting from origin will have decreasing $\mathbf{T}_\kappa$ score, where κ is a system containing a single point at origin, as shown in Figure 22.2. Otherwise the ray will be highly atypical. Atypicality is measured using univerisal uniform tests $\mathbf{t}$ over a computer measure space, introduced in Chapter 13. The computable metric space used is the set of vectors of norm 1. Note that all x with $\|x\| = 1$ has $\mathbf{T}_\kappa(x) = \infty$.

Remark 8 *Given a unit vector $v \in C$ and given a system κ, let $R_n(v) = \frac{1}{n}\int_{r=n}^{2n} \mathbf{T}_\kappa(r*v)dr$.*

Theorem 102 *Given a computable metric space $\mathcal{W} = (C, d, D)$ with computable measure space $(\mathcal{W}, L_C)$, if the ray of the unit vector $v \in C$, has the following lower bound of its $\mathbf{T}_\kappa$ measure, with $R_n(v) = \omega(1/\mathbf{m}(n)n)$, then it is very atypical, with $\mathbf{t}_{L_C}(v) = \infty$.*

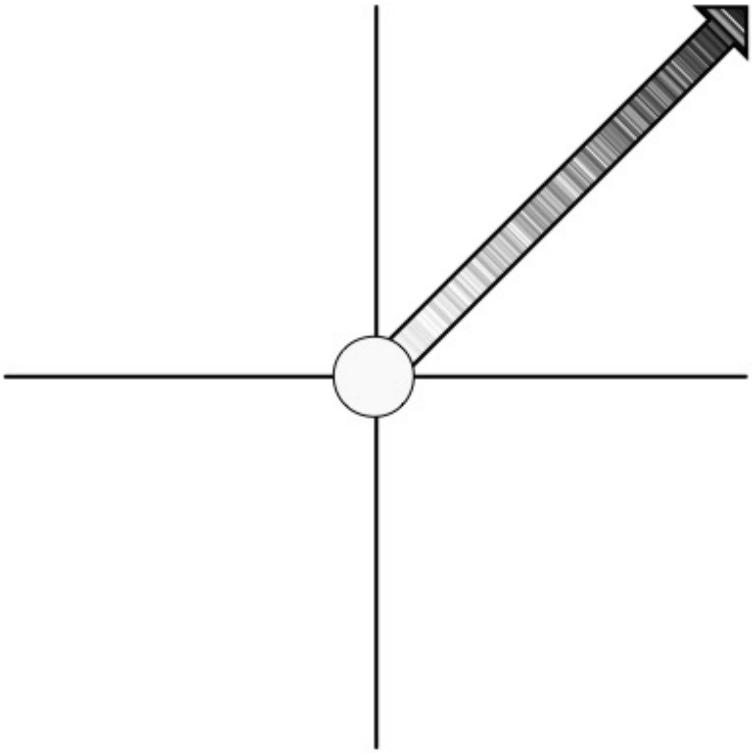


Figure 22.2: The system κ has a single mass point at the origin. All typical rays will start with high $\mathbf{T}_\kappa$ measure, and then the scores will drop off toward 0 as one follows the ray away from origin.

Proof. Let U_r be the uniform distribution over the subspace $\{x : r < \|x\| < 2r\}$. By Corollary 37, there is a c where if $1 \leq r$ then $U_r(\mathbf{T}_\kappa) < c/r$, where U_r be the uniform distribution over the subspace $\{x : r < \|x\| < 2r\}$. Thus $t_n(x) = \frac{n}{c} R_n(x) = \int_{r=n}^{2n} \mathbf{T}_\kappa(r*x)dr$ is a L_C-test over $\mathcal{W}$. Thus $\mathbf{t}_{L_C}(x) \stackrel{*}{>} \sum_n \mathbf{m}(n)t_n(x)$. Since $t_n(v) = \frac{n}{c}\omega(1/\mathbf{m}(n)n)$, $\mathbf{t}_{L_C}(v) \stackrel{*}{>} \sum_n \omega(1) = \infty$. □

22.6 Electrostatics

The results of this part of the manuscript can be readily applied to electrostatics. Coulomb's law parallels Newton's law of gravity, and a universal lower computable test $\mathbf{T}$ can be defined with respect to systems involving a finite number of negative point charges. However, further generalizations quickly reach roadblocks. For example, the magnitude of electric field of an idealized dipole is $\leq c/r^3$, for some constant c and large enough r, the distance from the origin. The corresponding universal test for the dipolte is $\mathbf{T} \in \omega(r^\epsilon)$, for any $\epsilon \in [0, 1)$. Thus for a finite set of negative charges $\lim_{r\to\infty} \mathbf{T}(r) = 0$, but for dipoles, $\lim_{r\to\infty} \mathbf{T}(r) = \infty$, contradicting the goal for coherent properties of $\mathbf{T}$.

Chapter 23

Subspaces, Orbits, and Spheres

In this chapter, properties of the universal test $\mathbf{T}_\kappa$ are proven over sufficiently computable subspaces of $\mathbb{R}^3$. Over such structures, the $\mathbf{T}$ is shown to be asymptotically balanced. Orbits are defined, which are one dimension rings in $\mathbb{R}^3$, and a lower bound of $\mathbf{T}_\kappa$ on rings is proved. Properties of $\mathbf{T}_\kappa$ are proved with respect to surfaces of spheres.

23.1 Subspaces

23.1.1 Computable Open Sets

Definition 62 *An open set $W \subseteq \mathbb{R}^3$ is computable if given an open rational box $s \in S$, $\mathbf{1}_W(s) = [s \subseteq W]$ is computable. Thus $\mathbf{1}_W(x) = [x \in W]$ is lower computable. Examples of computable open sets can be found in Figure 23.1.*

Theorem 103 *Let $V \subseteq \mathbb{R}^3$ be an open computable subset, κ be a computable system, and let $\kappa(V)$ be computable and finite. There is a constant $c \in \mathbb{N}$ where $2^{-n-\mathbf{K}(n)-c} < \kappa(\{x : x \in V, \mathbf{T}_\kappa(x) > 2^n\}) < 2^{-n+c}$.*

Proof. For Borel set X, let $\kappa'(X) = \kappa(X \cap V)$. So $\mathcal{R} = (\mathbb{R}^3, \|\cdot\|, \mathbb{Q}^3)$ is a computable metric space and $(\mathcal{R}, \kappa')$ is a computable measure space. This is because κ' is a $\mathbb{R}^3$-computable measure since

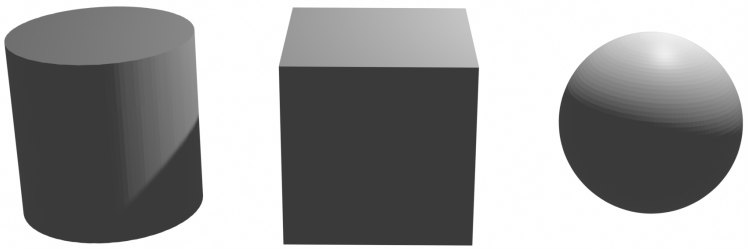


Figure 23.1: A graphical depiction three computable open sets, provided that their positions and shape parameters are computable. Any finite combination of the sets are also computable. There also exists non-compact computable open sets.

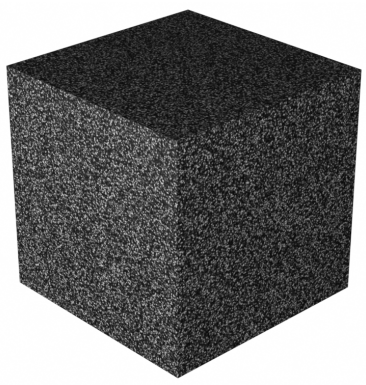


Figure 23.2: For compact open sets, there is a lower bound on the $\mathbf{T}_\kappa$ measure, with respect to the Lebesgue measure, L.

Proposition 15 applies because $\kappa'(V) = \kappa(V)$ is computable and for every $s \in S$,

$$\kappa'(s) = \kappa\left(\bigcup\{s' : s' \subseteq s \text{ and } s \subseteq V\}\right)$$

is lower computable because V is computable and κ is computable. Since V is computable, and $\mathbf{t}_{\kappa'}(V) < 1$, $\mathbf{t}_{\kappa'} \stackrel{*}{<} \mathbf{T}_\kappa$. Also $\mathbf{t}'_\kappa(x) \stackrel{*}{>} [x \in V]\mathbf{T}_\kappa(x)$. So over $x \in V$ $\mathbf{t}_{\kappa'} \stackrel{*}{=} \mathbf{T}_\kappa$. By Proposition 26, there is a $c \in \mathbb{N}$ where for all $n \in \mathbb{N}$

$$\kappa'(V)2^{-n-\mathbf{K}(n)-c} < \kappa'\{x : x \in V, \mathbf{t}_{\kappa'}(x) > 2^n/\kappa'(V)\} < \kappa'(V)2^{-n}.$$

The $\kappa'(V)$ term can be folded in the constant c. So,

$$2^{-n-\mathbf{K}(n)-c} < \kappa\{x : x \in V, \mathbf{T}_\kappa(x) > 2^n\} < 2^{-n+c}.$$

□

The following Corollary provides a lower bound for the $\mathbf{T}_\kappa$ measure in open computable sets, with respect to the Lebesgue measure, as shown in Figure 23.2.

Corollary 42 *Let V be a compact computable open set and let κ be a computable system with computable $L(V)$. Then there is a $c \in \mathbb{N}$ such that for all $n \in \mathbb{N}$, $L\{x : x \in V, \mathbf{T}_\kappa(x) > 2^n\} > 2^{-n-\mathbf{K}(n)-c}$.*

Proof. For Borel set W, the function $\lambda(W) = L(W \cap V)/L(V)$ is a $\mathbb{R}^3$-computable probability measure. By Corollary 29, there is a $c \in \mathbb{N}$ where

$$\lambda\{x : x \in V, \mathbf{t}_\kappa(x) > 2^n\} > 2^{-n-\mathbf{K}(n)-c}.$$

Using the reasoning in Theorem 103, $\mathbf{t}_\kappa \stackrel{*}{=} \mathbf{T}_\kappa$ over V. So

$$\lambda\{x : x \in V, \mathbf{T}_\kappa(x) > 2^n\} > 2^{-n-\mathbf{K}(n)-c}.$$

The corollary follows from the fact that in the domain of V, $L(V)\lambda(\cdot) = L(\cdot)$. □

Corollary 43 *Let V be an open compact computable set away from a computable system κ and let $\kappa(V)$ be computable. There is a constant $c \in \mathbb{N}$ where $2^{-n-\mathbf{K}(n)-c} < L\{x : x \in V, \mathbf{T}_\kappa(x) > 2^n\} < 2^{-n+c}$.*

Proof. Since V is compact and away from κ, there are constants $b_1, b_2 \in \mathbb{R}_{>0}$ such that $b_1 < \kappa(x) < b_2$ for all $x \in V$. So for Borel set $A \subseteq V$, $b_1 \int_A dxdydz < \kappa(A) < b_2 \int_A dxdydz$. This means $b_1 L(A) < \kappa(A) < b_2 L(A)$, for $A \subseteq V$. Applying this equation to Theorem 103 results in the corollary. □

23.1.2 Borel Compact Sets

Theorem 104 *Let V be a Borel compact set, κ be a computable system κ, and $\mathbb{R}^3$-computable non-atomic probability measure λ have support equal to V. Then there is a $c \in \mathbb{N}$ where $\lambda\{x : x \in V, \mathbf{T}_\kappa(x) > 2^n\} > 2^{-n-\mathbf{K}(n)-c}$.*

Proof. Let W be a open computable compact set containing V and let $\kappa'(A) = \kappa(A \cap W)$. Thus $\mathcal{R} = (\mathbb{R}^3, \|\cdot\|, \mathbb{Q}^3)$ is a computable metric space and $(\mathcal{R}, \kappa')$ is a computable measure space. Using the arguments in Theorem 103, for all $x \in W$, $\mathbf{t}_{\kappa'} \overset{*}{=} \mathbf{T}_\kappa(x)$. Applying Corollary 29 to λ and κ' provides a $c \in \mathbb{N}$ such that

$$\begin{aligned} 2^{-n-\mathbf{K}(n)-c} <& \lambda\{x : \mathbf{t}_{\kappa'}(x) > 2^n\}, \\ 2^{-n-\mathbf{K}(n)-c} <& \lambda\{x : x \in V, \mathbf{T}_\kappa(x) > 2^n\}. \end{aligned}$$

□

23.2 Orbits

Notions of typicality can be applied to orbits. Orbits are circular 1 dimensional paths typially around a mass point. If an orbit is at a computable distance from a mass point, then all its points will have infinite $\mathbf{T}$-scores. Otherwise, as a point orbits around a mass point, it will have oscilliations of typicality, similarily to oscillations of algorithmic fine grained entropy during dynamics.

Definition 63 (Orbit) *Let $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ be two perpendicular unit vectors in $\mathbb{R}^3$. Let κ be a system. An orbit is defined by $\{r\cos(\theta)\hat{\mathbf{x}} + r\sin(\theta)\hat{\mathbf{y}} + z : \theta \in [0, 2\pi], r \in \mathbb{R}_{>0}\}$, where $z \in \mathbb{R}^3$ is typically at a mass point. A representation $\alpha \in \mathrm{Rep}(O) \subset \{0,1\}^\infty$ of an orbit $O = (r, \hat{\mathbf{x}}, \hat{\mathbf{y}}, z)$ is any sequence such that $U(\alpha)$ outputs r, z, $\hat{\mathbf{x}}$, and $\hat{\mathbf{y}}$ to any degree of precision. An orbit is computable if it has a computable representation. The amount of information that an orbit O has with the halting sequence $\mathcal{H}$ is $\mathbf{I}(O : \mathcal{H}) = \inf_{\alpha \in \mathrm{Rep}(O)} \mathbf{I}(\alpha : \mathcal{H})$. We use the notation (r, z)-orbit to specify an orbit with arbitrary $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ vectors.*

Definition 64 (Rotational Transformation Group) *A rotational transformation group R^t is a transform group (from Definition 36) and is defined by $(z, \hat{\mathbf{x}}, \hat{\mathbf{y}})$, with center $z \in \mathbb{R}^3$ and an axis of rotational, with unit vectors $\hat{\mathbf{x}}, \hat{\mathbf{y}} \in \mathbb{R}^3$. It performs a rotation around the center aligned with the axis of rotation. In addition, for all $x \in \mathbb{R}^3$, $R^0 x = x$ and $R^t = R^{t+1}$. Obviously every orbit defines a rotational transformation group. A representation of a transformation group $R^t = (z, \hat{\mathbf{x}}, \hat{\mathbf{y}})$, is any sequence $\alpha \in \mathrm{Rep}(R^t)$ is any sequence such that $U(\alpha)$ outputs z, $\hat{\mathbf{x}}$, and $\hat{\mathbf{y}}$ to any degree of precision. The amount of information that a rotation group has with the halting sequence is $\mathbf{I}(R^t : \mathcal{H}) = \inf_{\alpha \in \mathrm{Rep}(R^t)} \mathbf{I}(\alpha :: \mathcal{H})$. For $x \in \mathbb{R}^3$, $\mathbf{I}((R^t, x) : \mathcal{H}) = \inf_{\alpha \in \mathrm{Rep}(R^t), \overrightarrow{x} \in \mathrm{Cauchy}(x)} \mathbf{I}(\langle \alpha, \langle \overrightarrow{x} \rangle \rangle : \mathcal{H})$.*

Claim 2 *Given a representation α of an orbit O, there is a computable rotational transformation group R^t around orbit O, such that for all $x \in O$, $R^0 x = R^1 x = x$. Furthermore given a program or representation of R^t, one can compute a rotation. That is, given a fast cauchy sequence sequence of x and an $t \in \mathbb{R}_{\geq 0}$ one can compute a fast Cauchy sequence for $R^t x$.*

Figure 23.3: Orbits will oscillate in $\mathbf{T}_\kappa$ value.

Corollary 44 *If any of the following conditions occur for an orbit O then $\mathbf{T}_\kappa(x) = \infty$, for all $x \in O$.*

(1) O is computable,

(2) O is a computable distance from a mass point of system κ,

(3) O is centered at a mass point with a computable axis of rotation,

Proof. (1) and (2) follow from Theorem 96. For (3), otherwise let $O = (r, \hat{\mathbf{x}}, \hat{\mathbf{y}}, z)$, where z is a mass point and $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ is computable. Let $r_1 < r < r_2$, $r_1, r_2 \in \mathbb{Q}_{>0}$. Let

$$W_q = \{x : r_1 < \|x - z\| < r_2, x \text{ differs from } (\hat{\mathbf{x}}, \hat{\mathbf{y}}) \text{ axis by } q \text{ radians}\}$$

Given n, one can find a q_n such that $\kappa(W_{q_n}) < 2^{-n}$ and define a test $t_n(x) = [x \in W_{q_n}]2^n$. So for all $x \in O$, $\mathbf{T}_\kappa(x) \stackrel{*}{>} \sum_n \mathbf{m}(n|\kappa)t_n(x) \stackrel{*}{>} \sum_n 2^{n-2\log n} = \infty$. □

Corollary 45 *Let κ be a system with mass point x_i. For all $c > 0$, there is an $r > 0$ such that for all (r, x_i) orbits O, $\mathbf{T}_\kappa(x) > c$, for all $x \in O$.*

Proof. This follows from Corollary 35. □

Remark 9 *We recall that the mutual information of a point $x \in \mathbb{R}^3$ with the halting sequence $\mathcal{H}$, is $\mathbf{I}(x : \mathcal{H})$, and is introduced in Definition 42.*

23.2.1 One Orbit

As described in the introduction, orbits will oscillate in $\mathbf{T}_\kappa$ values with larger fluctuations being more rare. A graphical depiction of Theorem 105 can be found in Figure 23.3

Theorem 105 *Let κ be a computable system with mass point z and $O = (r, \hat{\mathbf{x}}, \hat{\mathbf{y}}, z)$ be an orbit such that $\mathbf{I}(O : \mathcal{H}) < \infty$. There is a $c \in \mathbb{N}$ where for all $n \in \mathbb{N}$,*

$$2^{-n-c\mathbf{K}(n)} < L_1\{x : x \in O, \mathbf{T}_\kappa(x) > 2^n\}.$$

Proof. By Claim 2, given O, one can computable a (potentially uncomputable) rotational tranformation group R^t such that for all $x \in O$, $R^t x \in O$ for $t \in [0,1]$, $R^0 x = R^1 x = x$. This transformation group R^t is a rotation centered around mass point z and oriented with the axis $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$. Furthermore there is an x computable from O. Using Definition 64, $\mathbf{I}((R^t, x) : \mathcal{H}) <^+ \mathbf{I}(O : \mathcal{H}) < \infty$. Let $\beta \in \{0,1\}^\infty$ realize $\lceil \mathbf{I}((R^t, x) : \mathcal{H}) \rceil + 1$.

Let $B = \{y : \|y - z\| < R\}$ where $r < R$, $R \in \mathbb{Q}_{>0}$. For Borel set A, let $\kappa'(A) = \kappa(A \cap B)$. Thus $\mathcal{R} = (\mathbb{R}^3, \|\cdot\|, \mathbb{Q}^3)$ is a computable metric space and $(\mathcal{R}, \kappa')$ is a computable measure space. In addition B is a compact open computable set with computable $\kappa(B)$. Thus using arguments in the proof of Theorem 103, $\mathbf{t}_{\kappa'}(x) \stackrel{*}{=} \mathbf{T}_\kappa(x)$ for all $x \in B$. The application of Corollary 33, applied to κ' and the representation β of (R^t, x) that will realize Definition 49 using Claim 2, and noting that the transform group need not be measure preserving, with the dynamics starting at x results in a $c \in \mathbb{N}$ where

$$\begin{aligned} 2^{-n-c\mathbf{K}(n)} <& L_1\{t \in [0,1], \mathbf{t}_{\kappa'}(R^t x) > 2^n\} \\ 2^{-n-c\mathbf{K}(n)} <& L_1\{y : y \in O, \mathbf{t}_{\kappa'}(y) > 2^n\}. \end{aligned} \tag{23.1}$$

If the system κ only contains a single mass point z and $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ is computable, then R^t is a computable transformation tranform rotating around z. Furthermore R^t is measure preserving over the computable measure space $(\mathcal{R}, \kappa')$. Thus applying Theorem 24 results $c \in \mathbb{N}$ where

$$\begin{aligned} 2^{-\mathbf{K}(n)-n-c} <& L_1\{t : t \in [0,1], \mathbf{t}_{\kappa'}(R^t x) > 2^n\} < 2^{-n} \\ 2^{-\mathbf{K}(n)-n-c} <& L_1\{y : y \in O, \mathbf{t}_{\kappa'}(y) > 2^n\} < 2^{-n}. \end{aligned} \tag{23.2}$$

The fact that $\mathbf{t}_{\kappa'}(x) \stackrel{*}{=} \mathbf{T}_\kappa(x)$ for all $x \in O \subset B$ can be applied to Equation 23.1, proving the theorem. □

23.2.2 Two Orbits

In Theorem 105, it was shown that non-exotic orbits will oscillate in typicallity. Theorem 106 extends this result to two orbits. Given two orbits with two starting points that are not atypical or exotic, the typicality of the points will be out of sync as they orbit the mass points. Thus this theorem applies to cases where two points are orbiting the origin, but with very small (uncomputable) orbit error vectors. This is consistent with reality, as absolutely perfect orbits don't exist. A graphical depiction of Theorem 106 can be seen in Figure 23.4.

