Towards a General Theory of Algorithmic Physics

Samuel Epstein
samepst@jptheorygroup.org

July 25, 2023

Abstract

This article presents a survey of published and unpublished material of the intersection of algorithmic tnformation theory, quantum mechanics, and thermodynamics. It is, to the author's knowledge, the first of its type. Three different notions of the algorithmic content of quantum states are reviewed. Notions of algorithmic quantum typicality and mutual information are introduced. The relationship between algorithmic information and quantum measurements is explored. One of the surprising results is that an overwhelming majority of quantum (pure and mixed) states, when undertaking decoherence, will result in a classical probability with no algorithmic information. Thus most quantum states decohere into white noise. A quantum analouge of Martin Löf random sequence is reviewed. Algorithmic Information Theory presents new complications for the Many Worlds Theory, as it conflicts with the Independence Postulate. When algorithmically complicated processes are ruled out, measurements are required to produce distributions over quantum states that have cloneable information.

As for thermodynamics, new definitions of algorithmic coarse and fine grained entropy are introduced. Algorithmic fine grained entropy obeys the chain rule. The function oscillates during the course of dynamics. Small flucuation are common and larger fluctuations are more rare. Coarse grained entropy is shown to be an excellent approximation to fine grained entropy. It is shown to oscillate in the presence of dynamics as well. Synchroniized oscillations occur during the course of discrete dynamics with algorithmic fine grained entropy with respect to two volume measures. Properties of discrete ergodic dynamics are proven, in that, in the limit, a state will be in a partition cell with frequency equal to its measure.

Contents

Part I

Introduction

Chapter 1

Introduction

1.1 Two Brief Looks at Algorithmic Physics

This manuscript deals with the application of algorithmic information theory to physics, namely quantum information theory and thermodynamics. 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main references to this manuscript are [Gó1, Gac94, HR09, SBC01, Vit01, NS19, Vai98, Eve57, Eps20, Vit00, BvL01] and unpublished material from the author. In particular, the references of special import are [Gó1, Gac94] The reference [Gó1] 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 two interesting facts about algorithmic physics that are detailed in this manuscript.

1.1.1 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 6: $\mathbf{I}(\sigma:\rho)$. Almost all pure states $|\psi\rangle$ have low $\mathbf{I}(|\psi\rangle:|\psi\rangle)$. 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 7. Conservation inequalities prevent any type of post-processing of the measured information. As discussed in Chapter 10, the only means to infuse quantum self information is with a quantum measurement.

1.1.2 Algorithmic Thermodynamic Entropy has Oscillations.

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 14.

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 are constants c_1 and c_2 with

$$2^{-n-\mathbf{K}(n)-c_1} < L\{t \in [0,1] : \mathbf{H}_\mu(G^t\alpha) < \log \mu(\mathcal{X}) - n\} < 2^{-n+c_2}.$$

1.2 Conventions

The following section details the conventions in algorithmic information theory, which will be used in both the quantum mechanics section and the thermodynamics section. We use $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\{0,1\}$, $\{0,1\}^*$, and $\{0,1\}^\infty$ to denote natural numbers, integers, rational numbers, reals, bits, finite strings, and infinite sequences. $\{0,1\}^{*\infty} \stackrel{\mathrm{def}}{=} \{0,1\}^* \cup \{0,1\}^\infty$. $\|x\|$ 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. $[A] \stackrel{\mathrm{def}}{=} 1$ if statement A holds, else $[A] \stackrel{\mathrm{def}}{=} 0$. The set of finite bit-strings is denoted by $\{0,1\}^*$. 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.

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.

1.2.1 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 $\mathbf{K}_T(x|y) \stackrel{\mathrm{def}}{=} \inf\{\|p\| : T_y(p) = x\}$.

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 $y \in \{0,1\}^{*\infty}$. $\mathbf{K}(x|y) \stackrel{\mathrm{def}}{=} \mathbf{K}_U(x|y)$ is the Kolmogorov complexity of $x \in \{0,1\}^*$ relative to $y \in \{0,1\}^{*\infty}$.

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 $\mathcal{H}_i \stackrel{\text{def}}{=} [U(i)$ 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, [Gá01], $\mathbf{d}(x|P) =^+ \log \mathbf{t}_P(x)$. 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 [Gá01]) $\mathbf{d}(f(x)|fQ) <^+ \mathbf{d}(x|Q)$.

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 ([Eps22]) $\mathbf{I}(f(a);\mathcal{H}) <^+ \mathbf{I}(a;\mathcal{H}) + \mathbf{K}(f)$.

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

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

□

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 23 of Appendix B.

Lemma 2 ([Lev16, Eps21b]) $\mathbf{Ks}(x) <^{\log} \mathbf{I}(x;\mathcal{H})$.

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 $\mathbf{I}(f(\alpha):\beta) <^+ \mathbf{I}(\alpha:\beta) + \mathbf{K}(f)$.

1.2.2 Algorithmic Information Between Probabilities

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

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

Definition 3 (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 3 *Let ψ_a be an enumerable semi-measure, semi-computable relative to a.
$\sum_c 2^{\mathbf{I}(\langle a,c\rangle:b)} \psi_a(c) \overset{*}{<} 2^{\mathbf{I}(a:b)}/\mathbf{m}(\psi)$.*

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 $\mathbf{m}(a,b)/(\mathbf{m}(a)\mathbf{m}(b)) \overset{*}{>} \sum_c (\mathbf{m}(a,b,c)/(\mathbf{m}(b)\mathbf{m}(a,c)))\mathbf{m}(\psi)\psi_a(c)$, or $\sum_c (\mathbf{m}(a,b,c)/\mathbf{m}(a,c))\mathbf{m}(c|a) \overset{*}{<} \mathbf{m}(a,b)/\mathbf{m}(a)$, since $\mathbf{m}(c|a) \overset{*}{>} \mathbf{m}(\psi)\psi_a(c)$. Rewrite it $\sum_c \mathbf{m}(c|a)\mathbf{m}(a,b,c)/\mathbf{m}(a,c) \overset{*}{<} \mathbf{m}(a,b)/\mathbf{m}(a)$ or $\sum_c \mathbf{m}(c|a)\mathbf{m}(a)\mathbf{m}(a,b,c)/\mathbf{m}(a,c) \overset{*}{<} \mathbf{m}(a,b)$. The latter is obvious since $\mathbf{m}(c|a)\mathbf{m}(a) \overset{*}{<} \mathbf{m}(a,c)$ and $\sum_c \mathbf{m}(a,b,c) \overset{*}{<} \mathbf{m}(a,b)$. □

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

Proof. Using Lemma 4,

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

Using Lemma 5,

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

□

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 $\mathbf{I}_{\mathrm{Prob}}(p:p)$ 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 $\mathbf{I}_{\mathrm{Prob}}(p:p) = \mathbf{K}(x)$.*
- *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 $\mathbf{I}_{\mathrm{Prob}}(p:p) = \infty$.*

- *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* $\mathbf{I}_{\mathrm{Prob}}(p:p) = \mathbf{K}(x)$*, 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 2

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 survey 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 $\mathbf{I}(|\psi\rangle : \psi) \approx 0$. This has implications on the algorithmic content of measurements and decoherence. The main quantum mechanical areas covered in this manuscript are

- **Chapter 3:** This chapter covers the background material needed for the article. A new definition, the algorithmic information between probabilities, is introduced and shown to have information non-growth properties with respect to randomized processing.
- **Chapter 4:** Three different algorithmic measures of quantum states are covered. Their properties are described, including an addition inequality, a Quantum EL Theorem, and a generalized no-cloning theorem. Multiple relationships between the complexities are proven.
- **Chapter 5:** 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 6:** 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 7:** 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. An overwhelming majority of quantum pure states , when undertaking decoherence, will result in a classical probability with no algorithmic information.
- **Chapter 8:** 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 9:** This chapter shows the Many Worlds Theory and AIT are in conflict, as shown through the existence of a finite experiment that measures the spin of a large number of electrons. After the experiment there are branches of positive probability which contain forbidden sequences that break the Independence Postulate, a postulate in AIT.
- **Chapter 10:** 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:** Properties of the quantum information of basis states are proven.
- **Appendix B:** An extended coding theorem is proved with applications to proving inequalities of quantum complexities and the relation between dynamics and coarse grained entropy.