Definition 65 (Extended Universal Tests) *The set of cylinders in the Cantor space is denoted by $\mathcal{C}$, where $\zeta \in \mathcal{C}$ if there is an $x \in \{0,1\}^*$ such that $\zeta = \{\alpha : x \sqsubset \alpha \in \{0,1\}^\infty\}$. Given the space $\mathcal{Z} = \mathbb{R}^3 \times \mathbb{R}^3 \times \{0,1\}^\infty \times \{0,1\}^\infty$, a basis extended box $p \in P$ with respect to $\mathcal{Z}$, are two rational boxs and two intervals $p = (s_1, s_2, \zeta_1, \zeta)$, where $s_1, s_2 \in S$ and $\zeta_1, \zeta_2 \in \mathcal{C}$. A simple function $f : \mathcal{Z} \to \mathbb{R}_{\geq 0}$ is of the form $f_{s_1,s_2,\zeta_1,\zeta_2}(w,x,y,z) = v[w \in s_1, x \in s_2, y \in \zeta_1, z \in \zeta_2]$, with $v \in \mathbb{Q}_{>0}$. Simple functions can be enumerated producing the list $\{f_n\}$. A lower computable function F is of the form $F(w,x,y,z) = \sup_{n \in N} f_n(w,x,y,z)$ where N is an enumerable subset of $\mathbb{N}$. Given computable measure μ over $\{0,1\}^\infty$ and system κ, a lower computable function F is a (κ,κ,μ,μ)-test, or $f \in [\kappa,\kappa,\mu,\mu]$ if*

$$(\kappa,\kappa,\mu,\mu)(F) = \int_{\alpha,\beta \in \{0,1\}^\infty} \int_{x_1 \in \mathbb{R}^3, x_2 \in \mathbb{R}^3} F(x_1,x_2,\alpha,\beta), d\kappa(x_1) d\kappa(x_2) d\mu(\alpha) d\mu(\beta) \leq 1.$$

Given κ and μ, the set $[\kappa,\kappa,\mu,\mu]$ is enumerable, so there exists a universal lower computable test $\mathbf{T}_{(\kappa,\kappa,\mu,\mu)}(w,x,y,z) = \sum_{f_i \in [\kappa,\kappa,\mu,\mu]} \mathbf{m}(i|\kappa,\mu) f_i(w,x,y,z)$.

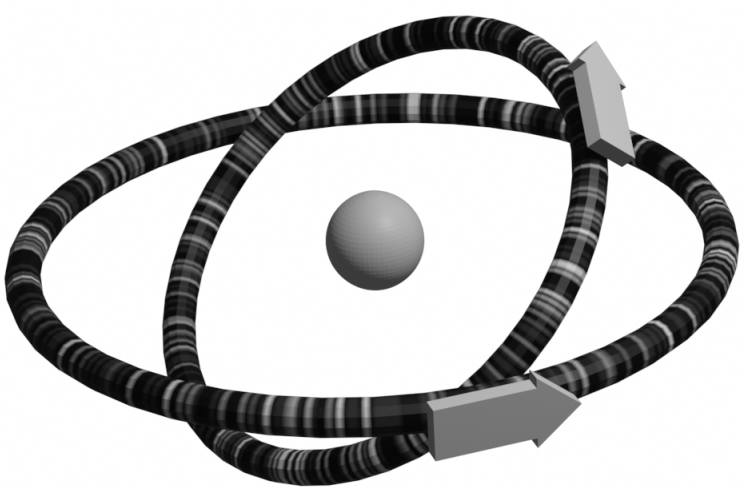


Figure 23.4: Two points rotating under two rotational transformation groups, centered around a single mass point at the origin. Provided that the two starting points and encodings of the transformation groups are typical and not exotic, the $\mathbf{T}_\kappa$ score of the two points will be out of sync as they rotate around the mass point.

Definition 66 *This definition introduces the mutual information between two points in $\mathbb{R}^3$ and two infinite sequences with the halting sequence $\mathcal{H}$. Let* Cauchy(x) *be all the fast Cauchy sequences converging to $x \in \mathbb{R}^3$, using the computable metric space $(R^3, \|\cdot\|, \mathbb{Q}^3)$. For $\zeta \in$ Cauchy(x), $\langle\zeta\rangle \in \{0,1\}^\infty$ is an encoding of the fast Cauchy sequence by an infinite encoding of each index of the ideal points converging to x. For points $x_1, x_2 \in \mathbb{R}^3$ and infinite sequences $\alpha, \beta \in \{0,1\}^\infty$, their multiple information with the halting sequence is*

$$\mathbf{I}((x_1, x_2, \alpha, \beta) : \mathcal{H}) = \inf_{(\xi,\zeta)\in(\text{Cauchy}(x_1),\text{Cauchy}(x_2))} \mathbf{I}(\langle\langle\xi\rangle, \langle\zeta\rangle, \alpha, \beta\rangle : \mathcal{H}).$$

Remark 10 *We recall Definition 41, where an r-interval $v \subseteq (0,1)$ is a finite collection of open intervals with rational endpoints.*

Theorem 106 *Let κ be a computable system with a single mass point at the origin, μ a computable probability measure over $\{0,1\}^\infty$ and R_1^t and R_2^t be rotational transformation groups orbiting the origin, with $(\alpha_1, \alpha_2) \in (\text{Rep}(R_1^t), \text{Rep}(R_2^t))$ and $x_1, x_2 \in \mathbb{R}^3$. If $\mathbf{T}_{(\kappa,\kappa,\mu,\mu)}(x_1, x_2, \alpha_1, \alpha_2) < \infty$ and $\mathbf{I}((x_1, x_2, \alpha_1, \alpha_2) : \mathcal{H}) < \infty$ then $\sup_{t\in[0,1]} |\mathbf{T}_\kappa(R_1^t x_1) - \mathbf{T}_\kappa(R_2^t x_2)| = \infty$.*

Proof. Assume not and let

$$d = \lceil \sup_{t\in[0,1]} |\mathbf{T}_\kappa(R_1^t x_1) - \mathbf{T}_\kappa(R_2^t x_2)| \rceil < \infty.$$

Let O_1 be the orbit of x_1 with R_1^t. Thus $\mathbf{I}(O_1 : \mathcal{H}) <^+ \mathbf{I}((x_1, x_2, \alpha_1, \alpha_2) : \mathcal{H}) < \infty$. Thus by Theorem 105, for all $x \in O_1$ and $c \in \mathbb{N}$ where

$$2^{-n-c\log n} < L_1\{t \in [0,1] : \mathbf{T}_\kappa(R^t x) > 2^n\},$$
$$2^{-n-c\log n} < L_1\{t \in [0,1] : \mathbf{T}_\kappa(R^t x_1) > 2^n\}.$$

Let $U_n = \{t : \mathbf{T}_\kappa(R^t x_1) > 2^n\}$. Given x_1 and α_1, one can enumerate an increasing r-interval $v \subseteq U_n$ and stop when $L(v) > 2^{-n-c\log n-1}$. By Lemma 26, there exists a rational $r \in v$, with

$$\mathbf{K}(r) <^{\log} n + \mathbf{I}(v;\mathcal{H}) <^{\log} n + \mathbf{I}((x_1,x_2,\alpha_1,\alpha_2):\mathcal{H}) + \mathbf{K}(n) <^{\log} n + \mathbf{I}((x_1,x_2,\alpha_1,\alpha_2):\mathcal{H}).$$

Let $A_n = \{x : \mathbf{T}_\kappa(x) > 2^n\}$. Thus $\kappa(A_n) < 2^{-n}$. Thus one can create the following $\mathcal{Z}$-test

$$\begin{aligned} t_n(x,y,\alpha,\beta) &= [x \in R_1^{-r} A_n, y \in R_2^{-r} A_{n-d}] 2^{2n-d} \\ \mathbf{K}(t_n) &<^+ \mathbf{K}(r,n) <^{\log} n + \mathbf{I}((x_1,x_2,\alpha_1,\alpha_2):\mathcal{H}) \\ &<^{\log} n, \end{aligned} \tag{23.3}$$

where Equation 23.3 is due to the fact that $\mathbf{I}((x_1,x_2,\alpha_1,\alpha_2):\mathcal{H})$ is finite. $\int t_n$ because R_1^t and R_t^2 are κ measure preserving. One can impliment t_n to be a lower computable function in the following manner. As open balls B with center x and radius s are enumerated that are subsets of A_n (or A_{n-d}), the algorithm for t_n rotates them with R_1^t (and also R_2^t), consistently with Claim 2.

As t_n reads the bits of α and β it produces a fast Cauchy sequence $\{x_n\}$ of points that converge to $R^{-r}x$ where $\|x_n - x\| < 2^{-n}$. The algorithm for test t_n creates as sequence of balls $\{B^i\}$, $B^i = \{y : \|y - x_i\| < s - 2^{-i}\}$ such that $B^i \subseteq B^{i+1}$ and $\lim_{i\to\infty} B^i = R^{-r}B$.

Thus for each $B \supset A_n$, $C \in A_{n-d}$, t_n enumerates two balls B^i and C^i (one for R_1^t and another for R_2^t) and if a point $(y,z) \in B^i \times C^i$, then $t(y,z,\alpha_1,\alpha_2) = 2^{2n-d}$. Let B_i and C_j be the balls enumerated by t_n where $\kappa(B^i) \le \kappa(B)$ and $\kappa(C^i) \le \kappa(C)$. Furthermore, let t_n numerate m balls after reading m bits of the encoded rotation matrices R_t^1 and R_2^t. The function $t_n \in [\kappa,\kappa,\mu,\mu]$ because

$$\begin{aligned} (\kappa,\kappa,\mu,\mu)(t_n) &= \int_{\alpha,\beta\in\{0,1\}^\infty} \int_{x_1,x_2\in\mathbb{R}^3} t_n(x_1,x_2,\alpha,\beta) d\kappa(x_1) d\kappa(x_2) d\mu(\alpha) d\mu(\beta) \\ &\le \sup_{n\to\infty} \sum_{a,b\in\{0,1\}^n} \mu(a)\mu(b) \int_{x_1,x_2\in(\bigcup_{i=1}^n B_i^a, \bigcup_{j=1}^n C_j^b)} 2^{2n-d} d\kappa(x_1) d\kappa(x_2) \\ &\le 2^{2n-d} \sup_{n\to\infty} \sum_{a,b\in\{0,1\}^n} \mu(a)\mu(b) \sum_{i,j} \int_{x_1,x_2\in(\bigcup_{i=1}^n B_i^a, \bigcup_{j=1}^n C_j^b)} d\kappa(x_1) d\kappa(x_2) \\ &\le 2^{2n-d} \sup_{n\to\infty} \sum_{a,b\in\{0,1\}^n} \mu(a)\mu(b) \kappa\left(\bigcup_{i=1}^n B_i^a\right) \kappa\left(\bigcup_{j=1}^n C_j^b\right) \\ &\le 2^{2n-d} \sup_{n\to\infty} \sum_{a,b\in\{0,1\}^n} \mu(a)\mu(b) \kappa\left(\bigcup_{i=1}^n B_i\right) \kappa\left(\bigcup_{j=1}^n C_j\right) \\ &\le 2^{2n-d} \sup_{n\to\infty} \sum_{a,b\in\{0,1\}^n} \mu(a)\mu(b) \sum_{i,j} \kappa(A_n)\kappa(A_{n-d}) \\ &\le 1, \end{aligned}$$

where B^a is the intermediate ball create by the t_n algorithm when just using the bits $a \in \{0,1\}^*$ (and similarly for C^b). By construction, such balls will have $\kappa(B^a) < 2^{-n}$ and $\kappa(C^b) < 2^{-n+d}$.

It must be that $t_n(x_1, x_2, \alpha_1, \alpha_2) \neq 0$ because $r \in U_n$. Thus

$$\begin{aligned}
\mathbf{T}_{(\kappa,\kappa,\mu,\mu)}(x_1, x_2, \alpha_1, \alpha_2) &\overset{*}{>} \sum_n \mathbf{m}(t_n|\kappa, \mu) t_n(x_1, x_2, \alpha_1, \alpha_2) \\
&\overset{*}{>} \sum_n \mathbf{m}(n, r_n) t_n(x_1, x_2, \alpha_1, \alpha_2) \\
&\overset{*}{>} \sum_n 2^{-n-O(\log n)} 2^{2n-d} \\
&= \infty,
\end{aligned}$$

causing a contradiction. □

23.3 Surfaces of Spheres

The same techniques used on orbits can also be applied to surfaces of spheres. If a sphere has finite mutual information with the halting sequence , then there will be a lower bound on the measure of the surface of the sphere of points with high atypicality $\mathbf{T}_\kappa$ scores.

Definition 67 *A sphere C is given by (x, r) where $x \in \mathbb{R}^3$ is its center and $r \in \mathbb{R}_{>0}$ is it radius. A representation of a sphere is any sequence in $\text{Rep}(C) \subset \{0,1\}^\infty$ that can produce (x, r) to any degree of precision $\mathbf{I}(C : \mathcal{H}) = \inf_{\alpha \in \text{Rep(C)}} \mathbf{I}(C : \mathcal{H})$.*

23.3.1 Uniform Sampling on a Sphere

Using inverse transform sampling, one can uniformly sample point on the unit sphere. This can be seen in Figure 23.5.

Proposition 38 *Given a representation of a sphere C there is a computable mapping from the distribution $\text{Uniform}([0,1],[0,1])$ to $\text{Uniform}(C)$. That is, two independent uniformly distributed numbers between 0 and 1 can be used to create the uniform measure over C.*

Proof. Let v be a point on the unit sphere C. We want the probability density $f(v)$ to be constant for a uniform distribution. So $f(v) = \frac{1}{4\pi}$ since $\int\int_C f(v)dA = 1$ and $\int\int_C dA = 4\pi$. So, using spherical coordinates,

$$f(v)dA = \frac{1}{4\pi} dA = f(\theta, \phi) d\theta d\pi.$$

Since $dA = \sin(\phi) d\phi d\theta$, it follows that $f(\theta, \phi) = \frac{1}{4\pi} \sin(\phi)$. The marginal distribution are

$$\begin{aligned}
f(\theta) &= \int_0^\pi f(\theta, \phi) d\phi = \frac{1}{2\pi} \\
f(\phi) &= \int_0^{2\pi} f(\theta, \phi) d\theta = \frac{\sin(\phi)}{2}.
\end{aligned}$$

The cumulative distribution of $f(\phi)$ is

$$F(\phi) = \int_0^\phi f(\hat{\phi}) d\hat{\phi} = \frac{1}{2}(1 - \cos(\phi)).$$

The inverse is $F^{-1}(u) = \arccos(1-2u)$. We then proceed with inverse transform sampling. Let U be the random number in $[0,1]$. Noting that $\Pr(U \leq F(\phi)) = F(\phi)$, we get $\Pr(F^{-1}(U) \leq \phi) = F(\phi)$. So one can generate a point on the uniform sphere, one samples θ from $2\pi \times U$ and sample ϕ from $F^{-1}(U)$. □

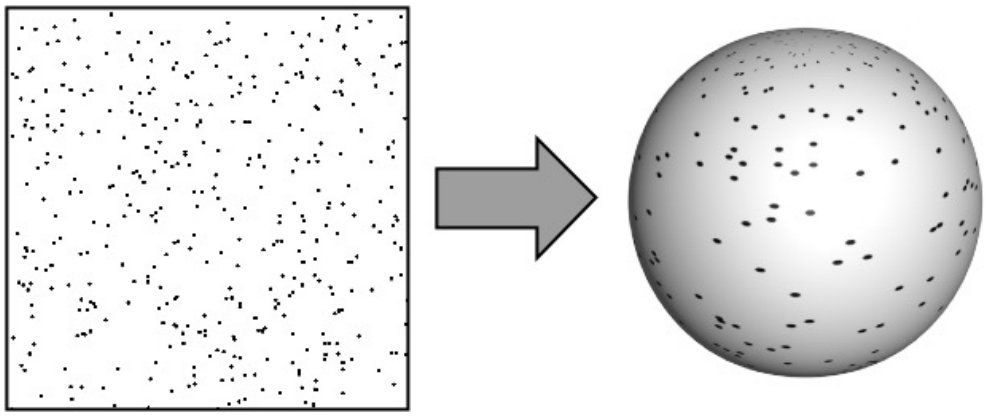


Figure 23.5: Using inverse transform sampling, random points in $[0,1] \times [0,1]$ can be mapped uniformly to the unit sphere. This process is used in the proof of Lemma 33.

23.3.2 Spheres and Computable Measure Spaces

Lemma 33 *Let $\mathcal{R} = (\mathbb{R}^3, \|\cdot\|, \mathbb{Q}^3)$ be the standard computable metric space over $\mathbb{R}^3$ and let $(\mathcal{R}, \mu)$ be a computable measure space, as defined in Chapter 13. Let $C = (x, r)$ be a sphere in $\mathbb{R}^3$, and let L_C be the uniform measure over the sphere. There is a $c \in \mathbb{N}$ such that if $L_C\{\alpha : \|\alpha - x\| = r, \mathbf{t}_\mu(\alpha) > 2^n/\mu(X)\} < 2^{-n-c\mathbf{K}(n)}$, then $\mathbf{I}(C : \mathcal{H}) = \infty$.*

Proof. We assume C has radius 1, the general case is straightforward. Let L_2 be the two dimensional Lebesgue measure. Fix n. Let $(\{0,1\}^\infty, \Gamma)$ be the Cantor space with the uniform measure. The binary representation (see Theorem 70) creates an isomorphism (ϕ, ϕ^{-1}) of computable probability spaces between the spaces $(\{0,1\}^\infty, \Gamma)$ and $([0,1] \times [0,1], L_2)$. It is the function $\phi(\gamma) = (0.\gamma[0,2,4,\ldots], 0.\gamma[1,3,5,\ldots])$. Let $J : [0,1] \times [0,1] \to C$ uniformly map the unit interval to C, as defined by Proposition 38. Thus for all 2D Borel sets $A \subseteq [0,1] \times [0,1]$, $\Gamma(\phi^{-1}(A)) = L_2(A)$. So

$$\begin{aligned}
&L_C\{\alpha : \|\alpha - x\| = r, \mathbf{t}_\mu(\alpha) > 2^n/\mu(X)\} < 2^{-n-c\mathbf{K}(n)} \\
&L_2\{t, s : \mathbf{t}_\mu(J(t,s)) > 2^n/\mu(X)\} < 2^{-n-c\mathbf{K}(n)} \\
&\Gamma\{\alpha : \mathbf{t}_\mu(J(\phi(\alpha)) > 2^n/\mu(X)\} < 2^{-n-c\mathbf{K}(n)}.
\end{aligned}$$

Let (δ, μ_δ) be a binary representation for $(\mathcal{R}, \mu)$. Let $\overline{\mu}_\delta = \mu_\delta/\mu_\delta(\{0,1\}^\infty)$ be a computable probability measure over $\{0,1\}^\infty$. Thus, due to Lemma 19 and Proposition 18, there is a $d \in \mathbb{N}$ with $\psi(\alpha) = \delta^{-1}(J(\phi(\alpha)))$ and

$$\begin{aligned}
2^{-n-c\mathbf{K}(n)} &> \Gamma\{\alpha : \mathbf{t}_\mu(J(\phi(\alpha)) > 2^n/\mu(X)\} \\
&> \Gamma\{\beta : \log \mathbf{t}_{\mu_\delta}(\psi(\alpha)) > 2^{n-d}/\mu(X)\} \\
&> \Gamma\{\beta : \mathbf{D}(\psi(\alpha)|\mu_\delta) > n + d - \log \mu(X)\} \\
&> \Gamma\{\beta : \mathbf{D}(\psi(\alpha)|\overline{\mu}_\delta) > n + d\}.
\end{aligned}$$

Let $W \in \{0,1\}^{\infty^{2^{n+c\mathbf{K}(n)-1}}}$ be a set of $2^{n+c\mathbf{K}(n)-1}$ infinite sequences with each sequence chosen independently according to the uniform distribution over $\{0,1\}^\infty$. The probability that all $\alpha \in W$

Figure 23.6: A graphical depiction of the $\mathbf{T}_\kappa$ measure over the surphace of the sphere. As a point travels across its surphace, its $\mathbf{T}_\kappa$ score will oscillate.

has $\mathbf{D}(\psi(\alpha)|\overline{\mu}_\delta) \leq n + d$ is

$$(1 - 2^{-n-c\mathbf{K}(n)})^{2^{n+c\mathbf{K}(n)-1}} \geq (1 - 2^{n+c\mathbf{K}(n)-1}2^{-n-c\mathbf{K}(n)}) \geq 1/2.$$

Let $\mathcal{U}$ be a distribution over $\{0,1\}^\infty$ that is the uniform measure applied independently to $2^{n+c\mathbf{K}(n)-1}$ encoded sequences. And let $\gamma \in \text{Rep}(C)$ minimize $\lceil \mathbf{I}(C : \mathcal{H}) \rceil + 1$ and $\lambda = \langle p \rangle \gamma$, where p is a program to compute $\mathcal{U}$. By Theorem 74,

$$\Pr_{\alpha \sim \mathcal{U}}[\mathbf{I}(\langle \alpha, \lambda \rangle : \mathcal{H}) > m] \stackrel{*}{<} 2^{-m+\mathbf{I}(\lambda:\mathcal{H})}.$$

Therefore by probabilistic arguments, there exist a set $W \in \{0,1\}^{\infty 2^{n+c\mathbf{K}(n)}}$ such that for all $\alpha \in W$,

$$\mathbf{D}(\psi(\alpha)|\overline{\mu}_\delta) \leq n + d \text{ and } \mathbf{I}(\psi(W) : \mathcal{H}) <^+ \mathbf{I}(W : \mathcal{H}) <^+ \mathbf{I}(\langle W, \lambda \rangle : \mathcal{H} <^+ \mathbf{I}(\lambda : \mathcal{H}) =^+ \mathbf{I}(C : \mathcal{H}).$$

Thus Lemma 23 applied to $\psi(W)$ and $\overline{\mu}_\delta$, results in

$$\begin{aligned}
\log |W| &< \max_{\beta \in \psi(W)} \mathbf{D}(\beta|\overline{\mu}_\delta) + 2\mathbf{I}((W : \mathcal{H}) + O(\mathbf{K}(|W|)). \\
&< \max_{\beta \in \psi(W)} \mathbf{D}(\beta|\overline{\mu}_\delta) + 2\mathbf{I}((C : \mathcal{H}) + O(\mathbf{K}(|W|)). \\
n + c\mathbf{K}(n) &< n + d + 2\mathbf{I}(C : \mathcal{H}) + O(\mathbf{K}(n)). \\
c\mathbf{K}(n) &< d + 2\mathbf{I}(C : \mathcal{H}) + O(\mathbf{K}(n)).
\end{aligned}$$

Thus for proper choice of c, $\mathbf{I}(C : \mathcal{H}) = \infty$. □

23.3.3 Typicality on a Sphere's Surface

Theorem 107 proves a lower bound on the $\mathbf{T}_\kappa$ score across points of a sphere's surface. A graphical depiction of this phenomena can be seen in Figure 23.6.