Chapter 3

Background

The reader is assumed to be familiar with both algorithmic information theory and quantum information theory, but we review the core terms. We use $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\{0,1\}$, $\{0,1\}^*$, and $\{0,1\}^\infty$ to denote natural numbers, integers, rational numbers, reals, 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 $[A] \stackrel{\text{def}}{=} 0$. The set of finite bit-strings is denoted by $\{0,1\}^*$. 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.

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.

3.1 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 $y \in \{0,1\}^{*\infty}$.

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, $\mathbf{t}_P(x) \overset{*}{>} \mathbf{m}(t)t(x)$. 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 4 (See [GÓ1]) $\mathbf{d}(f(x)|fQ) <^+ \mathbf{d}(x|Q)$.

Theorem 5 ([Lev84]) $\mathbf{I}(f(x) : y) <^+ \mathbf{I}(x : y)$.

Lemma 4 ([Eps22]) $\mathbf{I}(f(a);\mathcal{H}) <^+ \mathbf{I}(a;\mathcal{H}) + \mathbf{K}(f)$.

Lemma 5 *Let ψ_a be an enumerable semi-measure, semi-computable relative to a.* $\sum_c 2^{\mathbf{I}(\langle a,c\rangle:b)}\psi_a(c) \overset{*}{<} 2^{\mathbf{I}(a:b)}/\mathbf{m}(\psi)$.

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 $\mathbf{m}(a,b)/(\mathbf{m}(a)\mathbf{m}(b)) \overset{*}{>} \sum_c(\mathbf{m}(a,b,c)/(\mathbf{m}(b)\mathbf{m}(a,c)))\mathbf{m}(\psi)\psi_a(c)$, or $\sum_c(\mathbf{m}(a,b,c)/\mathbf{m}(a,c))\mathbf{m}(c|a) \overset{*}{<} \mathbf{m}(a,b)/\mathbf{m}(a)$, since $\mathbf{m}(c|a) \overset{*}{>} \mathbf{m}(\psi)\psi_a(c)$. Rewrite it $\sum_c \mathbf{m}(c|a)\mathbf{m}(a,b,c)/\mathbf{m}(a,c) \overset{*}{<} \mathbf{m}(a,b)/\mathbf{m}(a)$ or $\sum_c \mathbf{m}(c|a)\mathbf{m}(a)\mathbf{m}(a,b,c)/\mathbf{m}(a,c) \overset{*}{<} \mathbf{m}(a,b)$. The latter is obvious since $\mathbf{m}(c|a)\mathbf{m}(a) \overset{*}{<} \mathbf{m}(a,c)$ and $\sum_c \mathbf{m}(a,b,c) \overset{*}{<} \mathbf{m}(a,b)$. □

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 stochsticity measures are exotic, in that they have high mutual information with the halting sequence.

Lemma 6 ([Lev16, Eps21b]) $\mathbf{Ks}(x) <^{\log} \mathbf{I}(x;\mathcal{H})$.

3.1.1 Algorithmic Information Between Probabilities

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

Definition 4 (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 5 (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, $fp(x) = \sum_z f(x|z)p(z)$.*

Theorem 6 *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 4,

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

Using Lemma 5,

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

□

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 3

- *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 4 (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.*

3.2 Quantum Mechanics

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 $B^X = \text{Tr}_Y B$, $\text{Tr}(A \cdot B^X) = \text{Tr}((A \otimes I) \cdot B)$, which is used frequently throughout the paper. For positive 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 $\text{Tr}AC \leq \text{Tr}BC$. 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 $S(\sigma) = -\sum p_i \log p_i$.

A pure quantum state $|\phi\rangle$ and (semi)density matrix σ are called *elementary* if their real and imaginary components have rational coefficients. 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.

3.2.1 Quantum Operations

A quantum operation is the most general type of operation than can be applied to a quantum state. In Chapters 5 and 6, 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 $\mathrm{Tr}(\sigma) = \mathrm{Tr}(\varepsilon(\sigma))$.
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

$$\varepsilon(\sigma_A) = \mathrm{Tr}_B \left(U(\sigma_A \otimes \sigma_B)U^* \right). \tag{3.1}$$

A third way to characterize a quantum operation is through Kraus operators, which can be derived using an algebraic reworking of Equation 3.1. 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 $\sum_i M_i^* M_i \leq I$, where I is the identity matrix.

A quantum operation ε is elementary iff it admits a represented of the form in Equation 3.1 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 4

Quantum Complexity

4.1 Three Meausres of the Algorithmic Content of Individual 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, G0́1, Vit01]. BVL complexity [BvL01] measures the complexity of a pure quantum state $|\psi\rangle$ by the length of the smallest input to a universal quantum Turing machine that outputs a good approximation of $|\psi\rangle$. 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 analouge of of the universal semi-measure $\mathbf{m}$.

4.1.1 BVL Complexity

BVL complexity, introduced in [BvL01], uses a universal quantum Turing machine to define the complexity of a pure quantum state. The input and output tape of this machine consists of symbols of the type 0, 1, and #. The input is an ensemble $\{p_i\}$ of pure states $|\psi_i\rangle$ of the same length n, where $p_i \geq 0$ and $\sum_i p_i = 1$. This ensemble can be represented as a mixed state of n qubits. If, during the operation of the quantum Turing machine, all computational branches halt at a time t, then the contents on the output tape are considered the output of the quantum Turing machine. The BVL Complexity of a pure state, $\mathbf{Hbvl}[\epsilon](|\psi\rangle)$ is the size of the smallest (possibly mixed state) input to a universal quantum Turing machine such that fidelity between the output and $|\psi\rangle$ is at least ϵ. The fidelity between a mixed state output σ and $|\psi\rangle$ is $\langle\psi|\,\sigma\,|\psi\rangle$. We require that the input quantum state be elementary. We also require that universal quantum Turing machine be conditioned on the number of qubits n, on a classical auxillary tape.

4.1.2 Vitányi Complexity

Vitányi complexity [Vit01] is a measure of the algorithmic information content of a pure state $|\psi\rangle$. It is equal to the minimum of the Kolmogorov complexity of an approximating elementary pure state $|\phi\rangle$ summed with a score of their closeness. We use a slightly different definition than the original [Vit01], in that we use a classical machine and not a quantum machine for the approximation. Let N be the dimension of the Hilbert space.

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

4.1.3 Gács Complexity

Gács complexity [GÓ1] is defined using the following universal lower computable semi-density matrix, parametered by $x \in \{0,1\}^*$, with

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

The parameter N represents the dimension of the Hilbert space. We use $\boldsymbol{\mu}_X$ to denote the matrix $\boldsymbol{\mu}$ over the Hilbert space denoted by symbol X. 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 $\mathbf{Hg}(\sigma)$. This definition generalizes $\underline{H}$ in [GÓ1] to mixed states. Note than in [GÓ1], there is another measure of quantum algorithmic entropy, $\overline{H}$, which we will not cover in this paper. An infinite version of algorithmic entropy can be found at [BOD14].

4.2 Properties of Universal Matrix $\boldsymbol{\mu}$ 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 6. 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 $\mathbf{m}(x) \overset{*}{>} \mathbf{m}(p)p(x)$. This theorem is used throughout the paper, and will not be explicitly cited.

Theorem 7 ([GÓ1]) *Let $q \in \{0,1\}^*$, and the dimension of the Hilbert space, N, compute lower compute semi-density matrix A. Then $\boldsymbol{\mu} \overset{*}{>} \mathbf{m}(q|N)A$.*

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 computable from q. Thus since p is computable from q and N,

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

□

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

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

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

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, $\mathrm{Tr}_T |\psi\rangle\langle\psi|$, and add the resulting pure or mixed state to the sum ρ. Thus $\rho \overset{*}{<} \boldsymbol{\mu}_X$. Let $\sigma = \boldsymbol{\mu}_X \otimes |\psi\rangle\langle\psi|$, where $|\psi\rangle$ is a reference elementary state. Thus $\sigma \overset{*}{<} \boldsymbol{\mu}_{XY}$ so

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

□

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,

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

4.3 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. Due to the no-cloning theorem, [WZ82] one cannot clone a quantum state. The following theorem generalizes this no-go result, by showing there exist tensor products $|\psi\rangle^m$ that has signifigantly more $\mathbf{Hg}$ measure than $|\psi\rangle$. The following theorem presents a new proof to this result.