Theorem 107 *Let κ be a computable system For sphere $C = (z, r)$ and uniform measure L_C over the surface of C, if $\mathbf{I}(C : \mathcal{H}) < \infty$ then there is a $c \in \mathbb{N}$ where for all $n \in \mathbb{N}$,*

$$2^{-n-c\mathbf{K}(n)} < L_C\{\alpha : \alpha \in C, \mathbf{T}_\kappa(\alpha) > 2^n\}.$$

Proof. Let $B = \{y : \|y - z\| < R\}$ where $r < R$, $R \in \mathbb{Q}_{>0}$. For Borel set A, let $\kappa'(A) = \kappa(A \cap B)$. Thus $\mathcal{R} = (\mathbb{R}^3, \|\cdot\|, \mathbb{Q}^3)$ is a computable metric space and $(\mathcal{R}, \kappa')$ is a computable measure space. In addition B is a compact open computable set with computable $\kappa(B)$. The application of Lemma 33 results in a $c \in \mathbb{N}$ where for all n,

$$2^{-n-c\mathbf{K}(n)} < L_C\{\alpha : \|\alpha - z\| = r, \mathbf{t}_{\kappa'}(\alpha) > 2^n\}$$

Thus using arguments in the proof of Theorem 103, $\mathbf{t}_{\kappa'} \stackrel{*}{=} \mathbf{T}_{\kappa'} \stackrel{*}{=} \mathbf{T}_{\kappa}$, for all $x \in B$, proving the theorem. □

23.3.4 Spheres Around Systems with Single Mass Points

If there is a system with a single mass point at origin and a sphere around $(0, 0, 0)$, then one can prove a stronger result than Theorem 107. That is, with the halting sequence requirement removed. In addition the bounds are improved.

Theorem 108 *Let κ be a system with a single mass point at origin $w = (0, 0, 0)$ and $C = (w, r)$ be a sphere with an uncomputable radius r. There is a $c \in \mathbb{N}$ where for all $n \in \mathbb{N}$,*

$$2^{-n-\mathbf{K}(n)-c} < L_C\{\alpha : \|\alpha\| = r, \mathbf{T}_{\kappa}(\alpha) > 2^n\},$$

where L_C is the uniform measure over the surface of the sphere C.

Proof. Let $r_1 < r < r_2$ with $r_1, r_2 \in \mathbb{Q}_{>0}$. Let $\mathcal{W}$ be the computable metric space consisting of the unit sphere, as introduced in Definition 61. We define the Borel measure κ' over $\mathcal{W}$, where $\kappa'(B(v, \epsilon)) = \int_{\mathcal{B}(v,\epsilon)} \kappa(x, y, z)dxdydz$, where $\mathcal{B}(v, \epsilon) = \{r'y : \cos^{-1}(\overrightarrow{wy} \cdot \overrightarrow{wx}) > \epsilon, \|\overrightarrow{wy}\| = \|\overrightarrow{wx}\| = 1, r_1 < r', < r_2\}$. Since κ is a radial measure $\kappa'(B(v, \epsilon)) \propto L_C(r \cdot B(v, \epsilon))$.

By Corollary 25, for every $n > \max\{-\log \kappa'(S), 0\}$, there exists an open set A_n, such that $-\log \kappa'(A_n) =^+ n + \mathbf{K}(n)$. Thus one can upper compute $\mathbf{K}(n)$ and enumerate an open A'_n consisting of a finite union of basis sets $B(v, \epsilon)$ such that $-\log \kappa'(A'_n) =^+ n + \mathbf{K}(n)$. Let $A''_n = \{x : x \in r' \cdot A'_n : r_1 < r' < r_2\}$. Thus $-\log \kappa(A''_n) =^+ n + \mathbf{K}(n)$. Thus for given (r_1, r_2), one can create the κ-test $t(x) = \sup_{n \in \mathbb{N}}[x \in A''_n]2^{n-O(1)}$. Thus for $x \in A''_n$, $\mathbf{T}_{\kappa}(x) \stackrel{*}{>} 2^n$. Furthermore $-\log L_C(r \cdot A'_n) =^+ -\log \kappa'(A'_n) =^+ n + \mathbf{K}(n)$, proving the theorem. □

Chapter 24

Chain Rule

In this chapter, a chain rule is proved analagously to the proof of Theorem 109 in Chapter 14. The original proof is in [Gác21]. The difference is this chapter does not deal with computer measure spaces and κ is an infinite measure.

Definition 68 (Two Point Tests) *Given the space $\mathcal{Y} = \mathbb{R}^3 \times \mathbb{R}^3$, a basis extended box $q \in Q$ with respect to $\mathcal{Y}$, are two rational boxs $q = (s_1, s_2)$, where $s_1, s_2 \in S$. A simple function $f : \mathcal{Y} \to \mathbb{R}_{\geq 0}$ is of the form $f_{s_1,s_2}(x, y) = v[x \in s_1, y \in s_2]$, with $v \in \mathbb{Q}_{>0}$. Simple functions can be enumerated producing the list $\{f_n\}$. A lower computable function F is of the form $F(x, y) = \sup_{n \in N} f_n(x, y)$ where N is an enumerable subset of $\mathbb{N}$. Given computable measure μ over $\{0, 1\}^\infty$ and system κ, a lower computable function F is a (κ, κ)-test, or $f \in [\kappa, \kappa]$ if*

$$(\kappa, \kappa)(F) = \int_{x_1 \in \mathbb{R}^3, x_2 \in \mathbb{R}^3} F(x_1, x_2), d\kappa(x_1) d\kappa(x_2) \leq 1.$$

Given κ and μ, the set $[\kappa, \kappa]$ is enumerable, so there exists a universal lower computable test $\mathbf{T}_{(\kappa,\kappa)}(x, y) = \sum_{f_i \in [\kappa,\kappa]} \mathbf{m}(i|\kappa) f_i(x, y)$.

Remark 11 (Relativized Typicality) *Given a system κ, a typicality score of a point $x \in \mathbb{R}^3$ relativized to $\alpha \in \{0, 1\}^{*\infty}$, is $\mathbf{T}_\kappa(x|\alpha)$, which is equal to $\mathbf{T}_\kappa(x)$, except the universal Turing machine has α on an auxilliary tape. This similarly holds for $\mathbf{T}_{\kappa,\kappa}(x, y|\alpha)$. When a point $x \in \mathbb{R}^3$ is in the conditional of $\mathbf{T}_\kappa$, then the universal Turing machine is given acces to some standard representation of the point.*

Definition 69 (Newtonian Complexity) *Given system κ, $x, y \in \mathbb{R}^3$, $z \in \mathbb{R}^3 \cup \{0, 1\}^{*\infty}$,*

- $\mathbf{K}_\kappa(x|z) = -\log \mathbf{T}_\kappa(x|z)$,
- $\mathbf{K}_{(\kappa,\kappa)}(x, y|z) = -\log \mathbf{T}_{\kappa,\kappa}(x, y|z)$.

Newtonian complexity can take arbitrary values in $\mathbb{R} \cup \{-\infty\}$. It is the measure of the entropy of a point in Newtonian space. Computable points will have $-\infty$ Newtonian complexity. Newtonian complexity cannot take values of ∞, because $\mathbf{T}_\kappa$ is positive over the whole space $\mathbb{R}^3$.

Remark 12 *In this chapter we prove the following equivalent equations. Let κ be a system.*

- $\mathbf{K}_{\kappa,\kappa}(x, y) =^+ \mathbf{K}_\kappa(x) + \mathbf{K}_\kappa(y|x, \lceil \mathbf{K}_\kappa(x) \rceil)$,
- $\mathbf{T}_{\kappa,\kappa}(x, y) \stackrel{*}{=} \mathbf{T}_\kappa(x)\mathbf{T}_\kappa(y|x, \lceil -\log \mathbf{T}_\kappa(x) \rceil)$.

Definition 70 *Note that we use $\kappa^w f(w) = \int_{(x,y,z) \in \mathbb{R}^3} f(x, y, z) d\kappa(x, y, z)$.*

24.1 Derivation

Proposition 39 $\mathbf{K}_\kappa(x) <^+ -\log \kappa^y 2^{-\mathbf{K}_{\kappa,\kappa}(x,y)}$.

Proof. Let $f(x) = -\log \kappa^y 2^{-\mathbf{K}_{\kappa,\kappa}(x,y)}$. The function f is upper computable and has $\kappa^x 2^{-f(x)} \leq 1$. Due to the universal properties of $\mathbf{T}_\kappa$ and thus minimum property of $\mathbf{K}_\kappa$, the inequality is proven. □

Proposition 40 *For a computable function $f : N^2 \to \mathbb{N}$,*

$$\mathbf{K}_\kappa(x|y) <^+ \mathbf{K}(z) + \mathbf{K}_\kappa(x|f(y,z)).$$

Proof. The function

$$g_\kappa(x,y) = \sum_z 2^{-\mathbf{K}_\kappa(x|f(y,z))-\mathbf{K}(z)},$$

is lower computable and $\kappa^x g_\kappa(x,y) \leq \sum_z 2^{-\mathbf{K}(z)} \leq 1$. So $g_\kappa(x,y) \stackrel{*}{<} 2^{-\mathbf{K}_\kappa(x|y)}$. The left hand side is a summation, so the inequality holds for each element of the sum, proving the proposition. □

Proposition 41 *If $i < j$, then*

$$i + \mathbf{K}_\kappa(x|i) <^+ j + \mathbf{K}_\kappa(x|j).$$

Proof. Using Proposition 40, with $f(i,n) = i + n$, we have

$$\mathbf{K}_\kappa(x|i) - \mathbf{K}_\kappa(x|j) <^+ \mathbf{K}(j-i) <^+ j - i.$$

Definition 71 (G-test) *Let the universal Turing machine be relativized to system κ. Let $G : \mathbb{R}^3 \to \mathbb{Z} \cap \{-\infty\}$ be an upper computable function. We recall that the Kolmogorov complexity of an upper computable function $f : \mathbb{R}^3 \to \mathbb{R} \cup \{\infty\}$ is $\mathbf{K}(f)$, the length of the shortest program to upper compute it. A G-test is an upper computable function g from $\mathbb{R}^3 \times \mathbb{R}^3$ to $\mathbb{R}_{\geq 0} \cup \{\infty\}$ such that $\kappa^y g(x,y) \leq 2^{-G(x)}$.*

Proposition 42 *Let the universal Turing machine be relativized to system κ. Let $G : \mathbb{R}^3 \to \mathbb{Z} \cap \{-\infty\}$ be an upper computable function with $\mathbf{K}(G) = O(1)$. There is a universal G-test g where for any other G-test h, for all $x \in \mathbb{R}^3$, $h(x,\cdot) \stackrel{*}{<} g(x,\cdot)$.*

Proof. The algorithm for g is as follows. Given x, it lower computes $2^{-G(x)}$, it also enumerates all lower computable $f(x,\cdot)$ and adds them to the weighted sum with coefficient $\mathbf{m}(t|\kappa,x)$ if $\kappa(f)$ is not greater than the current lower computed value of $2^{-G(x)}$. If a test t has $\kappa(t) \leq 2^{-G(x)}$, then eventually $\mathbf{m}(t|\kappa,x)t$ will be completely added to the weighted sum. Thus g is a universal test.

Proposition 43 *Let the universal Turing machine be relativized to system κ. Let $G : \mathbb{R}^3 \to \mathbb{Z} \cap \{-\infty\}$ be an upper computable function. By Proposition 42, among G-tests $g(x,y)$ there is a maximal G-test f within a multiplicative constant. For all y,*

$$f(x,y) \stackrel{*}{=} 2^{-G(x)} \mathbf{T}_\kappa(y|x, G(x)).$$

Proof. To prove the inequality $\stackrel{*}{>}$, let $g(x,y,m) = \max_{i \geq m} 2^{-i}\mathbf{T}_\kappa(y|x,i)$. This function is lower computable, and decreasing in m. Let $g(x,y) = g(x,y,G(x))$, which is lower semicomputable since G is upper semi-computable. The multiplicative form of Proposition 41 implies

$$g(x,y,m) \stackrel{*}{=} 2^{-m}\mathbf{T}_\kappa(y|x,m)$$
$$g(x,y) \stackrel{*}{=} 2^{-G(x)}\mathbf{T}_\kappa(y|x,G(x)).$$

Since $\mathbf{T}_\kappa$ is a test,

$$\kappa^y 2^{-m}\mathbf{T}_\kappa(y|x,m) \leq 2^{-m}$$
$$\kappa^y g(x,y) \stackrel{*}{<} 2^{-G(x)},$$

which implies

$$g(x,y) \stackrel{*}{<} f(x,y)/\mathbf{m}(g|x) \stackrel{*}{<} f(x,y)2^{\mathbf{K}(G)} \stackrel{*}{<} f(x,y)$$

by the optimality of $f(x,y)$. Note that if $G(x) = -\infty$, $g(x,y) = \infty$ for all y and thus $f(x,y) = 2^{-G(x)}\mathbf{T}_\kappa(y|x,G(x)) = \infty$. We now consider the upper bound. For fixed x, $2^{G(x)}f(x,y)$ is a κ-test, conditioned on x and $G(x)$. So

$$2^{G(x)}f(x,y) \stackrel{*}{<} \mathbf{T}_\kappa(y|x,G(x))/\mathbf{m}(f|x,G(x)) \stackrel{*}{<} \mathbf{T}_\kappa(y|x,G(x))2^{\mathbf{K}(G)} \stackrel{*}{<} \mathbf{T}_\kappa(y|x,G(x)).$$

□

Theorem 109

$$\mathbf{K}_{\kappa\times\kappa}(x,y) =^+ \mathbf{K}_\kappa(x) + \mathbf{K}_\kappa(y|x,\lceil\mathbf{K}_\kappa(x)\rceil).$$

Proof. Let $f(x,y) = 2^{-\mathbf{K}_{\kappa,\kappa}(x,y)}$. Proposition 39 implies $\kappa^y f(x,y) \stackrel{*}{<} 2^{-\mathbf{K}_\kappa(x)+c}$. Let $G(x) = \lceil\mathbf{K}_\kappa(x)\rceil + c$ for proper choice of $c \in \mathbb{N}$. Note that if h is a lower computable function such that $\kappa^y h(x,y) \stackrel{*}{<} 2^{-\mathbf{K}_\kappa(x)}$, then $\kappa^x\kappa^y h(x,y) \stackrel{*}{<} \kappa^x\mathbf{T}_\kappa(x) \stackrel{*}{<} 1$, so $h \stackrel{*}{<} f$, so f is a universal G-test. Proposition 28 (noting $\mathbf{K}(-\mathbf{K}_\kappa) = O(1)$) gives

$$\mathbf{K}_\kappa(x,y) = -\log f(x,y) =^+ G(x) + \mathbf{K}_\kappa(y|x,G(x)).$$
$$\mathbf{K}_\kappa(x,y) = -\log f(x,y) =^+ \mathbf{K}_\kappa(x) + \mathbf{K}_\kappa(y|x,\lceil\mathbf{K}_\kappa(x)\rceil).$$

□

Exercise 31 *Prove that for system κ, $x,y \in \mathbb{R}^3$, $\mathbf{T}_{\kappa,\kappa}(x,y) \stackrel{*}{>} \mathbf{T}_\kappa(x)\mathbf{T}_\kappa(y)$.*

Part V

Black Holes

Chapter 25

Kolmogorov Complexity of Black Holes

25.1 The CV Correspondence

In this section, Algorithmic Information Theory is applied to the study of the interior of black holes. The main references for this chapter are [BSZ17, Sus20, BS18]. By abstracting black holes as quantum circuits, researchers can study the complexity of black holes by solely investigating complexity theoretic aspects of $SU(n)$, and thus can skip to Section 25.2. This chapter leads to the following conclusioin

> *The study of the Kolmogorov complexity of black holes can be reduced to the study of a random fictious particle in the $SU(n)$ space.*

This chapter provides proofs to some of the claims in [BS18] and introduces a new continuous model that generalizes the discrete case.

In 1935, Einstein and Rosen published a paper describing a wormhole, or "Einstein Rosen Bridge" (ERB) as seen in Figure 25.1. This connects the parallel universes of regions 1 and 3 of the Penrose's Diagram, though prospective explorers should be discouraged, as one must travel faster than the speed of light to traverse it. However it is theoretically possible for two adventerers to jump in at either side and meet at the middle. In [MS13], the "ER=EPR" principle was introduced.

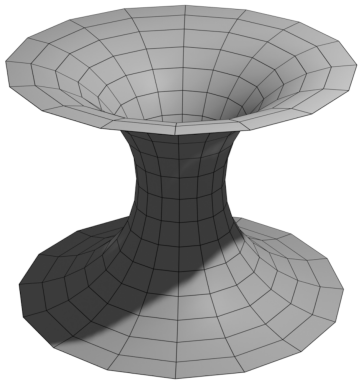


Figure 25.1: The Einstein Rosen Bridge.

This conjectures a link between the Einstein Rosen paper describing wormholes and the Einstein, Podolsky, Rosen paper describing entanglement. The principle states that entangled black holes share an ERB between them. This is one resolution to the AMPS Firewall paradox. The way in which two black holes can be entangled is when multiple EPR entangled pairs clump together and collapse into two black holes, as seen in Figure 25.2. The volume of ERBs increase over time. To

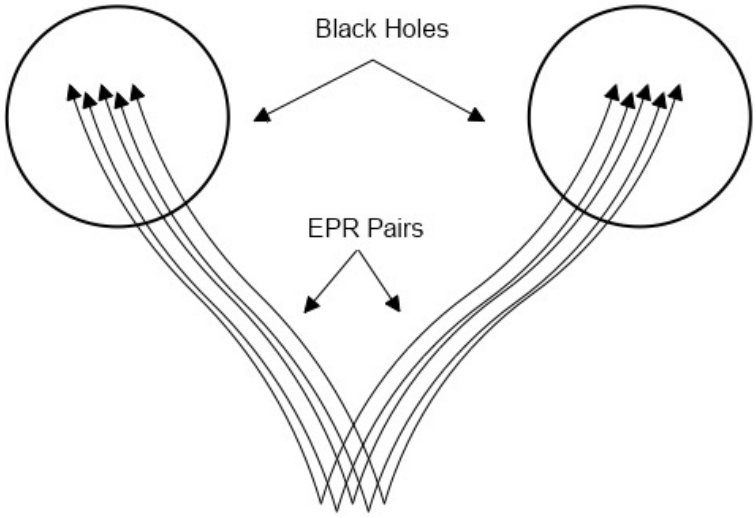


Figure 25.2: The formation of two entangled black holes. After a certain period of time, the mass of EPR particles will cause collapses into two black holes.

see this we look at the Penrose diagram of an Ads eternal black hole, or rather two entangled black holes connected by an ERB, as seen in Figure 25.4. In fact, the rate of of the volume growth is linear, with

$$\frac{dV(t)}{dt} \approx l_{ads} AT.$$

where t is the anchoring time, A is the horizon area, T is the black hole temperature, and l_{ads} is the AdS length scale. This can be seen in Figure 25.3. This follows directly from the AdS black hole metric tensor.

Exercise 32 *Prove, using the eternal AdS black hole metric, that the Einstein Rosen bridge will continue to grow linearly.*

According to classical generality the ERB will continue to grow forever. This is occurs for an exponential (in the number of qubits) amount of time. However at some point classical general relativity will break down due to the quantum reccurrence theorem. The quantum recurrence theorem states the non-integrable system with a finite density of states will be quasiperiodic with a recurrence time doubly exponential in the entropy, S. This applies to AdS black holes. One question is when does the expected value of the wormhole stop growing? In the black hole, there are only $\exp(S)$ number of mutually orthogonal states. Other states must be superpositions of the previous states, all which have sub-exponential wormhole length. Thus the expected value of the length of the wormhole must sharply stop growing at time $t \approx \exp S$. Thus the curve of the volume graph can be seen in Figure 25.5. The question answered is what other property of black holes is dual to this behavior? In [BSZ17, Sus20, BS18], this question was addressed. The answer cannot be the entropy, as thermalization occurs at a logarithmically shorter time span. As shown in the

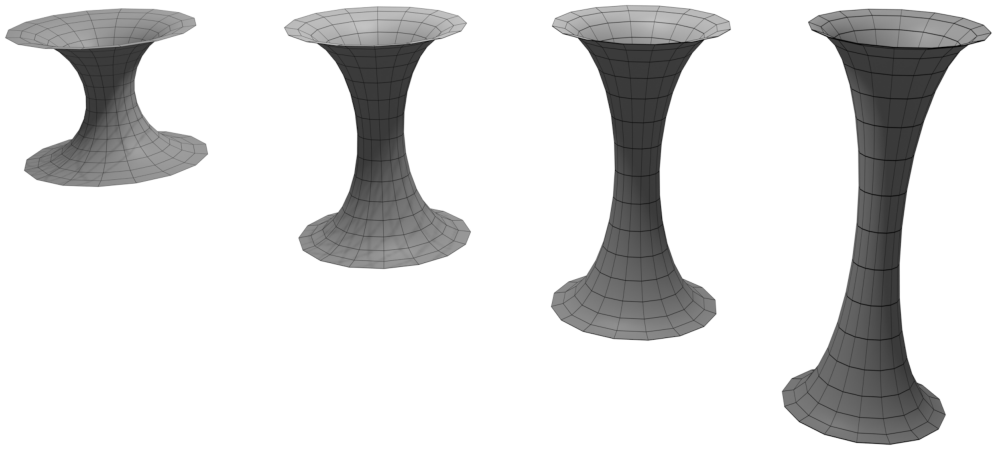


Figure 25.3: For an initial, exponential amount of time, the ERB will grow linearly. Eventually, classical general relativity will break down, and the ERB will reach a max volume.

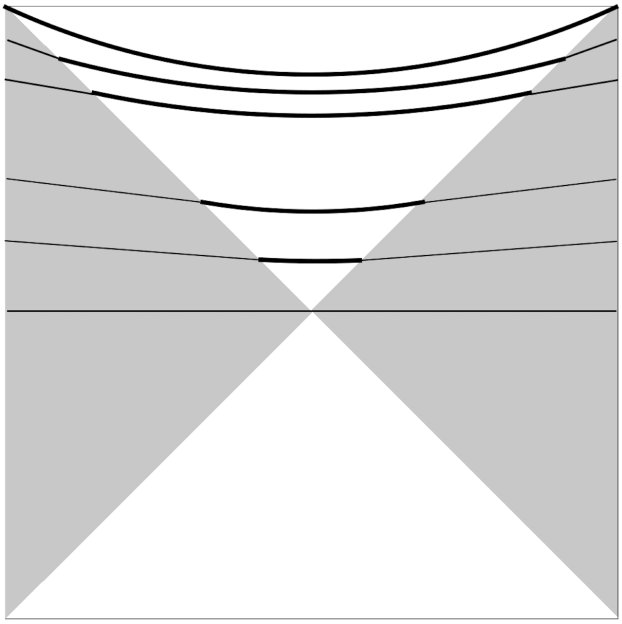


Figure 25.4: The linear time grow of an ERB in a Penrose diagram. Each horzontal line represents a time slice, with $t = 0$ being the center line where quadrants two and four connect. As time moves up, each slice has a longer volume between the two horizons. In classical general relativity, the limit is infinity at the singularity.

next section, one can define a notion of the complexity of the black hole, and this notion is has dual properties as the volume of an ERB. This is known in the literature as the *Complexity/Volume Correspondence*, CV Correspondence for short. In addition this is a statement of duality, and not causation. Recently, linear growth of the circuit complexity of quantum circuits was proven in [HFK+22], proving the linearity part of the CV Correspondence for circuit complexity. They do

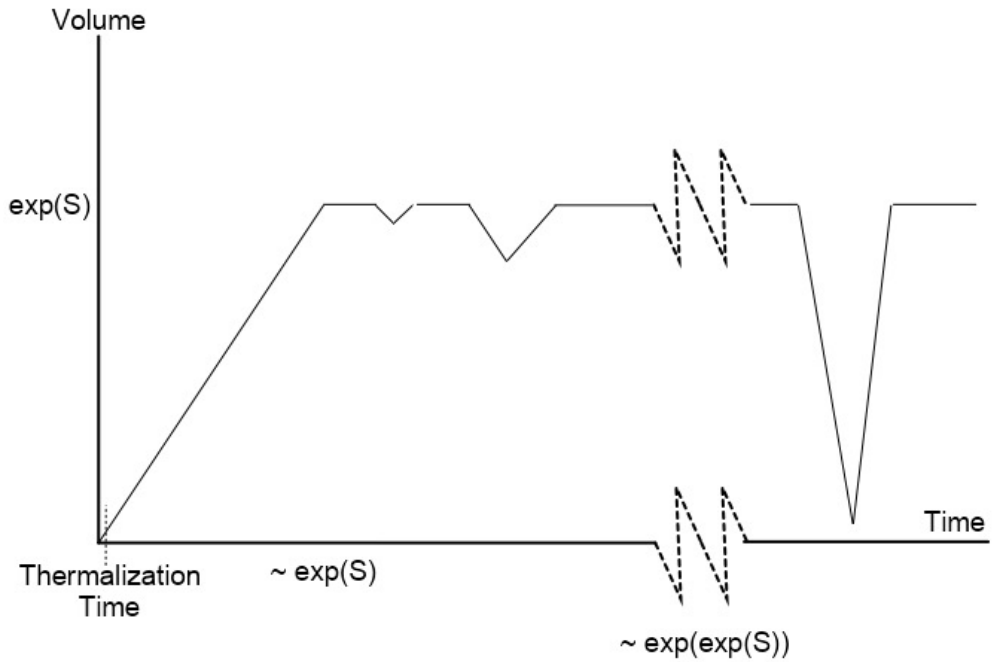


Figure 25.5: The volume vs. time graph of an ERB as seen in [BS18]. The growth continues linearly for an exponential amount of time until it eaches $\exp(S)$, where S is the entropy. Small pertubations occur as time continues and once every double exponential in S time, the volume is expected to dip to near its initial value. The thermalization of the black hole occurs at a much shorter time scale than the evolution of the volume of an ERB.

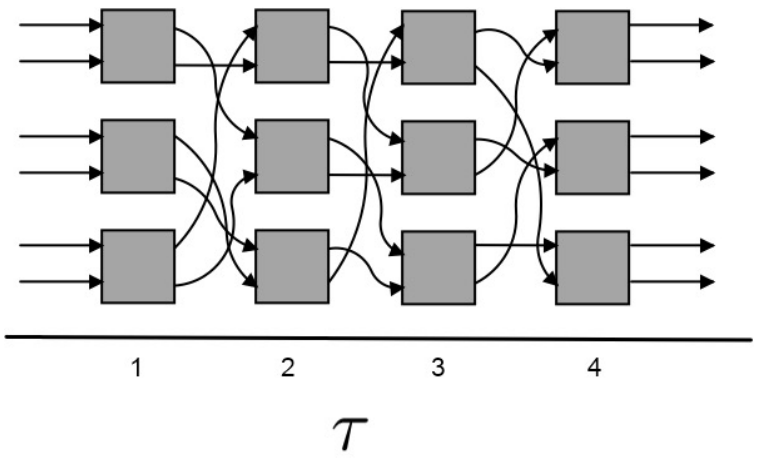


Figure 25.6: An example of a random circuit with $K = 6$, $k = 2$, and depth 4. At each round the qubits are randomly paired and sent through a random gate. τ is the Rindler time. This circuit is believed to be a fast scrambler.

not prove oscillations of complexity. The problem is still open for Kolmogorov complexity.

25.2 Black Hole Quantum Circuit Correspondence

Currently it is conjectured that black holes with entropty S can be represented by a random quantum circuits of $K \approx S$ qubits. This implies modelling black holes using a stochastic random model, i.e. with time dependent Hamiltonians. This is assuming the temperature is high enough such that every degree of freedom carries order one unit of entropy. In addition, whereas the black hole uses clock time, the quantum circuit uses Rindler time:

$$t_{\text{Rindler}} = 2\pi T t_{\text{Swarzschild}},$$

where T is the black hole temperature and $t_{\text{Swarzschild}}$ is the Swarzschild time from the perspective of an observer at infinity. Rindler time is preferred because it is dimensionless, like the time used in circuits in quantum complexity theory. Properties of black holes suggests that the quantum gates are not restricted by locality. In fact, the allowable interaction between the qubits is *k-local* and *all-to-all*, where $k \ll K$ is much less than the number of qubits. This is due to the fact that black holes are fast scramblers (see [SS08]). Fast scramblers mix up information so that any subsystem smaller than half the whole system has maximum entangled entropy. This is done in logarithm time. Random k-local quantum circuits are conjectured to be fast scramblers.

The random quantum circuit with $k = 2$ proceeds as follows. At each discrete time step, K qubits are randomly paired and each pair is sent through a randomly chosen gate. The specific gate set is not important, as long as it is universal. At the next step, the qubits are randomly regrouped, as shown in Figure 25.6.

This represents the discrete abstraction of black holes to quantum circuits. The CV Correspondence states that the behavior of complexity of these circuits over time is a dual to the volume growth of ERBs, that is they behave according to Figure 25.4. The next sections will be concerned with defining the Kolmogorov complexity of the circuits and giving support to its version of the CV Correspondence.

25.3 Kolmogorov Complexity of Circuits

In [BS18] two definitions of the complexity of circuits $\mathcal{C}$ were introduced, one involving quantum circuit complexity $\mathbf{K}_{\mathcal{Q}}(\mathcal{C})$ and the other involving Kolmogorov complexity, $\mathbf{K}(\mathcal{C})$. Complexity can be defined on states, but it is more convenient to define the complexity of quantum circuits in $SU(2^K)$. In this manuscript, we will focus solely on Kolmogorov complexity. One naive way to define $\mathbf{K}(\mathcal{C})$ is the minimal size of the program that will print out the circuit. However, this means that quantum circuits can have unbounded complexity. Instead, the desired approached is to give the same complexity to circuits that behave similarly, i.e. coarse graining the $SU(2^K)$ space.

First we start with the inner product metric between two unitary matrices $U, V \in SU(2^K)$,

$$d(U,V) = \arccos\left(\mathrm{Tr} U^* V\right) \tag{25.1}$$

where Tr is the normalized trace function. Since $\max_{U,V\in SU(2^K)} d(U,V) = \pi/2$, $SU(2^K)$ is a compact metric space. We then partition $SU(2^K)$ space into $\epsilon \in \mathbb{R}_{>0}$ ball of the same dimension $(2^{2K}-1)$. From [Sus20, BST03], we get that

$$\begin{aligned} \#\text{ of the number of unitaries } &\approx \left(\frac{2^K}{\epsilon^2}\right)^{4^K/2} \\ \log \#\text{ of the number of unitaries } &\approx \frac{4^K}{2} K \log 2 + 4^K \log \frac{1}{\epsilon}. \end{aligned} \tag{25.2}$$

Thus for small enough fixed ϵ and large enough K in the limit we can assume the number of unitaries is $\approx e^{4^K}$. We provide two definitions for the complexity of a unitary operator. In [Sus20, BS18, BSZ17], emphasis was put towards studying the circuit complexity, whereas this chapter will study the Kolmorogov complexity. The benefits of using Algorithmic Information Theory is that whether one chooses finite strings, infinite sequences, or computable metric spaces, as shown in this manuscript, oscillations of Kolmogorov complexity, randomness deficiency, or algorithmic entropy is guaranteed to occur in dynamics. The hard part is to show the linear increase of the expected value at the beginning of the dynamics. This was recently proven in the quantum circuit complexity case [HFK+22].

Definition 72 (Kolmogorov Complexity of a Unitary Operator) *We define an algorithm A that takes in K and ϵ as parameters and labels each epsilon ball of $SU(2^K)$ with a unique string of length 4^K. Furthermore we assume A is the simpliest program to perform this task. The Kolmogorov complexity of a unitary operator U is equal to the Kolmogorov complexity $\mathbf{K}(x)$ of the label x of the epsilon ball which U is in. The universal Turing machine is relativized to K. One may argue that the labelling is arbitrary, but given labeling algorithm B, $|\mathbf{K}^A(U) - \mathbf{K}^B(U)| <^+ \mathbf{K}(A,B)$. Thus all simple labellng algorithms will provide the approximately the same complexity. Furthermore the node corresponding to the epsilon ball that has the identity matrix $\mathbf{1}$ will have Kolmogorov complexity $O(1)$.*

Definition 73 (Circuit Complexity of a Unitary Operator) *Let G be the allowable gate set, that is k-local, all-to-all circuits. The circuit complexity of a unitary operator, $\mathbf{K}_{\mathcal{Q}}(U)$ is the minimum size of a sequence of gates $g_1, g_2, \ldots, g_n \in G$ such that $g_1 g_2 \ldots g_n U = \mathbf{1}$, where $\mathbf{1}$ is the identity operator.*

25.4 Graph Interpretation

The discretation of the $SU(2^K)$ space can be modeled with a undirected graph, termed $G(K)$. The dynamics of the black hole is approximated by a random walk on this graph. Each node corresponds

1. # of vertices is $\approx e^{4^K}$	Fact
2. $G(K)$ is regular and sparse with degree $\approx \left(\frac{2K}{e}\right)^{\frac{K}{2}}$	Fact
3. $G(K)$ is vertex transitive	Fact
4. The diameter of $G(K)$ is logarithmic in the number vertices	Conjecture
5. Loops of length less than 4^K are rare or absent.	Conjecture
6. $G(K)$ is an expander graph	Conjecture

Figure 25.7: The proven and conjectured properties of $G(K)$ representing unitary operators where the evolution of the quantum circuit is equivalent to a random walker.

to an epilson ball and its label is equal to the A-label of the epsilon ball. Two nodes share an edge if there is a permissible gate $g \in G$ that connects the epsilon balls. The graph $G(K)$ is undirected, implying if $g \in G$, then so is $g^* \in G$. Since the SU group space is homogeneous, $G(K)$ is vertex transitive. At each vertex, the number of possible choices in the discrete evolutationary step is the same, equalling

$$d \approx \frac{K!}{(K/2)!} \approx \left(\frac{2K}{e}\right)^{\frac{K}{2}}$$

Since the number of vertices of $G(K)$ is $\approx e^{4^K}$, the graph is sparse. Furthermore, assuming minimal collisions, the number of unitaries reached after D steps is conjectured to be

$$\text{Number of unitaries after } d \text{ steps } = d^D \approx \left(\frac{2K}{e}\right)^{DK/2}. \tag{25.3}$$

Thus approximately, the diameter is not larger than the number of steps it takes for all unitaries reached, so combining Equations 25.2 and 25.3.

$$\left(\frac{2K}{e}\right)^{\text{Diameter}/2} \leq \left(\frac{2^K}{\epsilon^2}\right)^{4^K/2}$$

$$\text{Diameter} \leq 4^K \left(2 + 3\frac{|\log \epsilon|}{\log K}\right).$$

Thus assuming minimal collisions, the diameter is logarithmic in the number of vertices of $G(K)$. All the properties described or conjectured implies that $G(K)$ is an expander graph. To recap, Figure 25.7 shows proved and conjectured properties about $G(K)$.

Assuming scarcity of small loops, locally at each vertex, the graph $G(K)$ looks like the exponential expansion shown in Figure 25.8. Thus a random walk on $G(K)$ corresponds to random motion in the $SU(2^K)$ space, as shown in Figure 25.9. Assume that all the conjectured properties of $G(K)$ hold. At each vertex, there are d choices for the random walk. Therefore, using the assumption that there are virtually no loops less than 4^K, the expected Kolmogorov complexity of the vertex will increase by $\log d \approx K \log K$ at each step until the max of 4^K is reached. Because we assume we are working with an regular expander graph, the distribution over the vertices from the random walk will converge to the uniform distribution in logarithmic time, that is 4^K steps. We will show this property in the next section. At this point, with a uniform stationary distribution, the Kolmogorov complexity of the random walker will oscillate, with a dip of n points of complexity will occur every e^{4^K} expected steps. Thus Figure 25.10 shows the behavior of the complexity of the system. The similarity of Figures 25.5 and 25.10 represent the Complexity-Volume Correspondence.

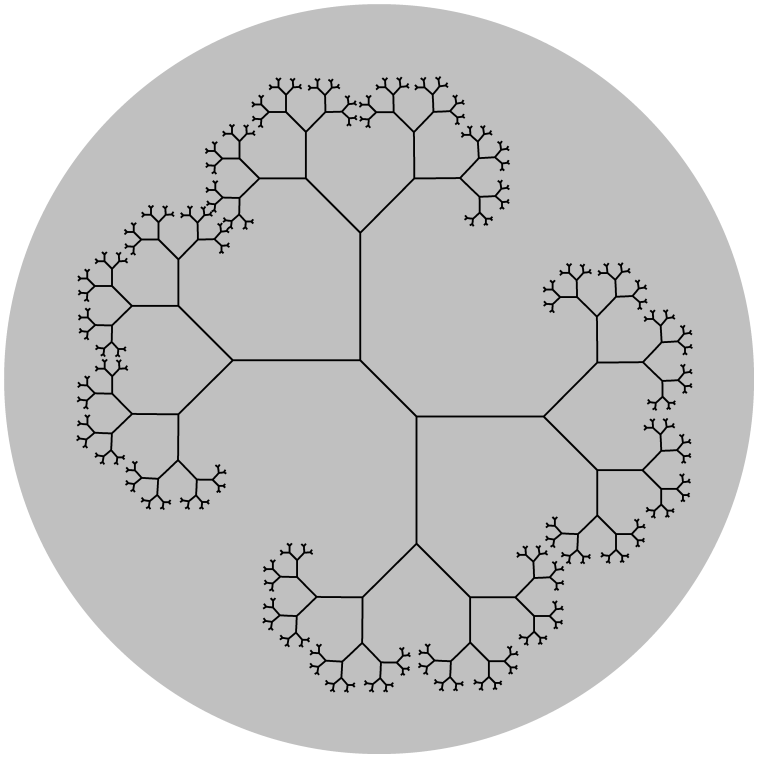


Figure 25.8: If $G(K)$ has only minimal loops of subexponential size, then locally at each vertex, $G(K)$ looks like the following graph, where there is an exponential expansion.

25.5 Sufficient Conditions for Linear Growth

In this section, we provide two sufficient conditions for linear growth the Kolmogorov complexity of black holes, namely, expander graphs and exponentially small return probabilities. When we use the graph $G(K)$, we assume the universal Turing machine U is relativized to K and thus by extension, $G(K)$. Let $M(K)$ be the random walk matrix of $G(K)$, where vertex i goes to vertex j with probability $M(K)_{i,j}$. We assume $G(K)$ has 2^n vertices.

Definition 74 (Expected Complexity) *Given probability π over the vertices of $G(K)$, its expected complexity value is $\mathbf{K}(\pi) = \sum_{\text{Vertex } v \in G(K)} \pi(v)\mathbf{K}(v)$.*

Let $d(K)$ be the vertex degree of $G(K)$. For probability π over $G(K)$, π^t is used to denote πM^t.

Proposition 44 *For any distribution π over $G(K)$, $\mathbf{K}(\pi) <^+ \mathbf{K}(\pi^1) <^+ \mathbf{K}(\pi) + \log d(K)$.*

Proof. The second inequality follows from the fact that if vertices v and w are connected by an edge then $\mathbf{K}(v) <^+ \mathbf{K}(w) + \log d(K)$. For the first inequality, we assume π is concentrated on a single vertex x, and the general case follows simply. Assume $\mathbf{K}(\pi^1) < \mathbf{K}(\pi) - b$, for some constant b to be determined later. Let x_i be the neighbors of x, for $i \in \{1, \ldots, d\}$. Thus $\frac{1}{d(K)} \sum_{i=1}^{d(K)} \mathbf{K}(x_i) \leq \mathbf{K}(x) - b$. So $\mathbf{m}(x) \stackrel{*}{>} \sum_{i=1}^{d(K)} 2^{-\mathbf{K}(x_i) - \log d(K)} \stackrel{*}{=} \mathbf{E}_{i \sim \text{Uniform}}[2^{-\mathbf{K}(x_i)}] \stackrel{*}{>} 2^{-\mathbf{E}_{i \sim \text{Uniform}}[\mathbf{K}(x_i)]} \stackrel{*}{>} 2^{-\mathbf{K}(x)+b}$. This is a contradiction for large enough b. □

The following corollary generalizes Proposition 44.

Corollary 46 *For any distribution π over $G(K)$, $\mathbf{K}(\pi) <^+ \mathbf{K}(\pi^t) + \mathbf{K}(t)$.*

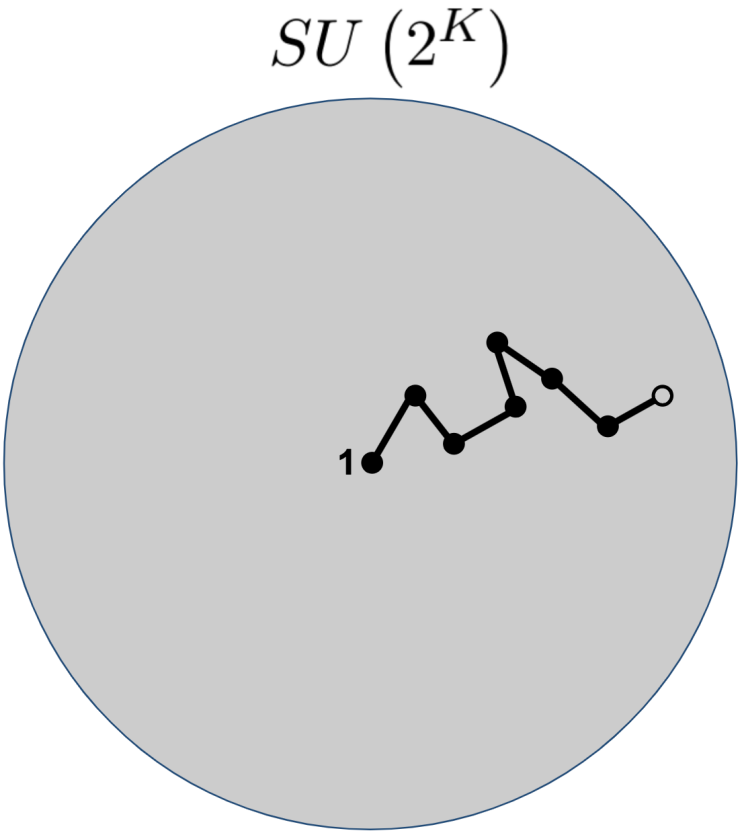


Figure 25.9: The random walk along the approximating graph $G(K)$ corresponds to the dynamics of an eternal AdS black hole in $SU(2^K)$ space. The random walk starts at the identity matrix $\mathbf{1}$.

Proof. We assume π is concentrated on a single vertex x, and the general case follows simply. Assume $\mathbf{K}(\pi^t) < \mathbf{K}(\pi) - b - \mathbf{K}(t)$, for some constant b to be determined later. Let p_i be the probability x travels to node x_i in a random walk of size t. This is also equal to the probability that x_i travels to x in t random steps. Thus $\sum_i p_i \mathbf{K}(x_i) \leq \mathbf{K}(x) - b - \mathbf{K}(t)$. So $\mathbf{m}(x) \overset{*}{>} \sum_i 2^{-\mathbf{K}(x_i) - \log p_i - \mathbf{K}(t)} \overset{*}{=} 2^{-\mathbf{K}(t)} \mathbf{E}_{i\sim p_i}[2^{-\mathbf{K}(x_i)}] \overset{*}{>} 2^{-\mathbf{K}(t)} 2^{-\mathbf{E}_{i\sim p_i}[\mathbf{K}(x_i)]} \overset{*}{>} 2^{-\mathbf{K}(x)+b}$. This is a contradiction for large enough b. □

25.5.1 Expander Assumption

In this section, we assume $G(K)$ is an expander graph, that is we assume properties 1, 2, 3, and 6 of Figure 25.7. Let $u(K)$ be the uniform distribution over the vertices of $G(K)$. We define the following expander property of graphs.