Theorem 11 ([Gó1]) $\log\binom{m+N-1}{m} <^+ \max_{|\psi\rangle}\mathbf{Hg}(|\psi\rangle^{\otimes m}) <^+ \mathbf{K}(m) + \log\binom{m+N-1}{m}$

Proof. 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 $\{1,\dots,N\}$. This is because for each such multiset $S=\{i_1,\dots,i_m\}$, 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

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

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

$$\mathrm{Tr}\boldsymbol{\mu}|\psi\rangle^m\langle\psi|^m \overset{*}{>} 2^{-c}. \tag{4.1}$$

Let Λ be the uniform distribution on the unit sphere of $\mathcal{H}_N$. 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 $U^m |\psi\rangle^m = |\rho\rangle^m$, 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 4.1, by $d\Lambda$ results in

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

□

4.4 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, $\mathbf{Hg}$, since as shown in Theorem 11, 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 $\mathbf{Hg}$.

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

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

using Theorem 9, which states $\mathrm{Tr}_Y\boldsymbol{\mu}_{XY} \overset{*}{=} \boldsymbol{\mu}_X$. 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 $\overset{*}{=} 2^{-\mathbf{Hg}(\rho)}$, 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 $2^{\mathbf{Hg}(\rho)}M_{\boldsymbol{\mu}_{2n}\rho} \overset{*}{<} \boldsymbol{\mu}_{\langle\rho,\mathbf{Hg}(\rho)\rangle}$. So we have that $-\log\mathrm{Tr}2^{\mathbf{Hg}(\rho)}M_{\boldsymbol{\mu}_{2n}\rho}\sigma = -\mathbf{Hg}(\rho) - \log T(\sigma, \rho) =^+ \mathbf{Hg}(\sigma \otimes \rho) - \mathbf{Hg}(\rho) >^+ -\log\boldsymbol{\mu}_{\langle\rho,\mathbf{Hg}(\rho)\rangle}\sigma =^+ \mathbf{Hg}(\sigma|\langle\rho, \mathbf{Hg}(\rho)\rangle)$. □

4.5 Subadditivity, Strong Subadditivity, Strong Superadditivity

Theorem 13 ([G0́1]) $\mathbf{Hg}(\sigma)$ *is subadditive, with* $\mathbf{Hg}(\sigma \otimes \rho) <^+ \mathbf{Hg}(\sigma) + \mathbf{Hg}(\rho)$.

Proof.

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

□

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}, $\mathbf{L}(\rho_{12}) + \mathbf{L}(\rho_{23}) < \mathbf{L}(\rho_{123}) + \mathbf{L}(\rho_2) + c$.

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 **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. □

4.6 Relation Between Complexities

4.6.1 Vitányi Complexity and Gács Complexity

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 $| \langle\psi|u\rangle |^2 \overset{*}{>} 2^{-2k}$, so

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

□

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 $\langle\langle|\psi\rangle\rangle\rangle$ 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, $\mathbf{I}$, with the halting sequence $\mathcal{H} \in \{0,1\}^\infty$.

Lemma 7 *For pure quantum state* $|\psi\rangle$,
$\min_{|\phi\rangle} \mathbf{K}(|\phi\rangle) - \log | \langle\psi|\phi\rangle |^2 <^{\log} -\log \sum_{|\phi\rangle} \mathbf{m}(|\phi\rangle)| \langle\psi|\phi\rangle |^2 + \mathbf{I}(\langle\langle|\psi\rangle\rangle : \mathcal{H})$.

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

$$\begin{aligned}
\min_{|\phi\rangle}\mathbf{K}(|\phi\rangle) - \log|\langle\psi|\phi\rangle|^2 &<^+ \min_{\theta\in\mathcal{D}}\mathbf{K}(|\theta\rangle) + f(|\theta\rangle) \\
&<^{\log} -\log\sum_{|\theta\rangle\in\mathcal{D}}\mathbf{m}(|\theta\rangle)2^{-f(|\theta\rangle)} + \mathbf{I}(\langle f\rangle;\mathcal{H}). &(4.2)\\
&<^{\log} g + \mathbf{I}(\langle f\rangle;\mathcal{H}) &(4.3)\\
&<^{\log} g + \mathbf{I}(\langle|\psi\rangle\rangle:\mathcal{H}) + \mathbf{K}(\langle f\rangle|\langle|\psi\rangle\rangle) &(4.4)\\
&<^{\log} g + \mathbf{I}(\langle|\psi\rangle\rangle:\mathcal{H}) + \mathbf{K}(g)\\
&<^{\log} -\log\sum_{|\phi\rangle}\mathbf{m}(|\phi\rangle)|\langle\psi|\phi\rangle|^2 + \mathbf{I}(\langle|\psi\rangle\rangle:\mathcal{H}).
\end{aligned}$$

Inequality 4.2 is due to Theorem 61. Inequality 4.3 is due to the definition of f and $\mathcal{D}$. Inequality 4.4 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 $\mathbf{Hg}(|\psi\rangle) <^+ \mathbf{Hv}(|\psi\rangle) <^{\log} \mathbf{Hg}(|\psi\rangle) + \mathbf{I}(|\psi\rangle:\mathcal{H}|n)$.

Proof. This follows directly from Lemma 7, relativized to n, and the fact that $\mathbf{Hg}(|\psi\rangle) =^+ -\log\mathrm{Tr}\boldsymbol{\mu}\,|\psi\rangle\langle\psi| =^+ -\log\sum_{|\phi\rangle}\mathbf{m}(|\phi\rangle|n)|\langle\phi|\psi\rangle|^2$.

4.6.2 BVL Complexity and Gács Complexity

Theorem 18 ([Eps20]) *For pure state* $|\psi\rangle\in\mathcal{Q}_n$, $\mathbf{Hg}(|\psi\rangle) <^+ \mathbf{Hbvl}[\epsilon](|\psi\rangle) + \mathbf{K}(\mathbf{Hbvl}^\epsilon(|\psi\rangle) - \log\epsilon$.

Proof. For each k and t in $\mathbb{N}$, let $\mathcal{H}_{k,t}$ be the smallest linear subspace spanning elementary k-qubit inputs to the universal quantum Turing machine M of size k that halt in t steps, outputing a n qubit mixed state. As shown in [Mul08], if $t\neq t'$, then $\mathcal{H}_{k,t}$ is perpendicular to $\mathcal{H}_{k,t'}$. Let $P_{k,t}$ be the projection onto $\mathcal{H}_{k,t}$. For each k and t, the universal quantum Turing machine defines a completely positive map $\Psi_{k,t}$ over $\mathcal{H}_{k,t}$, where $\Psi_{k,t}(\nu) = \rho$ implies that the quantum Turing machine, with semi-density matrix ν of length k on the input tape will output the n qubit mixed state ρ and halt in time t. Let ρ be a k qubit mixed state that minimizes $k = \mathbf{Hbvl}[\epsilon](|\psi\rangle)$ in time

t.

$$\rho \leq P_{k,t}$$
$$2^{-k}\rho \leq 2^{-k}P_{k,t}$$
$$\Psi_{k,t}2^{-k}\rho \leq \Psi_{k,t}2^{-k}P_{k,t}$$
$$\Psi_{k,t}2^{-k}\rho \leq \sum_t \Psi_{k,t}2^{-k}P_{k,t}$$

The semi density matrix $\sum_t \Psi_{k,t}2^{-k}P_{k,t}$ is lower computable relative to k, so

$$\mathbf{m}(k|N)2^{-k}\Psi_{k,t}\rho \leq \mathbf{m}(k|N)\sum_t \Psi_{k,t}2^{-k}P_{k,t} \overset{*}{<} \boldsymbol{\mu}$$

$$\mathbf{m}(k|N)2^{-k}\langle\psi|\,\Psi_{k,t}(\rho)\,|\psi\rangle \overset{*}{<} \langle\psi|\,\boldsymbol{\mu}\,|\psi\rangle$$
$$k + \mathbf{K}(k|N) - \log\epsilon >^{+} \mathbf{Hg}(|\psi\rangle).$$

□

Theorem 19 ([Eps20]) $\mathbf{Hbvl}[2^{-\mathbf{Hg}(|\psi\rangle)-O(\log\mathbf{Hg}(|\psi\rangle))}](|\psi\rangle) <^{\log} \mathbf{Hg}(|\psi\rangle)$.