Assumption 1 (Expander Graphs) *Let $\lambda(K)$ be the second largest eigenvalue of $M(K)$. Another equivalent definition is*

$$\lambda(K) = \max_{\pi} \frac{\|\pi M(K) - u(K)\|}{\|\pi - u(K)\|} = \max_{x \perp u(K)} \frac{\|xM(K)\|}{\|x\|}.$$

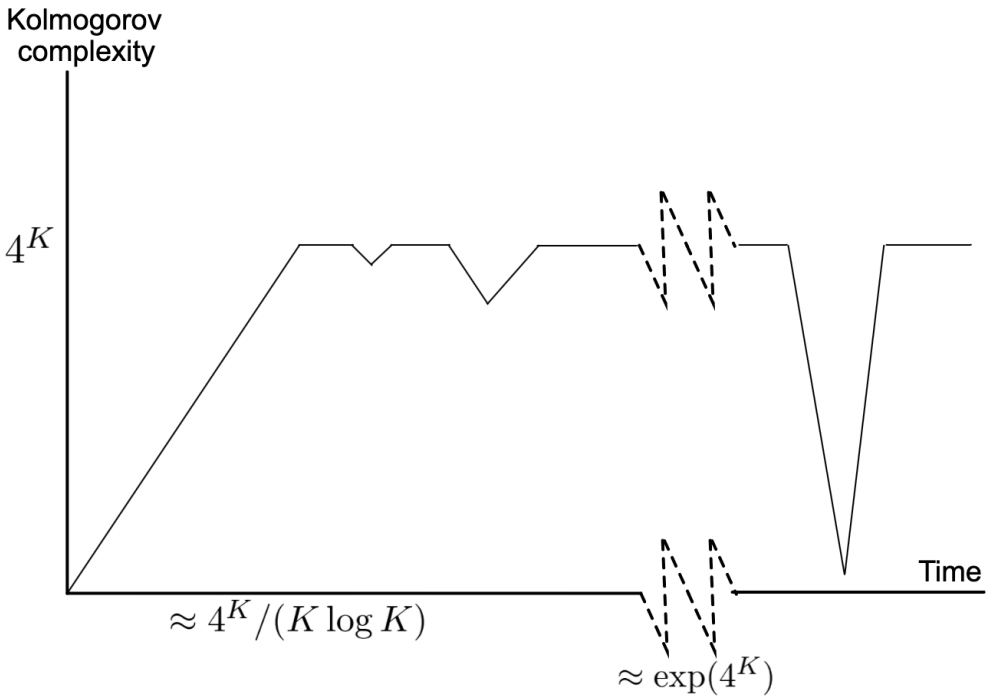


Figure 25.10: The time evolution of the Kolmogorov complexity of a quantum circuit modelling the black hole. The complexity/time graph parallels the evolution of the volume growth of the ERB, as shown in Figure 25.5.

Thus the maximum is over all probability π over the vertices of $G(K)$. The term $x \perp u(K)$ means the sum of the components of x is 0. By definition $\lambda(K) \in [0,1]$. The lower the $\lambda(K)$ score, the higher the expanding qualities of $G(K)$. In this section, it is assumed that the expansion property of the graphs is $\sup_k \lambda(K) = \lambda < 1$.

One useful property of expanders is that random walks will converge to the stationary probability in logarithmic time, as shown by the following well known theorem.

Theorem 110 *For every initial distribution π over the vertices of $G(K)$,*

$$|\pi^t - u(K)|_\infty \leq \lambda^t.$$

Definition 75 *The entropy of a distribution π over $G(K)$ is $\mathcal{H}(\pi) = -\sum_{i \in G(K)} \pi_i \log \pi_i$.*

Proposition 45 ([GŹ21]) $\mathbf{K}(\pi) \geq \mathcal{H}(\pi)$.

Proof. $\mathcal{H}$ is the size of the smallest π-expected length of a prefix free encoding of the vertices of $G(K)$. Since $\mathbf{K}$ is a prefix free code, it majorizes $\mathcal{H}$.

We assume a fixed K and remove its notation, and assume $|G| = 2^n$.

Theorem 111 *Given Assumption 1, for probability π^t concentrated on the node in G with the identity matrix, for $0 < t < n/(\log(1/\lambda))$, $t \log(1/\lambda) - 1 \leq \mathbf{K}(\pi^t) < t(\log d + O(1))$.*

Proof. The upper bound is from Proposition 44. For the lower bound, let $p_{\text{Max}} = \max_i p_i$, where p is a probability over G. It is easy to see that $\mathcal{H}(p) \geq p_{\text{Max}}$. Using Proposition 45 and Theorem 110, one gets for $0 < t \leq n/(\log/\lambda)$,

$$\mathbf{K}(\pi^t) \geq \mathcal{H}(\mathbf{p}^t) \geq -\log \pi_{\text{Max}}^t \geq -\log(\lambda^t + 2^{-n}),$$

and the theorem follows directly.

25.5.2 Return Probability Assumption

The graphs $G(K)$ are vertex transitive graphs. We fix a K and remove its notation, and let $|G| = 2^n$. Let $\psi(v, w, t)$ be the probability that a node v is on node w after a random walk of t steps.

Assumption 2 (Exponential Return Probability) *In this section we assume that there is a $\lambda \in (0, 1)$, such that for all random walks of G of size $0 < 1 < n/(\log(1/\lambda))$, $p(v, v, 2t) < \lambda^t$.*

Theorem 112 *Given Assumption 2, for probability π concentrated on the node in G with the identity matrix, for $0 < t < n/(\log(1/\lambda))$, we have that $t \log(1/\lambda) \leq \mathbf{K}(\pi^{2t}) < 2t(\log d + O(1))$.*

Proof. The upper bound is from Proposition 44. For the lower bound, due to [AS04], since G is vertex transitive, we have that $p(u, u, 2t) \geq p(u, w, 2t)$ for $v, w \in G$, $v \neq w$. Thus, using Proposition 45,

$$\mathbf{K}(\pi^{2t}) \geq \mathcal{H}(\pi^{2t}) \geq -\log p(u, u, 2t) \geq t \log(1/\lambda).$$

25.6 Weighted Random Walker

Assume that the stochastic random model is much more chaotic, and the dynamics are not modelled as a random walk on $G(K)$, but a walk along a recurrent markov model $M(K)$. In this section, we investigate what properties can be proven. If one assumes $M(K)$ has expander properties, then by Section 25.5.1, the expected complexity reaches the maximum in time $\approx 4^K$.

The following proposition shows that oscilliations in complexity have to occur, though it might be more frequent than the $G(K)$ case. Let p be the computable ergodic measure corresponding to the weighted walk on $M(K)$, and for simplicity suppose there are 2^n vertices. Each vertex is assigned a string $x \in \{0, 1\}^n$ and the complexity of the vertex is the Kolmogorov complexity of the string, $\mathbf{K}(x)$. We first prove the simple bounds and then prove the more complex tighter bounds.

25.6.1 Simple Bounds

Proposition 46 *There is a c where for probability p over $\{0, 1\}^n$, for all $m > \mathbf{K}(p) + c$, $p\{x : \mathbf{K}(x) < m\} > 2^{m - 2\mathbf{K}(m,p) - n - c}$.*

Proof. Order strings x of size n by $p(x)$ value, with largest values first, and breaking ties through any simple ordering on $\{0, 1\}^n$. It must be the first 2^ℓ strings X has $p(X) \geq 2^{\ell - n - 1}$ Otherwise the average value of $p(x)$, $x \in X$, is less than 2^{-n-1}. Thus for the remaining $2^n - 2^\ell$ strings Y, $p(y) < 2^{-n-1}$, So

$$\begin{aligned} p(\{0, 1\}^n) &= p(X) + P(Y) \\ &< 2^{\ell - n - 1} + (2^n - 2^\ell)(2^{-n-1}) \\ &= 2^{\ell - n - 1} + 2^{-1} - 2^{\ell - n - 1} \\ &= 1/2, \end{aligned}$$

which is a contradiction. Furthermore, the first 2^ℓ elements x have complexity $\mathbf{K}(x|p) <^+ \ell + \mathbf{K}(\ell)$ or $\mathbf{K}(x) <^+ \mathbf{K}(p, \ell) + \ell$. Let $m = \ell + \mathbf{K}(\ell, p) + O(1)$. By Proposition 48, $m - 2\mathbf{K}(m, p) <^+ \ell$. □

Proposition 47 *For every $c, n \in \mathbb{N}$, there exists $c' \in \mathbb{N}$ where for all $a, b \in \mathbb{N}$, if $a < b + n \log a + c$ then $a < b + 2n \log b + c'$.*

Proof.

$$\begin{aligned} \log a &< \log b + \log\log a + \log cn \\ 2\log a - 2\log\log a &< 2\log b + 2\log cn \\ \log a &< 2\log b + 2\log dn. \end{aligned}$$

Combining with the original inequality

$$\begin{aligned} a &< b + n\log a + c \\ a &< b + n(2\log b + 2\log dn) + c \\ &= y + 2n\log y + c', \end{aligned}$$

where $c' = 2n \log cn + c$. □

Proposition 48 *For all $d \in \mathbb{N}$ there is a $d' \in \mathbb{N}$ where if $x + \mathbf{K}(x, z) + d > y$ then $x + d' > y - 2\mathbf{K}(y, z)$.*

Proof. If $x + d > y$, then the lemma is satisfied, so $x + f \leq d$. Thus $y - x < \mathbf{K}(x, z) + d$ implies $\mathbf{K}(y - x) <^+ 2\log \mathbf{K}(x, z) + 2\log d$. Thus $\mathbf{K}(x, z) <^+ \mathbf{K}(y, z) + \mathbf{K}(y - x) <^+ \mathbf{K}(y, z) + 2\log \mathbf{K}(x, z) + 2\log d$. Applying Proposition 47, where $a = (x, z)$, $b = (y, z)$ and $c = 2\log d + O(1)$ and $n = 2$, we get a c' dependent on c and n where $\mathbf{K}(x, z) < \mathbf{K}(y, z) + 4\log \mathbf{K}(y, z) + c' < 2\mathbf{K}(y, z) + c' + O(1)$. So

$$\begin{aligned} x + \mathbf{K}(x, z) + d &> y \\ x + (2\mathbf{K}(y, z) + d' + O(1)) + d &> y \\ x + d'' &> y - 2\mathbf{K}(y, z), \end{aligned}$$

where $d'' = d' + O(1) + d$. □

25.6.2 Tighter Bounds

Theorem 113 *There is a $c \in \mathbb{N}$ where for probability p over $\{0, 1\}^n$, for $m > \mathbf{K}(p) + c$, $p\{x : \mathbf{K}(x) < m\} > 2^{m - n - 2\mathbf{I}(p;\mathcal{H}) - O(\mathbf{K}(n,m))}$.*

Proof. Without loss of generality, p can be assumed to have a range in powers of 2. Assume not, then there exist $\ell \in (\mathbf{K}(p) + c, n)$ such that $p\{x : \mathbf{K}(x) \leq \ell\} < 2^{-k}$, where $k = n - \ell - c - 2\mathbf{I}(p; \mathcal{H}) - O(\mathbf{K}(n, \ell))$ and c solely depends on the universal Turing machine. $\mathbf{K}(k) <^+ \mathbf{K}(n, \ell, c, \mathbf{I}(p; \mathcal{H}), \mathbf{K}(n, \ell))$. Suppose $\max\{p(x) : \mathbf{K}(x) > \ell\} \geq 2^{-k}$. Then

$$\mathbf{K}(p) + O(1) > \mathbf{K}\left(\arg\max_x p(x)\right) > \ell > \mathbf{K}(p) + c,$$

causing a contradiction, for choice of c dependent on U. Sample 2^{k-2} elements D without replacement according to p. p^* is the probability of D, where $\mathbf{K}(p^*) <^+ \mathbf{K}(p, n, \ell, \mathbf{K}(n, \ell), c, \mathbf{I}(p; \mathcal{H}))$. Even if every element x chosen has $p(x) = 2^{-k-1}$, the total p mass sampled is not greater than

$$2^{k-1}2^{k-2} \leq 2^{-3}.$$

The probability q that all $x \in D$ has $\mathbf{K}(x) > \ell$ is

$$q > \left(1 - 2^{-k}/(1 - 2^{-3} + 2^{-k})\right)^{2^{k-2}} > \left(1 - 2^{k+1}\right)^{2^{k-2}} = 1/2.$$

Thus, by Lemmas 2 and 1,

$$\Pr_{S \sim p^*}[\mathbf{I}(S; \mathcal{H}) > \mathbf{I}(p^*; \mathcal{H}) + m] \overset{*}{<} 2^{-m},$$

$$\Pr_{S \sim p^*}[\mathbf{I}(S; \mathcal{H}) > \mathbf{I}((p, n, \ell, \mathbf{K}(n, \ell), \mathbf{I}(p; \mathcal{H})); \mathcal{H}) + m] \overset{*}{<} 2^{-m}.$$

So by probabilistic arguments, there exists $D \subset \{0,1\}^n$, where for all $x \in D$, $\mathbf{K}(x) > \ell$ and

$$\mathbf{I}(D; \mathcal{H}) <^+ \mathbf{I}(p^*; \mathcal{H}) <^+ \mathbf{I}((p, c, n, \ell, \mathbf{K}(n, \ell), \mathbf{I}(p; \mathcal{H})); \mathcal{H}) <^+ \mathbf{I}(p; \mathcal{H}) + \mathbf{K}(\ell, n, \mathbf{I}(p; \mathcal{H}), c).$$

So by Lemma 21, applied to D and the uniform measure U_n over strings of length n,

$$\begin{aligned} k <& \max_{a \in D} \mathbf{d}(a|U_n) + \mathbf{I}(D; \mathcal{H}) + O(\mathbf{K}(\mathbf{I}(D; \mathcal{H}), k, U_n)) \\ n - \ell + c + O(\mathbf{K}(\ell, n)) + 2\mathbf{I}(p; \mathcal{H}) <& n - \ell + \mathbf{K}(n) + \mathbf{I}(p; \mathcal{H}) + O(\mathbf{K}(n, \ell, c, \mathbf{I}(p; \mathcal{H}))) \\ c <& O(\mathbf{K}(c)). \end{aligned}$$

which is a contradiction for large enough c dependent solely on the universal Turing machine U. □

25.7 Continuous Model

In this section, we examine the feasibility a continuous version to the discrete random walk on $G(K)$. The time is still discrete and the evolution is still k-local and all-to-all, but at each time step, each gate is a random 2^k unitary operator instead of being chosen from a finite gate set. Thus the state space is all of $SU(2^K)$ instead of an approximating graph.

To this end, let $\mathcal{X}(K) = (SU(2^K), d_K, Q(K))$ be a sequence of parameterized computable metric spaces. As stated before, $SU(2^K)$ is the set of all K-qubit unitary operators. The distance function $d_K : SU(2^K) \times SU(2^K) \to [0, \pi/2]$ is the inner product distance defined in Equation 25.1. The set of ideal points $Q(K)$ is all K-qubit operators with rational coefficients. The volume measure μ_K is the Haar measure over $SU(2^K)$ multiplied by e^{4^K}, which due to [PSZ20], is computable. Thus $(\mathcal{X}(K), \mu_K)$ is a computable probability measure space.

One property of $\mathbf{H}_{\mu_K}$ is that it can take arbitrary negative values which is not desirable. To this end we redefine $\mathbf{H}_{\mu_K}$. A capped, parameterized, uniform test takes in a natural number K and a measure μ over $SU(2^K)$, and outputs a μ-test over $SU(2^K)$, of the form $SU(2^K) \mapsto [0, 1]$. Notice that each test cannot have a value more than 1 which differs from the original definition of uniform tests. Using slightly modified reasoning of Lemma 18, there exists a universal test $\mathbf{t}_{K,\mu_K}(x)$, that multiplicatively dominates all capped parameterized uniform tests. The complexity of a unitary

operator U is defined using this test, with $\mathbf{H}_K(U) = -\log \mathbf{t}_{K,\mu_K}(U)$. By definition, for each K, $\mathbf{H}_K$ takes values between 0 and 4^K, which is the desired property to have.

The evolution operator randomly groups k qubits together then applies a random k-qubit unitary operator on them, distributed according to the Haar measure distribution over $SU(2^k)$.

However, as shown in [HFK+22], the set of unitary operators $U \in SU(2^K)$ formed from less than an exponential number of gates has measure 0 with respect to the Haar measure over $SU(2^K)$. This leads one to conjecture that $\mathbf{H}_K(U) = 0$, diverging from the linear complexity conjecture. Thus it appears that the best way to characterize black holes with algorithmic information theory is with coarse-graining.

Part VI

Independence Postulate

Chapter 26

Physics and the Independence Postulate

The Many Worlds Theory (**MWT**) was formulated by Hugh Everett [Eve57] as a solution to the measurement problem of Quantum Mechanics. Branching (a.k.a splitting of worlds) occurs during any process that magnifies microscopic superpositions to the macro-scale. This occurs in events including human measurements such as the double slit experiments, or natural processes such as radiation resulting in cell mutations.

One question is if **MWT** causes issues with the foundations of computer science. The physical Church Turing Thesis (**PCTT**) states that any functions computed by a physical system can be simulated by a Turing machine. A straw man argument for showing **MWT** and **PCTT** are in conflict is an experiment that measures the spin of an unending number of electrons, with each measurement bifurcating the current branch into two sub-branches. This results in a single branch in which the halting sequence is outputted. However this branch has Born probability converging to 0, and can be seen as a deviant, atypical branch.

In fact, conflicts do emerge between **MWT** and Algorithmic Information Theory. In particular, the Independence Postulate (**IP**) [Lev84, Lev13] is a finitary Church-Turing thesis, postulating that certain infinite and *finite* sequences cannot be found in nature, i.e. have high "addresses". One class of of forbidden sequences are large prefixes of the halting sequence. One such set of forbidden sequences is large prefixes of the halting sequence. If a forbidden sequence is found in nature, an information leak will occur. However **MWT** represents a theory in which such information leaks can occur.

Another promising area of research is Constructor Theory (**CT**) [Deu13] with main proponents David Deutsch and Chiara Marletto. **CT** aims to unify many areas of science with counterfactuals. Counterfactuals describe which processes that can occur or not occur. These counterfactuals are principles which it is conjectured that all laws of physics must adhere to. The basis tenet of **CT** is [Deu15]

> *All other laws of physics are expressible entirely in terms of statements about which physical transformations are possible and which are impossible, and why.*

In an online colloquium [Soc22], David Deutsch was asked if Gödel's Incompleteness Theorem or the halting problem would be incorporated into **CT**. David Deutsch responded with:

> *"No they wouldn't, at least we don't expect them to be added because those issues only arise in infinite sets and constructor theory regards... physical systems as always finite. It only makes statements about finite systems."*

However, among other things, **IP** is a finitary version of the halting problem, so we believe it must be reconciled with **CT**. The main issue is the following question. Is it possible or impossible to create or find a large prefix of the halting sequence? Is this even a well formed (answerable) question? For example, it is possible to create one such large prefix h one were to find any y of the same length and $y \oplus h$. But obviously such a construction seems lacking.

26.1 The Independence Postulate

In [Lev84, Lev13], the Independence Postulate, **IP**, was introduced:

> *Let $\alpha \in \{0,1\}^{*\infty}$ be a sequence defined with an n-bit mathematical statement (e.g., in Peano Arithmetic or Set Theory), and a sequence $\beta \in \{0,1\}^{*\infty}$ can be located in the physical world with a k-bit instruction set (e.g., ip-address). Then $\mathbf{I}(\alpha : \beta) < k + n + c$, for some small absolute constant c.*

Exercise 33 *Show that assuming **IP**, finite strings with recursive description that are much higher than higher-level math descriptions will have large addresses.*

Exercise 34 *Describe how conservation inequalities, such as Theorem 2 and Lemma 2 support **IP**.*

The **I** term is an information measure in Algorithmic Information Theory. For this chapter, the information term used is $\mathbf{I}(x : y) = \mathbf{K}(x) + \mathbf{K}(y) - \mathbf{K}(x, y)$, where **K** is the prefix-free Kolmogorov complexity. This definition of **I** can be used because the thought experiment only deal with finite sequences.

Let Ω_m be the first m bits of Chaitin's Omega (the probability that a universal Turing machine will halt). It is well known that $m <^+ \mathbf{K}(\Omega_m)$. Furthermore Ω_m can be described by a mathematical formula of size $O(\log m)$. Thus by **IP**, where $\Omega_m = \alpha = \beta$, Ω_m can only be found with physical addresses of size at least $m - O(\log m)$. Thus finding any sufficiently large sequence Ω_m is not physically possible. This is due to fact that the observable universe is 8.8×10^{26} meters across and a transistor can only be made to be 2×10^{-9} meters long. Thus if the minimum length of an address for a sequence is greater than a thousand, it cannot exist in nature. This sentiment was reflected in [Lev13], where sequences with small addresses are called "physical", and thus sequences with only high addresses are "unphysical". As we shall see in the next parts of this chapter, the sequence Ω_m for large enough m will cause trouble for both **MWT** and **CT**.

26.2 Many Worlds Theory

Some researchers believe there is an inherent problem in quantum mechanics. On one hand, the dynamics of quantum states is prescribed by unitary evolution. This evolution is deterministic and linear. On the other hand, measurements result in the collapse of the wavefunction. This evolution is non-linear and nondeterministic. This conflict is called the measurement problem of quantum mechanics.

The time of the collapse is undefined and the criteria for the kind of collapse are strange. The Born rule assigns probabilities to macroscopic outcomes. The projection postulate assigns new microscopic states to the system measured, depending on the the macroscopic outcome. One could argue that the apparatus itself should be modeled in quantum mechanics. However it's dynamics is deterministic. Probabilities only enter the conventional theory with the measurement postulates.

MWT was proposed by Everett as a way to remove the measurement postulate from quantum mechanics. The theory consists of unitary evolutions of quantum states without measurement collapses. For **MWT**, the collapse of the wave function is the change in dynamical influence of one part of the wavefunction over another, the decoherence of one part from the other. The result is a branching structure of the wavefunction and a collapse only in the phenomenological sense.

26.2.1 Branching Worlds

An example of a branching of universes can be seen in an idealized experiment with a single electron with spin $|\phi_\uparrow\rangle$ and $|\phi_\downarrow\rangle$. This description can be found in [SBKW10]. There is a measuring apparatus $\mathcal{A}$, which is in an initial state of $|\psi^{\mathcal{A}}_{\text{ready}}\rangle$. After $\mathcal{A}$ reads spin-up or spin-down then it is in state $|\psi^{\mathcal{A}}_{\text{reads spin }\uparrow}\rangle$ or $|\psi^{\mathcal{A}}_{\text{reads spin }\downarrow}\rangle$, respectively. The evolution for when the electron is solely spin-up or spin-down is

$$\begin{aligned}
|\phi_\uparrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{ready}}\rangle &\overset{\text{unitary}}{\longrightarrow} |\phi_\uparrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\uparrow}\rangle \\
|\phi_\downarrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{ready}}\rangle &\overset{\text{unitary}}{\longrightarrow} |\phi_\downarrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\downarrow}\rangle\,.
\end{aligned}$$

Furthermore, one can model the entire quantum state of an observer $\mathcal{O}$ of the apparatus, with

$$\begin{aligned}
&|\phi_\uparrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{ready}}\rangle \otimes |\xi^{\mathcal{O}}_{\text{ready}}\rangle \\
&\overset{\text{unitary}}{\longrightarrow} |\phi_\uparrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\uparrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{ready}}\rangle \\
&\overset{\text{unitary}}{\longrightarrow} |\phi_\uparrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\uparrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{reads spin }\uparrow}\rangle
\end{aligned}$$

$$\begin{aligned}
&|\phi_\downarrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{ready}}\rangle \otimes |\xi^{\mathcal{O}}_{\text{ready}}\rangle \\
&\overset{\text{unitary}}{\longrightarrow} |\phi_\downarrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\downarrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{ready}}\rangle \\
&\overset{\text{unitary}}{\longrightarrow} |\phi_\downarrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\downarrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{reads spin }\downarrow}\rangle\,.
\end{aligned}$$

For the general case, the electron is in a state $|\phi\rangle = a\,|\phi_\uparrow\rangle + b\,|\phi_\downarrow\rangle$, where $|a|^2 + |b|^2 = 1$. In this case, the final superposition would be of the form:

$$\begin{aligned}
&a\,|\phi_\uparrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\uparrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{reads spin }\uparrow}\rangle \\
&+b\,|\phi_\downarrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin }\downarrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{reads spin }\downarrow}\rangle\,.
\end{aligned}$$

This is a superposition of two branches, each of which describes a perfectly reasonable physical story. This bifurcation is one method on how the quantum state of universe bifurcates into two branches.

26.2.2 Müller's Theorem Revisited

Quantum computers have an interesting interpretation with respect to **MWT**. A quantum computer is realized by a number of qubits which can implemented in a number of ways such as trapped ions that behave as magnets. The qubits are isolated from the outside environment to make the decoherence time as long as possible. When the quantum computation begins, unitary transforms

are performs on the qubits, which in the context of **MWT**, causes an exponential branching on worlds, each containing a different qubit value. The operations of the quantum computer cause interference effects between the branches until a measurement at the end produces the same result for all branches. Müller's Theorem provides concrete limitations to this computational power.

Interaction between branches provides no benefit in compressing classical information.

26.2.3 Deriving the Born Rule

In the author's opinion, one of the main problems of **MWT** is its reconciliation of the Born rule, for which no proposed solution has universal consensus. In standard quantum mechanics, measurements are probabilistic operations. Measurements on a state vector $|\psi\rangle$, which is a unit vector over Hilbert space $\mathcal{H}$, are self-adjoint operators $\mathcal{O}$ on $\mathcal{H}$. Observables are real numbers that are the spectrum $\mathrm{Sp}(\mathcal{O})$ of $\mathcal{O}$. A measurement outcome is a subset $E \subseteq \mathrm{Sp}(\mathcal{O})$ with associated projector P_E on $\mathcal{H}$. Outcome E is observed on measurement of $\mathcal{O}$ on $|\psi\rangle$ with probability $P(E) = \langle\psi| P_E |\psi\rangle$. This is known as the Born rule. After this measurement, the new state becomes $P_E |\psi\rangle / \sqrt{\langle\psi| P_E |\psi\rangle}$. This is known as the projection postulate.

However, the Born rule and the projection postulate are not assumed by **MWT**. The dynamics are totally deterministic. Each branch is equally real to the observers in it. To address these issues, Everett first derived a typicality-measure that weights each branch of a state's superposition. Assuming a set of desirable constraints, Everett derived the typicality-measure to be equal to the norm-squared of the coefficients of each branch, i.e. the Born probability of each branch. Everett then drew a distinction between typical branches that have high typicality-measure and exotic atypical branches of decreasing typicality-measure. For the repeated measurements of the spin of an electron $|\phi\rangle = a |\phi_\uparrow\rangle + b |\phi_\downarrow\rangle$, the relative frequencies of up and down spin measurements in a typical branch converge to $|a|^2$ and $|b|^2$, respectively. The notion of typicality can be extended to measurements with many observables.

In a more recent resolution to the relation between **MWT** and probability, Deutsch introduced a decision theoretic interpretation [Deu99] that obtains the Born rule from the non-probabilistic axioms of quantum theory and non-probabilistic axioms of decision theory. Deutsch proved that rational actors are compelled to adopt the Born rule as the probability measure associated with their available actions. This approach is highly controversial, as some critics say the idea has circular logic.

Another attempt uses subjective probability [Vai98]. The experimenter puts on a blindfold before he finishes performing the experiment. After he finishes the experiment, he has uncertainty about what world he is in. This uncertainty is the foundation of a probability measure over the measurements. However, the actual form of the probability measure needs to be postulated:

Probability Postulate. *An observer should set his subjective probability of the outcome of a quantum experiment in proportion to the total measure of existence of all worlds with that outcome.*

Whichever explanation of the Born rule one adopts, the following section shows there is an issue with **MWT** and **IP**. There exist branches of substantial Born probability where information leaks occurs.

26.3 Violating the Independence Postulate

IP can be violated in the following idealized experiment measuring the spin $|\phi_\uparrow\rangle$ and $|\phi_\downarrow\rangle$ of m isolated electrons. We denote $|\phi_0\rangle$ for $|\phi_\uparrow\rangle$ and $|\phi_1\rangle$ for $|\phi_\downarrow\rangle$. The "address" (in the sense of **IP**) of this experiment (such as the physical address of the Large Hadron Collider) is $< O(\log m)$. The measuring apparatus will measure the spin of m electrons in the state $|\phi\rangle = \frac{1}{2}|\phi_\uparrow\rangle + \frac{1}{2}|\phi_\downarrow\rangle$. There is a measuring apparatus $\mathcal{A}$ with initial state of $|\psi^{\mathcal{A}}\rangle$, and after reading m spins of m electrons, it is in the state $|\psi^{\mathcal{A}}[x]\rangle$, where $x \in \{0,1\}^m$, whose ith bit is 1 iff the ith measurement returns $|\phi_1\rangle$.

The experiment evolves according to the following unitary transformation:

$$\bigotimes_{i=1}^{m} |\phi\rangle \otimes |\psi^{\mathcal{A}}\rangle \stackrel{\text{unitary}}{\longrightarrow} \sum_{a_1,\ldots,a_m \in \{0,1\}^m} 2^{-m/2} \bigotimes_{i=1}^{m} |\phi_{a_i}\rangle \otimes |\psi^{\mathcal{A}}[a_1 a_2 \ldots a_m]\rangle .$$

If the bits returned are Ω_m then a memory leak of size $m - O(\log m)$ has occurred, because Ω_m has been located by the address of the experiment, which is $O(\log m)$. Thus

$$\text{Born-Probability(a memory leak of size } m - O(\log m) \text{ occurred)} \geq 2^{-m}.$$

26.4 Reconciling MWT and IP

There are multiple variations of **MWT** when it comes to consistency across universes. In one formulation, all universes conform to the same physical laws. In another model, each universe has its own laws, for example different values of gravitational constant, etc. However, the experiment in the previous section shows that mathematics itself is different between universes, regardless of which model is used. In some universes, **IP** holds and there is no way to create information leaks. In other universes information leaks occur, and there are tasks where randomized algorithms fail but non-algorithmic physical methods succeeds. One such task is finding new axioms of mathematics. This was envisioned as a possibility by Gödel [Gö61], but there is a universal consensus of the impossibility of this task. Not any more! In addition, because information leaks are finite events, the Born probability of worlds containing them is not insignificant. In such worlds, **IP** cannot be formulated, and the the foundations of Algorithmic Information Theory itself become detached from reality.

Formulated another way, let us suppose the Born probability is derived from the probability postulate. We have a "blindfolded mathematician" who performs the experiment described above. Before the mathematician takes off her blindfold, she states the Independence Postulate. By the probability postulate, with measure 2^{-m} over all worlds, there is a memory leak of size $m - O(\log m)$ and **IP** statement by the mathematician is in error.

26.4.1 The Probability Rebuttal

As a rebuttal, one can, with non-zero probability, just flip a coin N times and get N bits of Chaitin's Omega. Or more generally, how does one account for a probability P over finite or infinite sequences learning information about a forbidden sequence β with good probability? Due to probabilistic conservation laws [Lev74, Lev84], we have

$$\Pr_{\alpha \sim P}[\mathbf{I}(\alpha : \beta) > \mathbf{I}(\langle P \rangle : \beta) + m] \stackrel{*}{<} 2^{-m}.$$

Thus the probability of a single event creating a leak is very small. However if many events occur, then the chances of a memory leak grows. However as there is many events, to locate one such leak, one will probably need a long address to find the leak, balancing out the **IP** equation.

This still leaves open the possibility of a memory leak occurring at an event with a small address. For example, say someone assigns a random 1 trillion bit sequence to every ip-address. What are the chances that one of them is $\Omega_{10^{12}}$? Since there are a small number of events that have a small address, the probability of a significant memory leak is extremely small. In physics on can postulate away events with extremely small probabilities. For example, the second law of thermodynamics states that entropy is non-decreasing, postulating away the extremely unlikely event that a large system suddenly decreases in thermodynamic entropy, i.e. a broken vase forming back to together.

26.4.2 Memory Leaks

There is no way to postulate away such memory leaks in **MWT**. Assuming the *probability postulate*, probability is a measure over the space of possible worlds. Thus when Bob now threatens to measure the spin of m particles, Alice now knows 2^{-m} of the resultant worlds will contain m bits of Chaitin's Omega, violating **IP**.

26.5 Constructor Theory

CT aims to define a set of principles, or counterfactuals that constrain how laws of the physics. For more motivation on **CT**, I refer the readers to [Deu13, Mar21]. Fundamental in **CT** are constructors which change substrates,

$$\text{Input Substrate} \xrightarrow{\text{Constructor}} \text{Output Substrate}.$$

A *constructor task* $\mathfrak{A}$ is a set of pairs each designating an input state for the task and an output state for that input. For example,

$$\mathfrak{A} = \{x_1 \to y_1, x_2 \to y_2, \dots\}.$$

A constructor is capable of performing task $\mathfrak{A}$, if whenever it is given a substrate in a input attribute of $\mathfrak{A}$, it transforms them to one of the output attribute that $\mathfrak{A}$ associates with that input.

Definition 76 *A task $\mathfrak{A}$ is impossible if there is a law of physics that forbids it being carried out with arbitrary accuracy and reliably by a constructor. Otherwise, $\mathfrak{A}$ is possible, with $\mathfrak{A}^{\checkmark}$.*

However a possible task is not guaranteed to occur. The central tenet of **CT** is that the laws of physics must conform to a set of counterfactuals or principles which are statements on whether tasks are possible or impossible. **CT** has been applied to large number of areas, including (but not limited to) classical information, quantum information theory, probability, life, and thermodynamics [Deu13, Deu15, Mar15, Mar16]. In the next section we will review the intersection of **CT** and information.

26.5.1 Information

In my opinion, the most successful endeavour of **CT** is its application to classical and quantum information. We review the work in [Deu15]. Any set of disjoint attributes is called a *varriable*. Whenever a substrate is in a state with attribute $x \in X$, where X is a variable, we say X is sharp with value x. Einstein's (1949) principle of locality has an intepretation as a counterfactual [Deu15],

There exists a mode of description such that the state of the combined system $\mathbf{S}_1 \oplus \mathbf{S}_2$ of any two substrates $\mathbf{S}_1$ and $\mathbf{S}_2$ is the ordered pair (x, y) of the states x of $\mathbf{S_1}$ and y of $\mathbf{S}_2$, and any construction undergone by $\mathbf{S}_1$ and not $\mathbf{S}_2$ can change only x and not y.

The theory of information relies on a **CT** theoretic description of *computation*. A *reversible computation* $\mathfrak{C}_\Pi(S)$ is the task of performing a permutation Π over some set S of at least two possible attributes of some substrate:

$$\mathfrak{C}_\Pi(S) = \bigcup \{x \to \Pi(x)\}\,.$$

A *computation variable* is a set S of two or more possible attrbutes for which $\mathfrak{C}_\Pi^\checkmark$ for all permutation Π over S. A *computation medium* is a substrate with at least one computation variable. With computation formalized, we are now ready to define information.

The *cloning task* for a set S of possible attributes of substrate $\mathbf{S}$ is the task

$$\mathfrak{R}_s(x_0) = \bigcup \{(x, x_0) \to (x, x)\}.$$

on substrate $\mathbf{S} \oplus \mathbf{S}$, where x_0 is some attribute realizable from naturally occurring resources. We will revisit this statement later. An *information variable* is a cloneable computation variable. An *information attribute* is a member of an information variable, and an *information medium* is a substrate that has at least one information variable.

A set X of possible attributes of a substrate $\mathbf{S}$ is *distinguishable* if

$$\left(\bigcup_{x \in X} \{x \to \Psi_x\}\right)^\checkmark,$$

where $\{\Psi_x\}$ is a information variable. If the original substrate continues to exist and the process stores its result in a second *output substrate*, the input variable X is *measurable*:

$$\left(\bigcup_{x \in X} \{(x, x_0) \to (y_x, `x'\}\right)^\checkmark.$$

The output substrate is prepared with a 'receptive' attribute x_0. We introduce the task of preparing a substrate. A variable X in a substrate $\mathbf{S}$ is *preparable* if there is a information medium $\mathbf{R}$ with an information variable W and a possible task $\mathfrak{A}^\checkmark$ such that for all $x \in X$ there is a $w(x) \in W$ such that $\{w(x) \to x\} \in \mathfrak{A}$. Thus

$$\left(\bigcup_{x \in X} \{w(x) \to x\}\right)^\checkmark.$$

26.6 Reconciling CT and IP

We discuss three issues between **IP** and **CT**. First, we recall from Section 26.3, that Ω_m is a forbidden string, and can only be found with a physical address of size at least $m - O(\log m)$. However the following principle [Deu15] states

***VI**. Any number of instance of any information medium, with any one of its information-instantiating attributes, is preparable from naturally occurring substrates.*

Take the information medium of a computer that has information variable $\{0,1\}^m$, for very large (but finite) m. By Principle **VI**, Ω_m can be prepared by naturally occurring substrates. However **IP** postulates that all such Ω_m cannot be found without an address almost the size of Ω_m. There is a difference in language from "naturally occurring substrates" and "high address" but the conflict remains.

We point out the second issue. An information variable S is a (cloneable) computation variable, such that every task permuting S with permutation Π is possible. Let the information variable S consist of all strings of length m for very large m. Let Π be a permutation that sends 0^m to Ω_m. By definition 76, this task is possible because there are no laws of physics which forbid this transformation to occur. However if this task is possible, 0^m is an address that can be compressed to a $O(\log m)$ address for Ω_m, contradicting **IP**. Thus by **IP** no information variables can exist.

We now point out the third issue. Let us say there is a task $\mathfrak{B}$ that maps a *preparable* variable S in a substrate **S** to a *measurable* variable T in substrate **T**. Since S is preparable, there is an information medium **R** with information variable R that can be mapped onto X with possible task $\mathfrak{A}$. Furthermore, since T is measurable, it can be mapped onto an information variable U in an information medium **U** with possible task $\mathfrak{C}$. In **CT**, one must conclude either the task $\mathfrak{A}$ is possible or impossible. Assume $\mathfrak{A}$ is possible. Then the combined task

$$(\mathfrak{ABC})^{\checkmark}$$

is possible. However if $\mathfrak{ABC}$ sends 0^m to Ω_m, then by **IP**, the combined task is impossible, and thus task $\mathfrak{B}$ is impossible. So now the question of possibility or impossibility of a task must take into account whether it can be combined with a preparer and a measurer to create an information leak. This is an intractable question that applies to every task with preparable input variables and measurable output variables.

26.6.1 The Halting Sequence Revisited

We revisit the question posed in the introduction:

Is it possible or impossible to create or find a large prefix of the halting sequence?

If we assert that this task is *possible*, then a violation of **IP** occurs. If we assert that this task is *impossible*, then this statement will be in conflict with how information variables are defined in **CT**. Thus there is a definite conflict between **IP** and **CT**.

26.7 Conclusion

As discussed in Section 26.4, **IP** postulates away a union of "bad" events. Such "forbidden" events break the inequality of **IP** and were initially called "information leaks". One can postulate away such leaks because the probability of single leak occurring is astronomically small [Lev13]. It remains to be seen how to reconcile **MWT**, **CT**, and **IP**. The simplest way to reconcile the **MWT** and **IP** is to just acknowledge there are branches where **IP** fails. Similarly, the easiest way to reconcile **CT** and **IP** is to discard one of them. Otherwise one would need statements like $\mathfrak{A}^{\checkmark *}$, which means that task $\mathfrak{A}$ is possible, but where $*$ is some additional theoretical equipment, such as an address system or a condition that a memory leak does not occur. However this reconciliation would be cumbersome, spoiling the elegance of **CT**. It remains to be seen how to overcome these obstacles.

Bibliography

[AS04] N. Alon and J. Spencer. *The Probabilistic Method*. Wiley, New York, 2004.

[BDH+12] L. Bienvenu, A. Day, M. Hoyrup, I. Mezhirov, and A. Shen. A constructive version of birkhoff's ergodic theorem for martin-löf random points. *Information and Computation*, 210:21–30, 2012.

[BDK+05] I Bjelaković, J Deuschel, T Krüger, R Seiler, R Siegmund-Schultze, and A Szkola. A quantum version of sanov's theorem. *Communications in Mathematical Physics*, 260(3):659–671, 2005.

[BH96] V. Buzek and M. Hillery. Quantum copying: Beyond the no-cloning theorem. *Phys. Rev. A*, 54:1844–1852, 1996.

[Bho21a] T. Bhojraj. Algorithmic randomness and kolmogorov complexity for qubits. *CoRR*, abs/2106.14280, 2021.

[Bho21b] T. Bhojraj. Prefix-free quantum kolmogorov complexity. *Theoretical Computer Science*, 875:65–80, 2021.

[BKM+06] F. Benatti, T. Krüger, M. Müller, R. Siegmund-Schultze, and A. Szko—a. Entropy and Quantum Kolmogorov Complexity: A Quantum Brudno's Theorem. *Communications in Mathematical Physics*, 265(2), 2006.

[BOD14] F. Benatti, S. K. Oskouei, and A. S. Deh Abad. Gacs Quantum Algorithmic Entropy in Infinite Dimensional Hilbert Spaces. *Journal of Mathematical Physics*, 55(8), 2014.

[Bru78] A. Brudno. The Complexity of the trajectories of a Dynamical System. *Russian Mathematical Surveys*, 33(1):197, 1978.

[BS18] A. Brown and L. Susskind. Second law of quantum complexity. *Phys. Rev. D*, 97, 2018.

[BST03] L. Boya, E. Sudarshan, and T. Tilma. Volumes of compact manifolds. *Reports on Mathematical Physics*, 52(3), 2003.

[BSZ17] A. Brown, L. Susskind, and Y. Zhao. Quantum complexity and negative curvature. *Phys. Rev. D*, 95, 2017.

[BV93] E. Bernstein and U. Vazirani. Quantum complexity theory. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing*, page 11–20, New York, NY, USA, 1993. Association for Computing Machinery.

[BvL01] A. Berthiaume, W. van Dam, and S. Laplante. Quantum Kolmogorov Complexity. *Journal of Computer and System Sciences*, 63(2), 2001.

[Deu99] David Deutsch. Quantum theory of probability and decisions. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 455(1988), 1999.

[Deu13] D. Deutsch. Constructor theory. *Synthese*, 190(18):4331–4359, 2013.

[Deu15] Marletto C. Deutsch, D. Constructor theory of information. *Proceedings of Royal Society A*, 471(2174), 2015.

[DJ13] A. Day and Miller. J. Randomness for non-computable measures. *Transactions of the American Mathematical Society*, 365(7):3575–3591, 2013.

[Eps19a] S. Epstein. Algorithmic no-cloning theorem. *IEEE Transactions on Information Theory*, 65(9):5925–5930, 2019.

[Eps19b] S. Epstein. Algorithmic no-cloning theorem. *IEEE Transactions on Information Theory*, 65(9), 2019.

[Eps19c] S. Epstein. On the algorithmic probability of sets. *CoRR*, abs/1907.04776, 2019.

[Eps19d] S. Epstein. On the algorithmic probability of sets. *CoRR*, abs/1907.04776, 2019.

[Eps20] Samuel Epstein. An extended coding theorem with application to quantum complexities. *Information and Computation*, 275, 2020.

[Eps21a] S. Epstein. On the algorithmic content of quantum measurements. *CoRR*, abs/2102.03905, 2021.

[Eps21b] Samuel Epstein. All sampling methods produce outliers. *IEEE Transactions on Information Theory*, 67(11):7568–7578, 2021.

[Eps22a] S. Epstein. The outlier theorem revisited. *CoRR*, abs/2203.08733, 2022.

[Eps22b] S. Epstein. Outliers, dynamics, and the independence postulate. *CoRR*, abs/2207.03955, 2022.

[Eps23a] S. Epstein. On the algorithmic information between probabilities. *CoRR*, abs/2303.07296, 2023.

[Eps23b] S. Epstein. On the existence of anomalies. *CoRR*, abs/2302.05972, 2023.

[Eps23c] S. Epstein. On the existence of anomalies, the reals case. *CoRR*, abs/2303.05614, 2023.

[Eps23d] S. Epstein. A quantum el theorem. *CoRR*, abs/2301.08348, 2023.

[Eps23e] S. Epstein. A quantum outlier theorem. *CoRR*, abs/2303.06256, 2023.

[Eve57] Hugh Everett. "relative state" formulation of quantum mechanics. *Rev. Mod. Phys.*, 29, 1957.

[FMN12] Greenberg N. Franklin, J., J. Miller, and K. Ng. Martin-löf random points satisfy birkhoff's ergodic theorem for effectively closed sets. *Proceedingys of the American Mathematical Society*, 140:623–3628, 2012.