Proof. We use reasoning from Theorem 7 in [Gács01]. From Theorem 7 in [Eps20] there exists a ρ such that $\mathbf{K}(\rho|N) - \log\langle\psi|\,\rho\,|\psi\rangle <^{\log} \mathbf{Hg}(|\psi\rangle)$. 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 program to the universal quantum Turing machine that will construct $\rho_{2^{m+2}}$. The input 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 $<^{+} \mathbf{K}(p|N) + m <^{\log} \mathbf{Hg}(|\psi\rangle)$. The fidelity of the approximation is $\langle\psi|\,\rho_{2^{m+2}}\,|\psi\rangle > 2^{-m-1} \geq 2^{-\mathbf{Hg}(|\psi\rangle)-O(\log\mathbf{Hg}(|\psi\rangle))}$. □

4.7 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.

Theorem 20 ([Lev16, Eps19c])
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 non exotic projections of large rank must have simple quantum pure states in their images. By non exotic, we mean the coding of the projection has low information with the halting sequence. 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 21 (Quantum EL Theorem [Eps23b]) *Fix an n qubit Hilbert space. Let P be a elementary projection of rank $> 2^m$. Then, relativized to (n,m), $\min_{|\phi\rangle\in\mathrm{Image}(P)}\mathbf{Hv}(|\phi\rangle) <^{\log} 3(n-m)+\mathbf{I}(\langle P\rangle;\mathcal{H})$.*

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 a projection of rank 2^m. We sample N independent pure states according to the uniform distribution Λ on the n qubit space. 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

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

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

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

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 6,

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

□

4.7.1 Computable Projections

Theorem 26 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 1 ([Eps23b]) *Fix an n qubit Hilbert space. Let P be a computable projection of rank $> 2^m$. Then, relativized to (n,m), $\min_{|\phi\rangle \in \mathrm{Image}(P)} \mathbf{Hv}(|\phi\rangle) <^{\log} 3(n-m) + \mathbf{I}(P;\mathcal{H})$.*

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 $\min_{|\psi\rangle \in \mathrm{Image}(P)} \mathbf{Hv}(|\psi\rangle) =^+ \min_{|\psi\rangle \in \mathrm{Image}(P_n)} \mathbf{Hv}(|\psi\rangle)$. Thus by Lemma 4, one gets that $\mathbf{I}(P_n;\mathcal{H}) <^+ \mathbf{I}(P;\mathcal{H})$. The corollary follows from Theorem 26. □

4.7.2 Quantum Data Compression

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 5

Quantum Typicality

5.1 Definition of Quantum Randomness Deficiency

In [G01́], 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 [G2́1]). 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 6 (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.

Theorem 22 *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 $T \overset{*}{<} \mathbf{t}_\sigma$ 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 23 ([G0́1]) *Relativized to elementary σ, $\log \mathbf{d}(\rho|\sigma) =^+ \log \mathrm{Tr}\rho\sigma^{-1/2}\boldsymbol{\mu}\sigma^{-1/2}\sigma$*

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

5.1.1 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 7 (Quantum Randomness Deficiency (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 7 equals Definition 6. By definition, $\mathfrak{T}_\sigma$ is universal, since for every lower computable σ-test ν, $\underline{\mathbf{m}}(\nu)\nu < \mathfrak{T}_\sigma$.

Theorem 24 *For n qbit density matrices σ, ρ, ν, and ξ, $\mathbf{d}(\sigma|\rho) + \mathbf{d}(\nu|\xi) <^+ \mathbf{d}(\sigma\otimes\nu|\rho\otimes\xi)$.*

Proof.

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

□

5.2 Conservation Over Quantum Operations

Proposition 1 *For semi-density matrix ν, relativized to a finite set of elementary matrices $\{M_i\}$, $\underline{\mathbf{m}}(\sum_i M_i^*\nu M_i|N) \stackrel{*}{>} \underline{\mathbf{m}}(\nu|N)$.*

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, $\mathbf{m}(q_M|N) \stackrel{*}{>} \mathbf{m}(q|N)$.

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

□

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

Theorem 25 (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 1, $\underline{\mathbf{m}}(\sum_i M_i^*\nu M_i|n) \stackrel{*}{>} \underline{\mathbf{m}}(\nu|n)$. So we have the following inequality

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

□

5.3 A Quantum Outlier Theorem

One recent result in the classical randomness deficiency case is that sampling methods produce outliers [Eps21b]. There are several proofs to this result, with one of them derived from the fact that large sets of natural numbers with low randomness deficiencies are exotic, in that they have high mutual information with the halting sequence.

In this paper, we prove a quantum version of this result. Projections of large rank must contain pure quantum states in their images that are outlying states. Otherwise, the projections are exotic, in that they have high mutual information with the halting sequence. Thus quantum coding schemes that use projections, such as Schumacher compression, must communicate using outlier quantum states. The classical and quantum theorems are analogous, but their proofs are very different!

Theorem 26 ([Eps23c]) *Relativized to an n qubit mixed state σ, for elementary 2^m rank projector P, $3m-2n <^{\log} \max_{|\phi\rangle\in\mathrm{Image}(P)} \mathbf{d}(|\phi\rangle\,|\sigma) + \mathbf{I}(\langle P\rangle;\mathcal{H})$.*

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, $\mathrm{Tr}\sigma T\leq 1$, and for a certain Q-probability of projections B, $\mathrm{Tr}TB$ 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.

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

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,

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

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 $\mathrm{Tr}AB > 2^{2m-n-3}$, and 0 otherwise. By the above equations, for all $A\in\mathrm{Support}(Q)$, $\int_\Omega [A,B]dR(B)\geq .5p$. So $\sum_A\int_\Omega [A,B]Q(A)dR(B)\geq .5p$. For Hermitian matrix A, $\{A\}$ is 1 if $\mathrm{Tr}A\sigma\leq 1$, 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 $\sum_R\{Q(R) : \mathrm{Tr}T_iR \geq 2^{2m-n-3}\} \geq p/4 = 2^{m-n-2}$. 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

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

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 6,

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

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 $(3m-2n)$. □

5.3.1 Computable Projections

Theorem 26 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 is a program that computes $P\}$.

Corollary 2 ([Eps23c]) *Relativized to an n qubit mixed state σ, for computable 2^m rank projector P, $3m - 2n <^{\log} \max_{|\phi\rangle \in \mathrm{Image}(P)} \mathbf{d}(|\phi\rangle\,|\sigma) + \mathbf{I}(\langle P\rangle;\mathcal{H})$.*

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 $\max_{|\psi\rangle\in\mathrm{Image}(P)} \mathbf{d}(|\psi\rangle\,|\sigma) =^+ \max_{|\psi\rangle\in\mathrm{Image}(P_n)} \mathbf{d}(|\psi\rangle\,|\sigma)$. Thus by Lemma 4, one gets that $\mathbf{I}(P_n;\mathcal{H}) <^+ \mathbf{I}(P;\mathcal{H})$. The corollary follows from Theorem 26. $\square$

Chapter 6

Quantum Information

6.1 Definition of Quantum Algorithmic Information

For a pair of random variables, $\mathcal{X}$, $\mathcal{Y}$, their mutual information is defined to be $\mathbf{I}(\mathcal{X}:\mathcal{Y}) = \mathcal{H}(\mathcal{X}) + \mathcal{H}(\mathcal{Y}) - \mathcal{H}(\mathcal{X},\mathcal{Y}) = \mathcal{H}(\mathcal{X}) - \mathcal{H}(\mathcal{X}/\mathcal{Y}) = \sum_{x,y} p(x,y)\log p(x,y)/p(x)p(y)$. This represents the amount of correlation between $\mathcal{X}$ and $\mathcal{Y}$. 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 $\mathcal{Y}$.

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.

As stated in Chapter 3, 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.

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 8 (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 27 $\mathbf{I}(\sigma:\rho) =^+ \mathbf{I}(\rho:\sigma)$.

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. □

6.2 Paucity of Self-Information

6.2.1 Pure States

For classical algorithmic information, for all $x \in \{0,1\}^*$,

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

As shown in this section, this property differs from the quantum case, where there exists quantum states with high descriptional complexity and negligible self information. In fact this is the case for most quantum states, where for most n qubit pure states $|\psi\rangle$,

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

The following theorem states that the information between two elementary states is not more than the combined length of their descriptions.

Theorem 28 *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

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

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

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

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

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

causing a contradiction. □

Theorem 29 ([Eps19b]) *Let Λ be the uniform distribution on the unit sphere of $\mathcal{H}_N$.*

1. $\mathbf{Hg}(I/N) =^+ \log N$,
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) is due to Theorem 28, with $\mathbf{I}(I/N : I/N) <^+ 2\mathbf{K}(I/N|N) <^+ 0$. (3) uses 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 $\int 2^{-\mathbf{Hg}(|\psi\rangle)} d\Lambda \overset{*}{=} \int \mathrm{Tr}\boldsymbol{\mu} |\psi\rangle\langle\psi| \, d\Lambda \overset{*}{=} \mathrm{Tr}\boldsymbol{\mu} \int |\psi\rangle\langle\psi| \, d\Lambda \overset{*}{=} N^{-1}$. (4) uses the proof of Theorem 11, 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

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

□

6.2.2 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 $\mathcal{H}_N$. 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 30 $\int 2^{\mathbf{I}(\sigma:\sigma)} d\mu(\sigma) <^+ 0$.

Proof.

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

The first term is not greater than

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

At this point, reasoning from the proof of Theorem 29 can be used to show that this term is $O(1)$. The second term is not greater than

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

Again, at this point, reasoning from the proof of Theorem 29 can be used to show that this term is $O(1)$.

6.3 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 5). Conservation inequalities have been extended to probabilistic transforms and infinite sequences. 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 31 (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 $\sum_i M_i\boldsymbol{\mu}M_i^*\overset{*}{<}\boldsymbol{\mu}$ and also $(\sum_i M_i\boldsymbol{\mu}M_i^*\otimes\boldsymbol{\mu})\overset{*}{<}\boldsymbol{\mu}\otimes\boldsymbol{\mu}$. 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 25 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

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

□

6.3.1 Algorithmic No-Cloning Theorem

The no-cloning theorem states that every unitary transform cannot clone an arbitrary quantum state. Hiowever 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 11 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 32 **([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 31. $\square$

Theorem 32, combined with the paucity of self-information in pure states (Theorem 29) shows that only a very sparse set of pure states can, given any unitary transform, duplicate algorithmic information.

6.3.2 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 3 *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 4 *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 25) and information (Theorem 31) over quantum operations, which includes the partial trace function.

6.3.3 Decoherence

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.

The single qubit example is as follows. The system is in state $|\psi_Q\rangle = \alpha|0\rangle + \beta|1\rangle$ and the environment is in state $|\psi_E\rangle$. The initial state is $|\psi_{QE}\rangle\,|\psi_Q\rangle \otimes |\psi_E\rangle = \alpha|0,\psi_E\rangle + \beta|1,\psi_E\rangle$. The combined system undergoes a unitary evolution U, becoming entangled, with the result $U|\psi_{QE}\rangle = \alpha|0,E_1\rangle + \beta|1,E_2\rangle$. The density matrix is $\rho_{QE} = |\alpha|^2|0,E_1\rangle\langle 0,E_1| + |\beta|^2|1,E_2\rangle\langle 1,E_2| + \alpha^*\beta|1,E_2\rangle\langle 0,E_1| + \alpha\beta^*|0,E_1\rangle\langle 1,E_2|$. The partial trace over the environment yields

$$\rho_Q = |\alpha|^2|0\rangle\langle 0|\langle E_1|E_1\rangle + |\beta|^2|1\rangle\langle 1|\langle E_2|E_2\rangle + \alpha^*\beta|1\rangle\langle 0|\langle E_2|E_1\rangle + \alpha\beta^*|0\rangle\langle 1|\langle E_1|E_2\rangle.$$

We have $\langle E_1|E_1\rangle = \langle E_2|E_2\rangle = 1$. Two environment-related terms are time dependent and can be described by an exponential decay function

$$\langle E_1|E_2\rangle = e^{-\gamma(t)}.$$

The larger the decay, the more off diagonal terms are suppressed. So

$$\rho_Q \approx \begin{pmatrix} |\alpha|^2 & \alpha^*\beta e^{-\gamma(t)} \\ \alpha\beta^* e^{-\gamma(t)} & |\beta|^2 \end{pmatrix}.$$

The above example can be generalzied to n qubit density matrices. Let $\mathrm{Decohere}(\sigma, t)$ be a decoherence operation that dampens the off-diagonal elements of σ with decay t. By definition, Decohere is a quantum operation. Randomness is conserved over decoherence. Thus if two states decohere, the first state does not increase in algorithmic atypicality with respect to the second state.

Corollary 5 $\mathbf{d}(\mathrm{Decohere}(\sigma, t)|\mathrm{Decohere}(\rho, t)) <^+ \mathbf{d}(\sigma|\rho)$.

This is a corollary to Theorem 25. When a state loses coherence into the environment will not gain information with any other state.

Corollary 6 *For semi-density matrices σ and ρ,* $\mathbf{I}(\mathrm{Decohere}(\sigma, t) : \mathrm{Decohere}(\rho, t)) <^+ \mathbf{I}(\sigma : \rho)$.

Chapter 7

Quantum Measurements

In quantum mechanics, measurements are modeled by POVMs. A POVM E is a finite or infinite set of positive definite matrices $\{E_k\}$ such that $\sum_k E_k = I$. 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, $\Phi_E(\sigma) = \sum_k E(\sigma(k)) |k\rangle \langle k|$.

7.1 Typicality and Measurements

Theorem 33 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.

Definition 9 (Deficiency, probabilities (Folkflore)) *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 $E\rho$,.

Theorem 33 ([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 $\underline{\mathbf{m}}(\nu|N) \overset{*}{>} 1$, $\mathbf{d}(E\sigma|E\rho) <^+ \mathbf{d}(\sigma|\rho)$.

7.2 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 4, 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 34 *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,

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

Corollary 7 *For density matrices ρ and σ, and $i,j \in \mathbb{N}$, relativized to POVMS E and F, $\mathbf{I}(i : j) + \log E_i(\rho)F_j(\sigma) <^{\log} \mathbf{I}(\rho : \sigma)$.*

7.3 Algorithmic Contents of Measurements

This sections shows the limitations of the algorithmic content of measurements of pure quantum states. Theorem 35 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 6 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 $\mathcal{I}$ 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 p 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 $\mathfrak{G}_{C\otimes D} = \sum_{A\otimes B\in\mathcal{C}_{C\otimes D}} \overline{\mathbf{m}}(A\otimes B/n)(A\otimes B)$ be an aggregation of upper computable $C\otimes D$ tests of the form $A\otimes B$, weighted by their upper probability.

Definition 10 *The upper information between semi-density matrices A and B is $\mathcal{I}(A : B) = \log \mathrm{Tr}\mathfrak{G}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(A\otimes B)$.*

Proposition 2 *$\mathcal{I}(I/N : I/N) = O(1)$.*

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

Lemma 8

- *Let* Λ *be the uniform distribution on the unit sphere of an* n *qubit space.* $\int 2^{\mathcal{I}(|\psi\rangle\,:\,|\psi\rangle)} d\Lambda = O(1)$,
- $\int 2^{\mathcal{I}(\sigma\,:\,\sigma)} d\mu(\sigma) = O(1)$.

Proof. The proof follows identically to that of Theorems 29 and 30, with reference to Proposition 2.

Lemma 9 **([Eps21a])** *Relativized to POVM* E, $\mathbf{I}_{\mathrm{Prob}}(E\sigma{:}E\sigma) <^{+} \mathcal{I}(\sigma{:}\sigma)$.

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

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

□

Theorem 35 **([Eps21a])** *Let* Λ *be the uniform distribution on the unit sphere of an* n *qubit space. Relativized to POVM* E, $\int 2^{\mathbf{I}_{\mathrm{Prob}}(E|\psi\rangle:E|\psi\rangle)} d\Lambda = O(1)$.