[GÖ61] Kurt Gödel. The modern development of the foundations of mathematics in the light of philosophy. *In: Kurt Gödel. Collected Works. Volume III. Oxford University Press.*, 1961.

[GÓ1] P. Gács. Quantum Algorithmic Entropy. *Journal of Physics A Mathematical General*, 34(35), 2001.

[GÁ21] Peter Gács. Lecture notes on descriptional complexity and randomness. *CoRR*, abs/2105.04704, 2021.

[Gac94] P. Gacs. The boltzmann entropy and randomness tests. In *Proceedings Workshop on Physics and Computation. PhysComp '94*, pages 209–216, 1994.

[GHR10] S. Galatolo, M. Hoyrup, and C. Rojas. Computing the speed of convergence of ergodic averages and pseudorandom points in computable dynamical systems. *Electronic Proceedings in Theoretical Computer Science*, 24, 06 2010.

[GHR11] P. Gács, M. Hoyrup, and C. Rojas. Randomness on computable probability spaces—a dynamical point of view. *Theory of Computing Systems*, 48(3):465–485, 2011.

[HFK+22] J. Haferkamp, P. Faist, N. Kothakonda, J. Eisert, and N. Halpern. Linear growth of quantum circuit complexity. *Nature Physics*, 18(5):528–532, 2022.

[HR09] M. Hoyrup and C. Rojas. Computability of probability measures and martin-löf randomness over metric spaces. *Information and Computation*, 207(7):830–847, 2009.

[Lem]

[Lev74] L. A. Levin. Laws of Information Conservation (Non-growth) and Aspects of the Foundations of Probability Theory. *Problemy Peredachi Informatsii*, 10(3):206–210, 1974.

[Lev76] L. A. Levin. Uniform tests of randomness. *Soviet Math. Dokl.*, 17(2):337–340, 1976.

[Lev84] L. A. Levin. Randomness conservation inequalities; information and independence in mathematical theories. *Information and Control*, 61(1):15–37, 1984.

[Lev13] L. A. Levin. Forbidden information. *J. ACM*, 60(2), 2013.

[Lev16] L. A. Levin. Occam bound on lowest complexity of elements. *Annals of Pure and Applied Logic*, 167(10):897–900, 2016. And also: S. Epstein and L.A. Levin, Sets have simple members, arXiv preprint arXiv:1107.1458, 2011.

[LO97] M. Li and Vitányi O. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer, 1997.

[Mar15] C. Marletto. Constructor theory of life. *Proceedings of Royal Society A*, 12(104), 2015.

[Mar16] C. Marletto. Constructor theory of thermodynamics. *CoRR*, abs/1608.02625, 2016.

[Mar21] C. Marletto. *The Science of Can and Can't*. Viking, 2021.

[MS13] J. Maldacena and L. Susskind. Cool horizons for entangled black holes. *Fortsch. Phys.*, 61:781–811, 2013.

[Mue07] M. Mueller. Quantum kolmogorov complexity and the quantum turing machine. *CoRR*, abs/0712.4377, 2007.

[Mul08] M. Muller. Strongly Universal Quantum Turing Machines and Invariance of Kolmogorov Complexity. *IEEE Transactions on Information Theory*, 54(2), 2008.

[Mul09] M. Muller. On the quantum kolmogorov complexity og classical strings. *International Journal of Quantum Information*, 07(04):701–711, 2009.

[NS19] A. Nies and V. Scholz. Martin-Löf Random Quantum States. *Journal of Mathematical Physics*, 60(9), 2019.

[PSZ20] A. Pauly, D. Seon, and M. Ziegler. Computing Haar Measures. In *28th EACSL Annual Conference on Computer Science Logic (CSL 2020)*, volume 152 of *Leibniz International Proceedings in Informatics (LIPIcs)*, 2020.

[SBC01] R. Schack, T.. Brun, and C. Caves. Quantum bayes rule. *Phys. Rev. A*, 64:014305, 2001.

[SBKW10] S. Saunders, J. Barrett, A. Kent, and D. Wallace. *Many Worlds?: Everett, Quantum Theory, & Reality*. OUP Oxford, 2010.

[Sch71] C. Schnorr. A unified approach to the definition of random sequences. *Mathematical systems theory*, 5(3), 1971.

[Sch95] B. Schumacher. Quantum coding. *Phys. Rev. A*, 51, 1995.

[Sch10] M. Schlosshauer. *Decoherence*. Springer Berlin, Heidelberg, 2010.

[Sei08] U. Seifert. Stochastic thermodynamics: principles and perspectives. *The European Physical Journal B*, 64(3):423–431, 2008.

[She15] M. Shell. *Thermodynamics and Statistical Mechanics: An Integrated Approach*. Cambridge Series in Chemical Engineering. Cambridge University Press, 2015.

[Soc22] Quantum Information Society. Prof david deutsch - quantum information in many worlds, 2022.

[SS08] Y. Sekino and L. Susskind. Fast scramblers. *Journal of High Energy Physics*, 2008(10), 2008.

[Sus20] L. Susskind. *Three Lectures on Complexity and Black Holes*. Springer Cham, 2020.

[Tao] T. Tao. What's new: A cheap version of the kabatjanskii-levenstein bound for almost orthogonal vectors. https://terrytao.wordpress.com/2013/07/18/a-cheap-version-of-the-kabatjanskii-levenstein-bound-for-almost-orthogonal-vectors/. Accessed: 2024-01-11.

[Vai98] L Vaidman. On schizophrenic experiences of the neutron or why we should believe in the many-worlds interpretation of quantum theory. *International Studies in the Philosophy of Science*, 12(3):245–261, 1998.

[Ver21] N. Vereshchagin. Proofs of conservation inequalities for levin's notion of mutual information of 1974. *Theoretical Computer Science*, 856, 2021.

[Vit00] P Vitányi. Three Approaches to the Quantitative Definition of Information in an Individual Pure Quantum State. In *Proceedings of the 15th Annual IEEE Conference on Computational Complexity*, COCO '00, page 263. IEEE Computer Society, 2000.

[Vit01] P.M.B. Vitanyi. Quantum kolmogorov complexity based on classical descriptions. *IEEE Transactions on Information Theory*, 47(6):2464–2479, 2001.

[Vov]

[Wil13] M. Wilde. *Quantum Information Theory*. Cambridge University Press, 2013.

[WZ82] W. Wootters and W. Zurek. A single quantum cannot be cloned. *Nature*, 299(5886):802–803, 1982.

[Zur03] W Zurek. Decoherence, einselection, and the quantum origins of the classical. *Rev. Mod. Phys.*, 75:715–775, 2003.

Part VII

Appendix

Appendix A

An Extended Coding Theorem

In [Lev16, Eps19c], a new inequality in the field of algorithmic information theory was proven. For a finite set of natural numbers D, it was shown that the size of the smallest description of an element of D, $\min_{x\in D} \mathbf{K}(x)$, is not much smaller than the negative logarithm of the algorithmic probability of the set, $-\log \sum_{x\in D} \mathbf{m}(x)$. This inequality holds for non-exotic sets whose encoding has little mutual information with the halting sequence, $\mathbf{I}(D;\mathcal{H}) = \mathbf{K}(D) - \mathbf{K}(D|\mathcal{H})$.

$$\min_{x\in D} \mathbf{K}(x) <^{\log} -\log \sum_{x\in D} \mathbf{m}(x) + \mathbf{I}(D;\mathcal{H}).$$

Due to algorithmic conservation laws, there are no algorithmic means to produce sets with arbitrary high mutual information with the halting sequence. In this appendix, we introduce an update on the above inequality, proving for non-exotic maps f between whole numbers with a finite domain, $\min_{x\in \mathrm{Dom}(f)} \mathbf{K}(x) + f(x)$ is close to the amount $-\log \sum_{x\in \mathrm{Dom}(f)} \mathbf{m}(x)2^{-f(x)}$. Exotic maps f have encodings with high mutual information with the halting sequence, $\mathbf{I}(f;\mathcal{H})$, with

$$\min_{x\in \mathrm{Dom}(f)} \mathbf{K}(x) + f(x) <^{\log} -\log \sum_{x\in \mathrm{Dom}(f)} \mathbf{m}(x)2^{-f(x)} + \mathbf{I}(f;\mathcal{H}).$$

The above inequality can be seen as an extended coding theorem.

A.1 Left-Total Machines

The notion of the "left-total" universal algorithm is needed for the proof of both the mixed state and pure state coding theorems. We say $x \in \{0,1\}^*$ is total with respect to a machine if the machine halts on all sufficiently long extensions of x. More formally, x is total with respect to T_y for some $y \in \{0,1\}^{*\infty}$ iff there exists a finite prefix free set of strings $Z \subset \{0,1\}^*$ where $\sum_{z\in Z} 2^{-\|z\|} = 1$ and $T_y(xz) \neq \perp$ for all $z \in Z$. We say (finite or infinite) string $\alpha \in \{0,1\}^{*\infty}$ is to the "left" of $\beta \in \{0,1\}^{*\infty}$, and use the notation $\alpha \lhd \beta$, if there exists a $x \in \{0,1\}^*$ such that $x0 \sqsubseteq \alpha$ and $x1 \sqsubseteq \beta$. A machine T is left-total if for all auxiliary strings $\alpha \in \{0,1\}^{*\infty}$ and for all $x, y \in \{0,1\}^*$ with $x \lhd y$, one has that $T_\alpha(y) \neq \perp$ implies that x is total with respect to T_α. An example can be seen in Figure A.1.

For the remaining part of this chapter, we can and will change the universal self delimiting machine U into a universal left-total machine U' by the following definition. The algorithm U' enumerates all strings $p \in \{0,1\}^*$ in order of their convergence time of $U(p)$ and successively assigns them consecutive intervals $i_p \subset [0,1]$ of width $2^{-\|p\|}$. Then U' outputs $U(p)$ on input p' if the open interval corresponding to p' and not that of $(p')^-$ is strictly contained in i_p. The open interval in

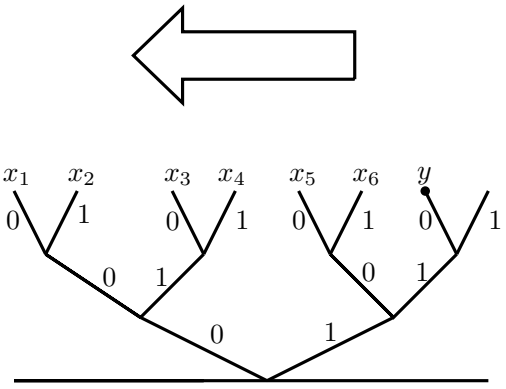


Figure A.1: The above diagram represents the domain of a left total machine T with the 0 bits branching to the left and the 1 bits branching to the right. For $i \in \{1..5\}$, $x_i \lhd x_{i+1}$ and $x_i \lhd y$. Assuming $T(y)$ halts, each x_i is total. This also implies each x_i^- is total as well.

[0,1] corresponding with p' is $([p']2^{-\|p'\|}, ([p']+1)2^{-\|p'\|})$ where $[p]$ is the value of p in binary. For example, the value of both strings 011 and 0011 is 3. The value of 0100 is 4. The same definition applies for the machines U'_α and U_α, over all $\alpha \in \{0,1\}^{*\infty}$. We now set U to equal U'.

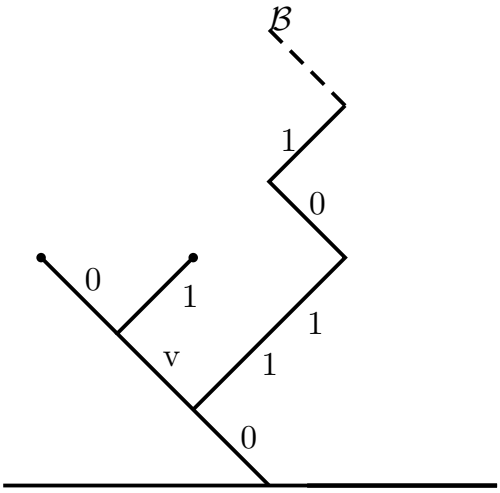


Figure A.2: The above diagram represents the domain of the universal left-total algorithm U, with the 0 bits branching to the left and the 1 bits branching to the right. The strings in the above diagram, $0v0$ and $0v1$, are halting inputs to U with $U(0v0) \neq \perp$ and $U(0v1) \neq \perp$. So $0v$ is a total string. The infinite border sequence $\mathcal{B} \in \{0,1\}^\infty$ represents the unique infinite sequence such that all its finite prefixes have total and non total extensions. All finite strings branching to the right of $\mathcal{B}$ will cause U to diverge.

Without loss of generality, the complexity terms of Chapter 1 are defined in this section with respect to the universal left total machine U. The infinite border sequence $\mathcal{B} \in \{0,1\}^\infty$ represents the unique infinite sequence such that all its finite prefixes have total and non total extensions. The term "border" is used because for any string $x \in \{0,1\}^*$, $x \lhd \mathcal{B}$ implies that x total with respect to

U and $\mathcal{B} \lhd x$ implies that U will never halt when given x as an initial input. Figure A.2 shows the domain of U with respect to $\mathcal{B}$. The border sequence is computable from $\mathcal{H}$.

For all total strings $b \in \{0,1\}^*$, we define the semimeasure $\mathbf{m}_b(x) = \sum\{2^{-\|p\|} : U(p) = x,\ p \lhd b \text{ or } b \sqsubseteq p\}$. If b is not total then $\mathbf{m}_b(x) = \perp$ is undefined. Thus the algorithmic weight $\mathbf{m}_b$ of a string x is approximated using programs that either extend b or are to the left of b.

A.2 Extended Coding Theorem

Lemma 34 *Let f be a elementary map and m be a elementary semi measure. Let $a \in \mathbb{W}$ vary over* $\mathrm{Dom}(f)$. *Then* $\min_a f(a) + \mathbf{K}(a|m) <^{\log} -\log \sum_a m(a)2^{-f(a)} + \mathbf{Ks}(f|m)$.

Proof. If m is not a proper probability measure, and R is the support of m, we modify m to give an arbitary $b \in \mathbb{W}$, the value of $1 - m(R)$. So m can be assumed to be an elementary probability measure. Since all terms in the theorem are conditioned on $\langle m \rangle$, we will also condition all complexity terms in the proof on $\langle m \rangle$ and drop its notation. More formally, $U(x)$ is used to denote $U_{\langle m \rangle}(x)$, $\mathbf{K}(x)$ is used to denote $\mathbf{K}(x|m)$, and $\mathbf{Ks}(f)$ is used to denote $\mathbf{Ks}(f|m)$.

For any elementary map g, let $g_n = g^{-1}(n) \cap \mathrm{Supp}(m)$ and let $g_{\leq n} = \cup_{i=0}^{n} g_i$, for $n \in \mathbb{W} \cup \{\infty\}$. Let $s = \lceil -\log \sum_{a \in f_{\leq \infty}} m(a)2^{-f(a)} \rceil$. Using the reasoning of Markov's inequality,

$$\sum_{a \in f_{\leq \infty}} m(a)2^{-f(a)} \geq 2^{-s}, \tag{A.1}$$

$$\sum_{a \in f_{\leq \infty} \setminus f_{\leq s}} m(a)2^{-f(a)} \leq \sum_{a \in f_{\leq \infty} \setminus f_{\leq s}} m(a)2^{-s-1} \leq 2^{-s-1}, \tag{A.2}$$

$$\sum_{a \in f_{\leq s}} m(a)2^{-f(a)} \geq 2^{-s-1}. \tag{A.3}$$

Equation (A.1) follows from the definition of s and Equation (A.3) follows from Equations (A.1) and (A.2). We now turn our attention to creating an elementary probability measure Q with the following properties:

1. f is typical of Q and Q is simple, i.e. there is a $v \in \{0,1\}^*$ with $U(v) = \langle Q \rangle$ and $\|v\| + 3\log\max\{\mathbf{d}(f|Q, v), 1\}$ is not much larger than $\mathbf{Ks}(f)$.

2. All strings in the support of Q encode elementary functions g whose range contain a lot of values that are not greater than s, with $\sum_{a \in g_{\leq s}} m(a)2^{-g(a)} \geq 2^{-s-1}$.

To accomplish this goal, we start with the program $v' \in \{0,1\}^*$ and elementary probability measure Q' that realizes the stochasticity of f, with $U(v') = \langle Q' \rangle$, and also with the relation $\mathbf{Ks}(f) = \|v'\| + 3\log\max\{\mathbf{d}(f|Q', v'), 1\}$. Note that this implies $\langle f \rangle \in \mathrm{Supp}(Q')$. Let Q be the elementary probability measure equal to Q' conditioned on the set of (encoded) elementary maps g such that $\sum_{a \in g_{\leq s}} m(a)2^{-g(a)} \geq 2^{-s-1}$. Thus $Q(\langle g \rangle) = [g \in S]Q'(g)/Q'(S)$, where $S \subset \{0,1\}^*$, the support of Q, is defined as $S = \{\langle g \rangle : g \in \mathrm{Supp}(Q'), \sum_{a \in g_{\leq s}} m(a)2^{-g(a)} \geq 2^{-s-1}\}$. This Q is computable from v' and s. Using this fact, define the Q program $v \in \{0,1\}^*$, to be of the form $v = v_0 v_s v'$, where $v_0 \in \{0,1\}^*$ is helper code of size $O(1)$, and $v_s \in \{0,1\}^*$ is a shortest U-program

for s. So $\|v\| <^+ \|v'\| + \mathbf{K}(s)$. We define $d = \max\{\mathbf{d}(f|Q, v), 1\}$ and we have that

$$\begin{aligned}
\|v\| &<^+ \|v'\| + \mathbf{K}(s), \\
\|v\| + 3\log d &<^+ \|v'\| + \mathbf{K}(s) + 3\log d \\
&<^+ \|v'\| + \mathbf{K}(s) + 3\log(\max\{-\log Q(f) - \mathbf{K}(f|v), 1\}) \\
&<^+ \|v'\| + \mathbf{K}(s) + 3\log(\max\{-\log Q'(f) - \mathbf{K}(f|v), 1\}) && \text{(A.4)} \\
&<^+ \|v'\| + \mathbf{K}(s) + 3\log(\max\{-\log Q'(f) - \mathbf{K}(f|v') + \mathbf{K}(v|v'), 1\}) && \text{(A.5)} \\
&<^+ \|v'\| + \mathbf{K}(s) + 3\log(\max\{-\log Q'(f) - \mathbf{K}(f|v') + \mathbf{K}(s), 1\}) && \text{(A.6)} \\
&<^{\log} \|v'\| + \mathbf{K}(s) + 3\log(\max\{-\log Q'(f) - \mathbf{K}(f|v'), 1\}), \\
\|v\| + 3\log d &<^{\log} \mathbf{Ks}(f) + \mathbf{K}(s). && \text{(A.7)}
\end{aligned}$$

Equation (A.4) follows from $Q(f) = Q'(f)/Q'(\mathrm{Supp}(Q))$, and thus $-\log Q(f) \leq -\log Q'(f)$. Equation (A.5) follows from the inequality $\mathbf{K}(f|v') <^+ \mathbf{K}(f|v) + \mathbf{K}(v|v')$. Equation (A.6) follows from v being computable from v' and v_s, and thus $\mathbf{K}(v|v') <^+ \mathbf{K}(s)$.

We now create a small set of lists of numbers A that will intersect with the range of a large percentage of the support of Q. We do so by using the probabilistic method. Let $c \in \mathbb{N}$ be a constant solely dependent on the universal Turing machine U to be determined later. We use an elementary measure w_n over lists A^n of (possibly repeating) whole numbers of size $cd2^{s+1-n}$ where $w_n(A^n) = \prod_{i=1}^{cd2^{s+1-n}} m(A^n_i)$. For a set of $s+1$ lists $A = \{A^n\}_{n=0}^s$, we a measure w over A, where $w(A) = \prod_{n=0}^s w_n(A^n)$.

For a set of lists A and elementary function g, let $\mathbf{1}(g, A) = 1$ if $g_n \cap A^n = \emptyset$ for all $n \in [0, s]$, and $\mathbf{1}(g, A) = 0$, otherwise. Thus

$$\begin{aligned}
\mathbf{E}_{g\sim Q}\mathbf{E}_{A\sim w}[\mathbf{1}(g, A)] &= \sum_g Q(g) \prod_{n=0}^{s} (1 - m(g_n))^{|A^n|} \\
&\leq \sum_g Q(g) \prod_{n=0}^{s} \exp\{-|A^n| m(g_n)\} && \text{(A.8)} \\
&= \sum_g Q(g) \exp\left\{-\sum_{n=0}^{s} |A^n| m(g_n)\right\} \\
&= \sum_g Q(g) \exp\left\{-\sum_{n=0}^{s} cd2^{s+1-n} m(g_n)\right\} \\
&= \sum_g Q(g) \exp\left\{-cd2^{s+1} \sum_{n=0}^{s} m(g_n) 2^{-n}\right\} \\
\mathbf{E}_{g\sim Q}\mathbf{E}_{A\sim \lambda}[\mathbf{1}(g, A)] &\leq \sum_g Q(g) \exp\{-cd\} = \exp\{-cd\}. && \text{(A.9)}
\end{aligned}$$

Equation (A.8) follows from the inequality $(1-a) \leq e^{-a}$ over $a \in [0, 1]$. Equation (A.9) follows from the definition of the support of Q, where $g \in \mathrm{Supp}(Q)$ iff $\sum_{a \in g_{\leq s}} m(a) 2^{-g(a)} \geq 2^{-s-1}$. By the probability argument, there exists a set of lists $A = \{A^n\}_{n=0}^s$ such that $|A^n| = cd2^{s+1-n}$ and

$$\mathbf{E}_{g\sim Q}[\mathbf{1}(g, A)] \leq \exp\{-cd\}.$$

There exists a brute force search algorithm that on input c, d, v, outputs A. Note that the strings s and $\langle Q \rangle$ are computable from v. This algorithm computes all possible sets of lists $A' = \{A'^n\}_{n=0}^{s}$, $|A'^n| = cd2^{s+1-n}$, $A'^n \subseteq \mathrm{Supp}(Q)$ and outputs the first A' such that $\mathbf{E}_{g\sim Q}[\mathbf{1}(g, A')] \leq \exp\{-cd\}$. The existence of such an A' is guaranteed by Equation (A.9). So

$$\mathbf{K}(A) <^+ \mathbf{K}(c, d, v). \tag{A.10}$$

We now show that there is an n where $f_n \cap A^n \neq \emptyset$. To do so, we show that any function g in the support of Q whose range does not interesct with A, i.e. $\mathbf{1}(g, A) = 1$ will have a very high deficiency of randomness with respect to Q and v. For all such g and proper choice of c solely dependent on U,

$$\begin{aligned}
\mathbf{d}(g|Q, v) &= \lfloor -\log Q(g) \rfloor - \mathbf{K}(g|v) \\
&> -\log Q(g) - (-\log \mathbf{1}(g, A)\lfloor e^{cd} \rfloor Q(g) + \mathbf{K}(\mathbf{1}(\cdot, A)\lfloor e^{cd} \rfloor Q(\cdot)|v)) - O(1) && \text{(A.11)} \\
&> cd \log e - \mathbf{K}(\mathbf{1}(\cdot, A)\lfloor e^{cd} \rfloor Q(\cdot)|v) - O(1) \\
&> cd \log - \mathbf{K}(A, c, d|v) - O(1) \\
&> cd \log e - \mathbf{K}(c, d) > d. && \text{(A.12)}
\end{aligned}$$

With c being chosen, it is removed from consideration for the rest of the proof, with $c \in \emptyset(1)$. Equation A.11 is due to the fact that for any elementary semimeasure P, $\mathbf{K}(x) <^+ \mathbf{K}(P) - \log P(x)$. Equation A.12 is due to Equation A.10. So $\mathbf{1}(f, A) = 0$, otherwise by the above equation, $\mathbf{d}(f|Q, v) > d$, causing a contradiction. So there exists $n \in [0, s]$ with $a \in f_n \cap A^n$ and

$$\begin{aligned}
\mathbf{K}(a) &<^+ \log |A^n| + \mathbf{K}(A^n) \\
&<^+ \log |A^n| + \mathbf{K}(A) + \mathbf{K}(A^n|A) \\
&<^+ (\log d + s - n) + \mathbf{K}(d, v) + \mathbf{K}(n) && \text{(A.13)} \\
&=^+ \log d + s - f(a) + \mathbf{K}(d, v) + \mathbf{K}(f(a)) \\
\mathbf{K}(a) + f(a) &<^+ \log d + s + \mathbf{K}(v) + \mathbf{K}(d) + \mathbf{K}(f(a)) \\
\mathbf{K}(a) + f(a) &<^{\log} s + \|v\| + 3 \log d && \text{(A.14)} \\
\mathbf{K}(a) + f(a) &<^{\log} s + \chi(f) && \text{(A.15)} \\
\min_{a \in f_{\leq \infty}} \mathbf{K}(a) + f(a) &<^{\log} -\log \sum_{a \in f_{\leq \infty}} m(a) 2^{-f(a)} + \chi(f). && \text{(A.16)}
\end{aligned}$$

Equation (A.13) follows from Equation (A.10), and from $c \in O(1)$. Equation (A.14) follows from $\mathbf{K}(x) <^{\log} \|x\|$ for $x \in \{0,1\}^* \cup \mathbb{W}$. Equation (A.15) follows directly from Equation (A.7). Equation (A.16) follows from the definition of s and its form proves the theorem.

Proposition 49 *For border prefix $b \sqsubseteq \mathcal{B}$, $\mathbf{K}(b|\mathcal{H}) <^+ \mathbf{K}(\|b\|)$ and $\|b\| <^+ \mathbf{K}(b)$.*

Proof. The border $\mathcal{B}$ is computable from the halting sequence $\mathcal{H}$, so it follows easily $\mathbf{K}(b|\mathcal{H}) <^+ \mathbf{K}(\|b\|)$. We recall that $\Omega = \sum_x \mathbf{m}(x)$ is Chaitin's Omega, the probability that U will halt. It is well known that the binary expansion $\Omega' \in \{0,1\}^\infty$ of Ω is Martin Löf random. Given $b \sqsubset \mathcal{B}$, $\|b\| \in \{0,1\}^n$, one can compute $\hat{\Omega} = \sum\{2^{-\|y\|}[U(y) \neq \perp] : y \lhd b\}$ with differs from Ω in the summation of programs which branch from $\mathcal{B}$ at positions $n+1$ or higher. Thus $\Omega - \hat{\Omega} \leq 2^{-n}$. So $n <^+ \mathbf{K}(\Omega'[0..n-1]) <^+ \mathbf{K}(\Omega'[0..n-1], b) <^+ \mathbf{K}(\Omega'[0..n-1]|b) + \mathbf{K}(b) <^+ \mathbf{K}(b)$.

Proposition 50 *If $b \in \{0,1\}^*$ is total and b^- is not total, then b^- is a border prefix, with $b^- \sqsubset \mathcal{B}$.*

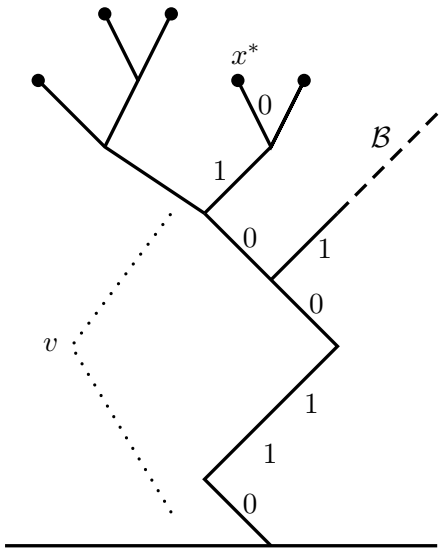


Figure A.3: The above figure shows an example of the domain of left-total U with the terms used in Lemma 35. $x^* = 0110010$ and $v = 01100$. Since v is total and v^- is not, v^- is a prefix of the border sequence $\mathcal{B}$. In the above example, assuming all halting extensions of v produce a unique output, $|\text{Support}(Q)| = 5$, and $Q(x) = 2^{-\|x^*\|+\|v\|} = 0.25$.

Proof. If $b \in \{0,1\}^*$ is total and b^- is not, then b^- has a total extension b^-0 and a non total extension b^-1, thus by the definition of the border sequence, $b^- \sqsubset \mathcal{B}$. The following lemma shows that non-stochastic strings x are "exotic," i.e. have high $\mathbf{I}(x\,;\mathcal{H})$ information with the halting sequence.

Lemma 35 *For $x \in \{0,1\}^*$, $\mathbf{Ks}(x) <^{\log} \mathbf{I}(x\,;\mathcal{H})$.*

Proof. Let $U(x^*) = x$, $\|x^*\| = \mathbf{K}(x)$, and v be the shortest total prefix of x^*. We define the elementary probability measure Q such that $Q(a) = \sum_w 2^{-\|w\|}[U(vw) = a]$. Thus Q is computable relative to v. In addition, since $v \sqsubseteq x^*$, one has the lower bound $Q(x) \geq 2^{-\|x^*\|+\|v\|} = 2^{-\mathbf{K}(x)+\|v\|}$. Therefore

$$\begin{aligned}
\mathbf{d}(x|Q,v) &= \lfloor -\log Q(x) \rfloor - \mathbf{K}(x|v) \\
&\leq \mathbf{K}(x) - \|v\| - \mathbf{K}(x|v) \\
&<^+ (\mathbf{K}(v) + \mathbf{K}(x|v)) - \|v\| - \mathbf{K}(x|v) \\
&<^+ (\|v\| + \mathbf{K}(\|v\|) + \mathbf{K}(x|v)) - \|v\| - \mathbf{K}(x|v), \\
\mathbf{d}(x|Q,v) &<^+ \mathbf{K}(\|v\|). \qquad \text{(A.17)}
\end{aligned}$$

Since v is total and v^- is not total, by proposition (50), v^- is a prefix of the border sequence

$\mathcal{B}$ (see Figure A.3). In addition, Q is computable from v. Therefore

$$\begin{aligned}
\mathbf{K}(x|\mathcal{H}) &<^+ \mathbf{K}(x|Q) + \mathbf{K}(Q|\mathcal{H}) \\
&<^+ \mathbf{K}(x|Q) + \mathbf{K}(v|\mathcal{H}) \\
&<^+ -\log Q(x) + \mathbf{K}(\|v\|) \qquad \text{(A.18)} \\
&<^+ \mathbf{K}(x) - \|v\| + \mathbf{K}(\|v\|), \\
\|v\| &<^+ \mathbf{K}(x) - \mathbf{K}(x|\mathcal{H}) + \mathbf{K}(\|v\|), \\
\|v\| &<^{\log} \mathbf{I}(x;\mathcal{H}). \qquad \text{(A.19)}
\end{aligned}$$

Equation (A.18) is due to Proposition (49). Since Q is computable from v, one gets $\mathbf{Ks}(x) <^+ \mathbf{K}(v) + 3\log(\max\{\mathbf{d}(x|Q,v),1\}) <^+ \|v\| + \mathbf{K}(\|v\|) + 3\log(\max\{\mathbf{d}(x|Q,v),1\})$. Due to Equation A.17, one gets $\mathbf{Ks}(x) \leq \|v\| + O(\mathbf{K}(\|v\|)) <^{\log} \|v\|$. Due to Equation A.19, one gets $\mathbf{Ks}(x) <^{\log} \mathbf{I}(x;\mathcal{H})$.

Theorem 114 *For elementary map f, $\min_{a\in\mathrm{Dom}(f)} f(a)+\mathbf{K}(a) <^{\log} -\log\sum_{a\in\mathrm{Dom}(f)} \mathbf{m}(a)2^{-f(a)} + \mathbf{I}(\langle f\rangle;\mathcal{H})$.*

Proof. Let $s = \lceil 1 - \log\sum_{a\in\mathrm{Dom}(f)} \mathbf{m}(a)2^{-f(a)}\rceil$ and let $S(z) = \lceil -\log\sum_{a\in\mathrm{Dom}(f)} \mathbf{m}_z(a)2^{-f(a)}\rceil$ be a partial recursive function from strings to rational numbers. S is defined solely on total strings, where $S(z)\neq\perp$ iff z is total. For total strings z, z^-, one has that $\mathbf{m}_{z^-}(x) \geq \mathbf{m}_z(x)$ and therefore $S(z^-) \leq S(z)$. Let b be the shortest total string with the property that $S(b) < s$. This implies $S(b^-) = \perp$ and thus b^- is not total. So by proposition (50), $b^- \sqsubseteq \mathcal{B}$ is a prefix of border. Lemma 34, with U containing b on an auxilliary tape, with $m(a) = \mathbf{m}_b(a)$, provides $a \in \mathbb{W}$ such that $\mathbf{K}(a|m,b) + f(a) <^{\log} s + \mathbf{Ks}(f|m,b)$. Since $\mathbf{K}(m|b) = O(1)$, we have Equation (A.20). Lemma (35), conditional on b, results in Equation (A.21), with

$$\mathbf{K}(a|b) + f(a) <^{\log} s + \mathbf{Ks}(f|b), \qquad \text{(A.20)}$$
$$\mathbf{K}(a|b) + f(a) <^{\log} s + \mathbf{I}(f;\mathcal{H}|b), \qquad \text{(A.21)}$$
$$\mathbf{K}(a|b) + f(a) <^{\log} s + \mathbf{K}(f|b) - \mathbf{K}(f|b,\mathcal{H}). \qquad \text{(A.22)}$$

Using the fact that $\mathbf{K}(a) <^+ \mathbf{K}(a|b) + \mathbf{K}(b)$, we get $\mathbf{K}(a) - \mathbf{K}(b) <^+ \mathbf{K}(a|b)$, and combined with Equation (A.22), we get Equation (A.23). Equation (A.24) is due to the chain rule $\mathbf{K}(b) + \mathbf{K}(f|b) <^{\log} \mathbf{K}(f) + \mathbf{K}(b|f)$. Equation (A.25) follows from the inequality $\mathbf{K}(f|\mathcal{H}) <^+ \mathbf{K}(f|b,\mathcal{H}) + \mathbf{K}(b|\mathcal{H})$.

$$\mathbf{K}(a) + f(a) <^{\log} s + \mathbf{K}(b) + \mathbf{K}(f|b) - \mathbf{K}(f|b,\mathcal{H}), \qquad \text{(A.23)}$$
$$\mathbf{K}(a) + f(a) <^{\log} s + \mathbf{K}(f) + \mathbf{K}(b|f) - \mathbf{K}(f|b,\mathcal{H}), \qquad \text{(A.24)}$$
$$\mathbf{K}(a) + f(a) <^{\log} s + \mathbf{K}(f) + \mathbf{K}(b|f) - \mathbf{K}(f|\mathcal{H}) + \mathbf{K}(b|\mathcal{H}), \qquad \text{(A.25)}$$
$$\mathbf{K}(a) + f(a) <^{\log} s + \mathbf{I}(f;\mathcal{H}) + (\mathbf{K}(b|f) + \mathbf{K}(b|\mathcal{H})). \qquad \text{(A.26)}$$

The remaining part of the proof shows that $\mathbf{K}(b|f) + \mathbf{K}(b|\mathcal{H}) = O(\log(s + \mathbf{K}(b)))$. This is sufficient to proof the theorem due to its logarithmic precision and by the right hand side of the inequality of Equation (A.23) being larger than $s + \mathbf{K}(b)$ (up to a logarithmic factor). Since b is a prefix of border, due to proposition (49), one gets that $\mathbf{K}(b|\mathcal{H}) < O(\mathbf{K}(\|b\|)) < O(\log\|b\|) < O(\log\mathbf{K}(b))$. Thus combined with Equation (A.26) and also Equation (A.23), one gets

$$\mathbf{K}(a) + f(a) <^{\log} s + \mathbf{I}(f;\mathcal{H}) + \mathbf{K}(b|f). \qquad \text{(A.27)}$$

We now prove $\mathbf{K}(b|f) <^{+} \mathbf{K}(s, \|b\|)$. This follows from the existence of an algorithm, that when given f, s, and $\|b\|$, computes $S(b')$ for all $b' \in \{0,1\}^{\|b\|}$ ordered by $\lhd$, and then outputs the first b' such that $S(b') < s$. This output is b otherwise there exists total $b' \lhd b$, with $\|b'\| = \|b\|$, and $S(b') < s$. This implies the existence of total string b'^{-} such that $S(b'^{-}) < s$. This contradicts the definition of b being the shortest total string with $S(b) < s$. So $\mathbf{K}(b|f) <^{+} \mathbf{K}(s, \|b\|)$ and thus one gets the final form of the theorem, as shown below. Equation (A.28) is again due to the right hand side of Equation (A.23).

$$\begin{aligned} \mathbf{K}(a) + f(a) &<^{\log} s + \mathbf{I}(f; \mathcal{H}) + \mathbf{K}(s, \|b\|), \\ \mathbf{K}(a) + f(a) &<^{\log} s + \mathbf{I}(f; \mathcal{H}), \\ \min_{a \in \mathrm{Dom}(f)} \mathbf{K}(a) + f(a) &<^{\log} -\log \sum_{a \in \mathrm{Dom}(f)} \mathbf{m}(a) 2^{-f(a)} + \mathbf{I}(f; \mathcal{H}). \end{aligned} \tag{A.28}$$

Corollary 47 (EL Theorem) *For finite* $D \subset \{0,1\}^*$, $\min_{x\ inD} \mathbf{K}(x) <^{\log} -\log \sum_{x \in D} \mathbf{m}(x) + \mathbf{I}(x; \mathcal{H})$.

Appendix B

Mixed State Quantum Coding Theorem

In Lemma 5, a coding theorem was introduced for pure states. In this appendix, a mixed state coding theorem is proven. We assume that the universal Turing machine U is *left-total*, as defined in Appendix A. We define $\mathcal{B}$ to be the infinite border sequence, also defined in Appendix A. The information term $\mathbf{I}$ is from Definition 2. The following proof uses notation from quantum mechanics, defined in Part II of the manuscript.

Theorem 115 *For each density matrix σ with $k = \lceil -\log \sum_\rho \mathbf{m}(\rho)\mathrm{Tr}\rho\sigma \rceil$, there is an elementary density matrix ρ such that*

1. $\mathbf{K}(\rho) - \log \mathrm{Tr}\rho\sigma <^{\log} k$,
2. $\mathbf{K}(\rho) <^{\log} \mathbf{I}(\langle\sigma\rangle : \mathcal{H}) + \mathbf{K}(k)$.

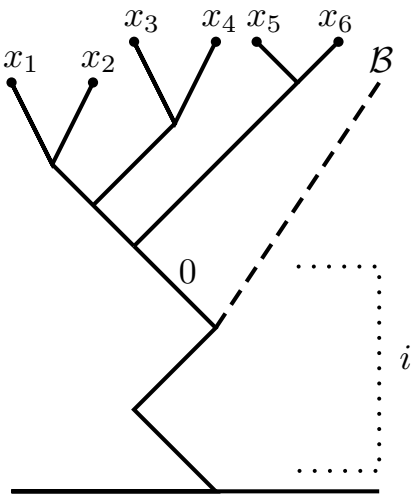
Proof. By definition of $\boldsymbol{\mu}$, we have that $k =^+ -\log \mathrm{Tr}\boldsymbol{\mu}\sigma$. For a given sequence x, we define the following semi-density matrix with the form $\nu[x] = \sum\{2^{-\|y\|} |\phi\rangle \langle\phi| : x \sqsubseteq y, U(y) = |\phi\rangle\}$. Thus $\mathrm{Tr}\nu[x] \leq 2^{\|x\|}$. For $i \in \mathbb{N}$, if $\mathcal{B}[i] = 0$, let ν_i be a 0 matrix, otherwise if $\mathcal{B}[i] = 1$, then $\nu_i = \nu[\mathcal{B}[0..i-1]0]$. This definition makes sense because $\mathcal{B}[0..i-1]0$ is total whenever $\mathcal{B}[i] = 1$. A visual description of ν_i can be seen in Figure B.1.

So $2^{-k} \stackrel{*}{=} \mathrm{Tr}\boldsymbol{\mu}\sigma \stackrel{*}{=} \sum_{i=1}^{\infty} \mathrm{Tr}\nu_i\sigma$. Since the trace of $\nu[i]$ is the weighted sum of prefix free extensions of a string of length i, (or 0), $\mathrm{Tr}\nu_i \leq 2^{-i}$. So $2^{-k} \stackrel{*}{=} \sum_{i=1}^{k+1} \mathrm{Tr}\nu_i\sigma + \sum_{i=k+2}^{\infty} \mathrm{Tr}\nu_i\sigma \stackrel{*}{<} \sum_{i=1}^{k+1} \mathrm{Tr}\nu_i\sigma + 2^{-k-1}$.

So $\sum_{i=1}^{k+1} \mathrm{Tr}\nu_i\sigma \stackrel{*}{>} 2^{-k-1}$. So there is an j, $1 \leq j \leq k+1$ such that $\mathrm{Tr}\nu_j\sigma > c2^{-k-1}/(k+1)$, where c is a positive rational constant solely dependent on the universal Turing machine U. Let j be the smallest index where this occurs.

We define ρ to be equal to $\nu_j/\mathrm{Tr}\nu_j$. ρ is an elementary density matrix, which is computable from the first $j-1$ bits of the border sequence, with $\mathbf{K}(\rho) <^+ \mathbf{K}(\mathcal{B}[0..j-1]) <^+ j + 2\log k$. Since $\mathrm{Tr}\nu_j \leq 2^{-j}$, we have that $-\log \mathrm{Tr}\rho\sigma <^+ k - j + \log k$. So $\mathbf{K}(\rho) - \log \mathrm{Tr}\rho\sigma <^+ k + 3\log k$, proving (1).

It remains to prove that $\mathbf{K}(\mathcal{B}[0..j-1]) <^{\log} \mathbf{I}(\langle\sigma\rangle : \mathcal{H}) + \mathbf{K}(k)$. Using the definition of $\mathbf{I}$, we have that $\mathbf{K}(\mathcal{B}[0..j-1]) - \mathbf{K}(\mathcal{B}[0..j-1] \,|\, \langle\sigma\rangle) - \mathbf{K}(\mathcal{B}[0..j-1] \,|\, \mathcal{H}) <^+ \mathbf{I}(\langle\sigma\rangle : \mathcal{H})$. Since the border sequence is computable from the halting sequence $\mathbf{K}(\mathcal{B}[0..j-1] \,|\, \mathcal{H}) <^+ \mathbf{K}(j)$, therefore it is sufficient to prove $\mathbf{K}(\mathcal{B}[0..j-1] \,|\, \langle\sigma\rangle) <^+ \mathbf{K}(k, j)$.



$$U(x_1) = |\theta_1\rangle$$
$$U(x_2) = \text{Garbage}$$
$$U(x_3) = |\theta_3\rangle$$
$$U(x_4) = |\theta_4\rangle$$
$$U(x_5) = \text{Garbage}$$
$$U(x_6) = |\theta_6\rangle$$

$$\nu[i] = \textstyle\sum_{j\in\{1,3,4,6\}} 2^{-\|x_j\|} |\theta_j\rangle\langle\theta_j|$$

Figure B.1: The above figure shows an example v_i, where $\mathcal{B}[i] = 1$. Two extensions of $\mathcal{B}[0..i-1]0$ do not produce elementary pure states when given as input to the universal Turing machine U. They are x_2 and x_5. The semi density matrix v_i has $\text{Tr}v_i \leq 2^{-i}$ and is the weighted sum of elementary density matrices $|\theta_i\rangle\langle\theta_i|$.

Given k, j and $\langle\sigma\rangle$ there is a program that can enumerate $\nu[x]$ over all total strings x of length j. This program also uses $\langle\sigma\rangle$ to compute from below $\text{Tr}\nu[x]\sigma$, until it finds a string y such that $\text{Tr}\nu[y]\sigma > c2^{-k-1}/(k+1)$. Note that there is a unique string y with this property and $y = \mathcal{B}[0..j-1]0$. Otherwise $y = \mathcal{B}[0..h-1]0z$, for some $h < j$, $\|h\| + \|z\| = j$, with $\mathcal{B}[h] = 1$. However, then $\text{Tr}\nu_h > c2^{-k-1}/(k+1)$, causing a contradiction for the definition of j. Thus the program can find $\mathcal{B}[0..j-1]$, completing the proof for (2). □