Proof. By Lemma 9, $2^{\mathcal{I}(|\psi\rangle:|\psi\rangle)} \overset{*}{>} 2^{\mathbf{I}_{\mathrm{Prob}}(E|\psi\rangle:E|\psi\rangle)}$. From Lemma 8, $\int 2^{\mathcal{I}(|\psi\rangle:|\psi\rangle)} d\Lambda = O(1)$. The integral $\int 2^{\mathbf{I}_{\mathrm{Prob}}(E|\psi\rangle:E|\psi\rangle)} d\Lambda$ is well defined because $2^{\mathbf{I}_{\mathrm{Prob}}(E|\psi\rangle:E|\psi\rangle)} = \mathrm{Tr}\sum_{i,j}\mathbf{m}(i,j)\nu_{i,j}(|\psi\rangle\langle\psi| \otimes |\psi\rangle\langle\psi|)$, for some matrices $\nu_{i,j}$ which can be integrated over Λ. □

Theorem 36 *Relativized to POVM* E, $\int 2^{\mathbf{I}_{\mathrm{Prob}}(E\sigma:E\sigma)} d\mu(\sigma) = O(1)$.

Proof. By Lemma 9, $2^{\mathcal{I}(\sigma:\sigma)} \overset{*}{>} 2^{\mathbf{I}_{\mathrm{Prob}}(E\sigma:E\sigma)}$. From Lemma 8, $\int 2^{\mathcal{I}(\sigma:\sigma)} d\mu(\sigma) = O(1)$. □

An implication of Theorems 35 and 36 is that for an overwhelming majority of quantum states, the probabilities induced by a measurement will have negligible self information.

7.3.1 Algorithmic Content of Decoherence

Decoherence was explained in Section 6.3.3. In the idealized case, decoherence transforms an arbitrary density matrix σ into a classical probability, with the off-diagonal terms turned to 0. Let p_σ be the classical probability that σ decoheres to, with $p_\sigma(i) = \sigma_{ii}$. The following corollary to Theorem 35, for an overwhelming majority of pure or mixed states σ, p_σ is noise, that is, has negligible self-information. The corollary 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 8 *Let Λ be the uniform distribution on the unit sphere of an n qubit space.*

- $\int 2^{\mathbf{I}_{\mathrm{Prob}}\left(p_{|\psi\rangle}:p_{|\psi\rangle}\right)} d\Lambda = O(1)$,
- $\int 2^{\mathbf{I}_{\mathrm{Prob}}(p_\sigma:p_\sigma)} d\mu(\sigma) = O(1)$.

7.4 PVMs

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 $P|\psi'\rangle(i) = 1$. To look at the effects of a measurement operation on the algorithmic information theoretic properties of a state, take a PVM, $F = \{\Pi_i\}_{i=1}^{2^{n-c}}$, where n is the number of qubits of the Hilbert space. Let Λ_F be the distribution of pure states when F is measured over the uniform distribution Λ over n qubit spaces. Thus Λ_F represents the F-collapsed states from Λ.

Theorem 37 $n - 2c <^+ \log \int 2^{\mathbf{I}_{\mathrm{Prob}}(F:|\psi\rangle:F|\psi\rangle)} d\Lambda_F$.

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

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

Chapter 8

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 infinite system size. This model is an infinite quantum spin chain. 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 a NS random state. 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.

8.1 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 $Z(x^* x) = Z(x x^*)$, 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 $\rho(X) = \mathrm{Tr} S X$, 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 $Z\lceil n$ 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 $\tau\lceil n = \tau_n$. 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.

8.1.1 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 38 ([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 [GÓ1]). Furthermore let $G^e_m = \langle p^e_{m,r}\rangle_{r\in\mathbb{N}}$. For each $k, n \in \mathbb{N}$, let $q^n_k = \bigvee_{e+n+1\leq k} p^e_{e+n+1,k}$. Thus $q^n_k \leq q^n_{k+1}$ and $\tau q^n_k \leq \sum_e \tau(p^e_{e+n+1,k}) \leq 2^{-n}$. The universal NS test is $R^n = \langle q^n_k\rangle_{k\in\mathbb{N}}$. Since $\tau(R^r) \leq 2^{-n}$, $\rangle R^n\rangle$ is a NS test. For a set e,

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

□

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

8.2 Closure Properties

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

Theorem 39 *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.

8.3 Gács Complexity and NS Random Sequences

In this section, we characterize NS random states in terms of Gács complexity, $\mathbf{Hg}$.

Theorem 40 *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 $\mathbf{m}(q|a_n)\sigma \overset{*}{<} \boldsymbol{\mu}_{a_n}$. 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 41 *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 $\mathrm{Rank}(p_n^r) \leq 2^{n-r}$. Thus $\inf_r Z(L^r) = \delta > 0$. For each r, there exists an n such that $\mathrm{Tr}(p_n^r z_n) \geq \delta$, where $z_n = Z \upharpoonright n$. Since $2^{r-n}p_n^r$ is a computable semi-density matrix given n and r, $\mathbf{m}(r|n)2^{r-n}p_n^r \overset{*}{<} \boldsymbol{\mu}$. So $\mathbf{m}(r|n)2^{r-n}\delta \overset{*}{<} \mathrm{Tr}\boldsymbol{\mu}z_n$, 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 Theorem 40, $\sup_{n\in A} n - \mathbf{Hg}(Z \upharpoonright n) = \infty$.

8.4 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 $\mathrm{Tr} pZ{\upharpoonright}n > k$, 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 $[Z]$.

8.5 Quantum Operation Complexity

In a canonical algorithmic information theory example, Alice wants to send a single text message x to Bob. Alice sends a program q to Bob such that $x = U(q)$, where U is a fixed universal Turing machine. The cost of the transmission is the length of q. 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, represented as an density matrix over $\mathbb{C}^{2^n}$, or an element of $S(\mathcal{M}_n)$. 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 classical bits $q \in \{0,1\}^*$ on the classical channel, describing an elementary quantum operation η, where $U(q) = [\eta]$. Bob then applies η to ρ to produce $\sigma' = \eta(\rho)$. Bob is not required to produce σ exactly. Instead the fidelity of the attempt is measured by the trace distance between σ and σ'. The trace distance D between two matrices A and B is $D(A,B) = \frac{1}{2}\|A-B\|_{\mathrm{Tr}}$, with $\|A\|_{\mathrm{Tr}} = \mathrm{Tr}|A|$. We use $\mathcal{O}_{m,n}$ to denote the set of elementary quantum operations that take m qubit quantum states to $n \geq m$ qubit quantum states.

Definition 11 *For n qubit density matrix σ, the quantum operation complexity at accuracy ϵ is* $\mathbf{Hoc}^\epsilon(\sigma) = \min\{\mathbf{K}([\eta]) + m : \eta \in \mathcal{O}_{m,n}, \xi \in S(\mathcal{M}_m), D(\sigma, \eta(\xi)) < \epsilon\}$.

8.6 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. Theorem 42 builds upon the proof of the Theorem 4.4 in [NS19] using quantum operation complexity $\mathbf{Hoc}$.

Theorem 42 *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, $\mathbf{Hoc}^\epsilon(Z{\upharpoonright}_n) > n - r$.*
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 $\mathbf{Hoc}^{\sqrt{\epsilon}}(Z{\upharpoonright}_n) < n - r$.*

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 $\rho \in S(\mathcal{M}_n)$ such that there exists a quantum operation $\eta \in \mathcal{O}_{z,n}$ and pure state $\sigma \in S(\mathcal{M}_z)$ such that $\rho = \eta(\sigma)$ and $\mathbf{K}_t([\eta]) + z \leq n - r$. 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 $\mathcal{M}_t$ and $p(r,t)$ is computable from r and t and $p(r,t) \leq p(r,t+1)$.

Let $b(y,n,z)$ be the number of programs of length y which outputs an encoding of an elementary quantum operation $\eta \in \mathcal{O}_{z,n}$. Let $b(y,n)$ be the number of programs of length y which outputs an encoding of an elementary quantum operation $\eta \in \mathcal{O}_{z,n}$, for any $z \leq n$. So

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

So for NS Σ^0_1 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 $\mathbf{Hoc}^{\epsilon}(Z\lceil n) \leq n-r$. So there is an elementary quantum operation $\eta \in \mathcal{O}_{z,n}$ and input $\rho \in S(\mathcal{M}_z)$ such that $\mathbf{K}([\eta]) + z \leq n-r$ and $D(Z\lceil_n, \eta(\rho)) < \epsilon$. So $\eta(\rho)$ is in the range $p(n,r,t)$ for some t and so $\text{Tr}\eta(\rho)p(n,r,t) = 1$. This implies $1-\epsilon < Z(p(n,r,t)) \leq Z(G^r)$. 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 $\leq n$. Let $\langle L^r\rangle$ be the universal NS test, where each L^r is enumerated by $\{p^r_t\}$, with $p^r_t \in \mathcal{M}_{n(r,t)}$. Assume there is an infinite number of r where $\text{Tr}Z\lceil n(r,\mathbf{bb}(r/2))p^r_{\mathbf{bb}(r/2)} > 1-\epsilon$. 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 $\{u_i\}$ and kernel spanned by $\{v_i\}$. Thus $2^{-r} \geq \tau(p)$. Let $p' \geq p$ with $p' \in \mathcal{M}_n$ such that each u_i is in the range of p' and $\{v_i\}_{i=1}^k$ is in the range of p' such that k is minimized such that $\tau(p') = 2^{-r}$. Thus $\text{Tr}Z\lceil n(p') > 1-\epsilon$. The eigenvectors of p' are $\{w_i\}_{i=1}^{2^{n-r}}$ and its kernel is spanned by the vectors $\{y_i\}_{i=1}^{2^n-2^{n-r}}$. Let $z' = \text{Proj}(Z\lceil n; p')$ be a density matrix with eigenvalues $v_i \in \mathbb{R}$ corresponding to eigenvectors w_i. For $i \in [1,2^n]$, let $B(i) \in \{0,1\}^*$ be an encoding of n bits of the number i, with $B(1) = 0^{(n)}$, $B(2) = 10^{(n-1)}$, and $B(2^n) = 1^{(n)}$. 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 3 ([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 $D(\sigma,\sigma') \leq \sqrt{1-\alpha}$.*

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 $F(\sigma,\sigma') \geq \alpha^{-\frac{1}{2}}\langle\psi_\sigma|p|\psi_\sigma\rangle = \alpha^{\frac{1}{2}}$, where F is fidelity, with $F(\sigma,\sigma') = \text{Tr}\sqrt{\sqrt{\sigma'}\sigma\sqrt{\sigma'}}$. Thus the proposition follows from $D(\sigma,\sigma') \leq \sqrt{1-F(\sigma,\sigma')}$. □

Thus for the diagonal $2^{n-r} \times 2^{n-r}$ matrix σ with entries $\{v_i\}_{i=1}^{2^{n-r}}$, $z' = U(\sigma \otimes |0^r\rangle\langle 0^r|)U^*$. By Proposition 3, since $1 - \epsilon < \mathrm{Tr}(p' Z{\upharpoonright}n)$ and $z' = \mathrm{Proj}(z_n; p')$, it must be that $D(z', Z{\upharpoonright}n) < \sqrt{\epsilon}$. Thus using quantum operation $\eta = (U, |0^r\rangle\langle 0^r|, \emptyset)$ and input σ,

$$
\begin{aligned}
\mathbf{Hoc}^{\epsilon}(Z{\upharpoonright}n) &\leq \mathrm{Dim}(\sigma) + \mathbf{K}([\eta]) \\
&\leq n - r + \mathbf{K}([(U, |0^r\rangle\langle 0^r|, \emptyset)]) \\
&<^{+} n - r + \mathbf{K}(n, r) \\
&<^{+} n - r + \mathbf{K}(\mathbf{bb}(r/2), r) \\
&<^{+} n - r + r/2 + \mathbf{K}(r) \\
&<^{+} n - r/3.
\end{aligned}
$$

Thus for every r there exists an n where $\mathbf{Hoc}^{\epsilon}(Z{\upharpoonright}n) < n - r$. 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 $q < \mathbf{bb}(r/2)$, $\mathrm{Tr} Z_{n(r,q)} p^r_{n(r,q)} \leq 1 - \epsilon$. 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 $\mathrm{Tr} Z_{n(r,s)} p^r_{n(r,s)} > 1 - \epsilon$. This is because the set of rational numbers Q such that $q > 1 - \epsilon$ for all $q \in Q$ can be enumerated and the set $V = \{v : \mathrm{Tr} Z_{n(r,v)} p^r_{n(r,v)} > q, q \in Q\}$ can be enumerated using the infinite encoding $[Z] \in \{0,1\}^{\infty}$. The returned s is the first enumerated element of V. 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 9 *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, $\mathbf{Hoc}^{\epsilon}(Z \upharpoonright n) > n - r$.*

Proof. Assume Z is NS random. Then for all $0 < \epsilon < 1$, Z passes each NS test at order $1 - \epsilon$. Then by Theorem 42 (1), for all $0 < \epsilon < 1$ there is an r where for all n, $\mathbf{Hoc}^{\epsilon}(Z{\upharpoonright}n) > n - r$. 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 42 (2), for $\epsilon = \sqrt{\delta}$, for all r, there is an n where $\mathbf{Hoc}^{\epsilon}(Z{\upharpoonright}n) < n - r$.

8.7 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 BVL complexity (using the fidelity measure). The results provide two bounds with respect to two variants of **Hbvl**: approximate-scheme complexity and finite accuracy complexity.

In this subsection we provide a quantum variant of Brudno's theorem with respect to quantum communication complexity $\mathbb{R}^{\epsilon}$. Differently from the **Hbvl** 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 qbits, 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 $\Psi\lceil n$ to denote Ψ restricted to the finite subalgebra $\mathcal{A}_{\{1,\dots,n\}}$ of $\mathcal{A}_\infty$. The right shift T is a $*$-automorphism on $\mathcal{A}_\infty$ uniquely defined by its actions on local observables $T : a \in \mathcal{A}_{\{m,\dots,n\}} \mapsto \mathcal{A}_{\{m+1,\dots,n+1\}}$. 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.

Lemma 10 *Let R_j be the smallest subspace spanned by pure states produced by elementary quantum operations $\eta \in \mathcal{O}_{z,n}$with $\mathbf{K}(\eta) + z < j$. Then $Dim(R_j) < 2^j$.*

Proof. Let $b(y, z)$ be the number of programs of length y that outputs an elementary quantum operation $\eta \in \mathcal{O}_{x,z}$ over the Hilbert space $\mathcal{H}_{2^n}$. Let $b(y)$ be the number of programs of length y that outputs an elementary quantum operation $\mathcal{O}_{z,n}$ over the Hilbert space $\mathcal{H}_{2^n}$.

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

Theorem 43 *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. *$\Psi\lceil n(P_n) > 1 - \delta$.*
2. *For all one dimensional projectors $p \le P_n$, $\mathbf{Hoc}^\epsilon(p)/n \in (h - \delta, h + \delta)$.*

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, $\Psi\lceil n(P'_n) > 1 - \delta'$, for all one dimensional projectors $p' \le P'_n$, $2^{-n(h+\delta')} < \Psi\lceil n(p') < 2^{-n(h-\delta')}$, 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 an elementary quantum operation $\eta \in \mathcal{O}_{g,n}$, where $\mathbf{K}(\eta) + g < n(h - \delta'')$. Let Q_n be the projector onto R_n. By Lemma 10, $\mathrm{Dim}(R_n) < 2^{n(h-\delta'')}$. 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, $\Psi\lceil n(P_n) \ge \Psi\lceil n(P'_n) - \mathrm{Dim}(R_n)2^{-n(h-\delta')} > 1 - \delta' - 2^{n(h-\delta'')}2^{-n(h-\delta')} = 1 - \delta' - 2^{n(\delta'-\delta'')} > 1 - \delta$, for large enough n.

By definition, since P_n is orthogonal to R_n, for all ϵ, for all one dimensional projectors $p \le P_n$, $\mathbf{Hoc}^\epsilon(p) \ge n(h - \delta'') > n(h - \delta)$. Furthermore, all such p can be produced from an elementary quantum operation η that maps $\lceil n(h + \delta')\rceil$ length pure states into S_n. Therefore for large enough n, $\mathbf{Hoc}^\epsilon(p) \le \mathbf{K}(\eta) + \lceil n(h + \delta')\rceil <^+ \mathbf{K}(n, h) + \lceil n(h + \delta')\rceil < n(h + \delta)$.

8.8 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 the Carathéodory's Extension Theorem, each such pre-measure can be uniquely extended to a measure Γ over $\{0,1\}^\infty$. in Chapter 7, measurements of finite collections of qubits are studied. This section deals with measurement measurement systems, which can be applied to infinite quantum states.

Definition 12 (Meaurement System ([Bho21])) *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 defiinition 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 $Z \upharpoonright n$. For the first bit, we use the standard definition of measurement, where

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

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

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

So

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

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

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

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

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

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

Theorem 44 ([Bho21]) *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 $[\![A_i^m]\!] \subseteq [\![A_{i+1}^m]\!]$, $\mathrm{Range}(p_i^m) \subseteq \mathrm{Range}(p_{i+1}^m)$. 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, $\{0,1\}^\infty \setminus \mathrm{MLR} \subseteq S^m$. Since by assumption $\mu_Z^B(\{0,1\}^\infty \setminus \mathrm{MLR}) > \delta$, 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 p be the pre-measure associated witth μ_Z^B. So we have

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

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

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

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

Theorem 45 ([Bho21]) *There are states that are Bhojraj random and not NS Random.*

8.9 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 [Bho21], it was proven that NS randomness is equivalent to NS Solovay randomness.

Proposition 4 *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 46 *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^0_1 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 47 *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 $G^r = \langle p^r_n \rangle_{n \in \mathbb{N}}$, 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 $\{D^r_n\}_{n \in \mathbb{N}}$, $D^r_n \subset \mathbb{N}$, with infinite repetition.

$$H^r = \{q^r_n\}, q^r_n = \bigvee_{m<n} q^r_m \bigvee \left(\bigwedge_{t \in D^r_n} p^t_n \right).$$

Each H^r can be seen to be an NS Σ^0_1 set. Furthermore $\tau(H^r) \leq \sum_{t>r} Z(G^t) \leq \sum_{t>r} 2^{-t} = 2^{-r}$. 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 $Z(p^r_{n_r}) \geq \delta''$, for some $\delta' < \delta'' < \delta$. We reorder the NS Solovay test $\langle G^r \rangle$ such that r ranges over solely R. Let $z_r = p^r_{n,r}$. Let $D_{n,k}$ be the set of all unordered subsets of $\{1, \ldots, n\}$ of size k. For $k > n$ let $F_{n,k} = \emptyset$. Let

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

So for all $n \in \mathbb{N}$, using Proposition 4,

$$\begin{aligned} & n(\delta'' - \delta') \\ \leq & \sum_{r=1}^{n} (Z(z^r) - Z(H^k)) \\ \leq & \sum_{r=1}^{n} Z(z^r \setminus H^k) \\ \leq & \sum_{r=1}^{n} Z\left(\bigvee_{s=1}^{k} F_{n,s} \wedge z^r \right) \end{aligned} \tag{8.1}$$

Equation 8.1 is due to the fact that for $s > k$ there is a t where we have $\mathrm{Range}(F_{n,s}) \leq \mathrm{Range}(q^k_s)$. Let $F_{n,s,r} = F_{n,s} \wedge z^r$, with for a fixed $s \leq k$, $\sum_{i=1}^{n} Z(F_{n,s,i}) \leq s$.

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

This is a contradiction for large enough n. □

Corollary 10 *A quantum state is NS random if and only if it is NS Solovay random.*

Chapter 9

The Many Worlds Theory

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**) is a finitary Church-Turing thesis, postulating that certain infinite and *finite* sequences cannot be found in nature, a.k.a. have high "addresses". 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. This blog entry covers the main arguments of this conflict.

9.1 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.

9.1.1 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

$$\begin{aligned}
|\phi_\uparrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{ready}}\rangle &\xrightarrow{\text{unitary}} |\phi_\uparrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin}\uparrow}\rangle \\
|\phi_\downarrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{ready}}\rangle &\xrightarrow{\text{unitary}} |\phi_\downarrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin}\downarrow}\rangle\,.
\end{aligned}$$

Furthermore, one can model the entire quantum state of an observer $\mathcal{O}$ of the apparatus, with

$$\begin{aligned}
&|\phi_\uparrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{ready}}\rangle \otimes |\xi^{\mathcal{O}}_{\text{ready}}\rangle \\
\xrightarrow{\text{unitary}}\ &|\phi_\uparrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin}\uparrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{ready}}\rangle \\
\xrightarrow{\text{unitary}}\ &|\phi_\uparrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin}\uparrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{reads spin}\uparrow}\rangle \\
\\
&|\phi_\downarrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{ready}}\rangle \otimes |\xi^{\mathcal{O}}_{\text{ready}}\rangle \\
\xrightarrow{\text{unitary}}\ &|\phi_\downarrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin}\downarrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{ready}}\rangle \\
\xrightarrow{\text{unitary}}\ &|\phi_\downarrow\rangle \otimes |\psi^{\mathcal{A}}_{\text{reads spin}\downarrow}\rangle \otimes |\xi^{\mathcal{O}}_{\text{reads spin}\downarrow}\rangle\,.
\end{aligned}$$

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.

9.1.2 Deriving the Born Rule

In my 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.

9.2 Violating 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 $\mathbf{I}(\alpha:\beta) < k+n+c_{\mathrm{IP}}$, for some small absolute constant c_{IP}.*

The $\mathbf{I}$ term is an information measure in Algorithmic Information Theory. For this blog, the information term we use is $\mathbf{I}(x:y) = \mathbf{K}(x)+\mathbf{K}(y)-\mathbf{K}(x,y)$, where $\mathbf{K}$ is the prefix-free Kolmogorov complexity. We can use this definition of $\mathbf{I}$ because we 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). We have $m <^+ \mathbf{K}(\Omega_m)$. 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 addresses of size at least $m - O(\log m)$.

IP can be violated in the following idealized experiment measuring the spin $|\phi_\uparrow\rangle$ and $|\phi_\downarrow\rangle$ of N 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 is $< O(\log N)$. The measuring apparatus will measure the spin of N 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 N spins of N electrons, it is in the state $|\psi^{\mathcal{A}}[x]\rangle$, where $x \in \{0,1\}^N$, whose ith bit is 1 iff the ith measurement returns $|\phi_1\rangle$.

The experiment evolves according to the following unitary transformation:

$$\bigotimes_{i=1}^{N} |\phi\rangle \otimes |\psi^{\mathcal{A}}\rangle \overset{\text{unitary}}{\longrightarrow} \sum_{a_1,\ldots,a_N \in \{0,1\}^N} 2^{-N/2} \bigotimes_{i=1}^{N} |\phi_{a_i}\rangle \otimes |\psi^{\mathcal{A}}[a_1 a_2 \ldots a_n]\rangle\,.$$

If the bits returned are Ω_N then a memory leak of size $N - O(\log N)$ has occurred, because Ω_N has been located by the address of the experiment, which is $O(\log N)$. Thus

$$\text{Born-Probability(a memory leak of size } N - O(\log N) \text{ occurred)} \geq 2^{-N}.$$

9.3 Conclusion

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 gravity, 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^{-N} over all worlds, there is a memory leak of size $N - O(\log N)$ and **IP** statement by the mathematician is in error.

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

$$\Pr_{\alpha\sim P}[\mathbf{I}(\alpha:\beta) > \mathbf{I}(\langle P\rangle:\beta)+m] \overset{*}{<} 2^{-m}.$$

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 occuring at an event with a small address. Since there are a small number of events that have a small address, the probability of a signifcant 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.

There is no way to postulate 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 N particles, Alice now knows 2^{-N} of the resultant worlds will contain N bits of Chaitin's Omega, violating **IP**.

Chapter 10

Conclusion

10.1 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, we have

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

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

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

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 9. Information non growth laws can be extended to signals [Eps23a] which can be modeled as probabilities over $\mathbb{N}$ or Euclidean space[1]. The "signal strength" of a probability p over $\mathbb{N}$ is measured by its self information.

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

A signal, when undergoing randomized processing f (see Section 3.1.1), will lose its cohesion. Thus any signal going through a classical channel will become less coherent.

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

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

[1] In [Eps23a] probabilities over $\{0,1\}^\infty$ and T_0 second countable topologies were also studied.

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 a classical channel $X \to Y$. So using the above inequality, cohesion will deteriorate regardless of $X's$ probability, with

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

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 $E\sigma(k) = \mathrm{Tr}\sigma E_k$. However from [Eps23a], for elementary (even enumerable) probabilities $E\sigma$,

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

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 35 and 36 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 6.2.2),

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

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 8 states that for an overwhelming majority of pure or mixed states σ, p_σ is noise, that is, has negligible self-information.

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

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, we've 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, 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 37 states

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

10.2 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

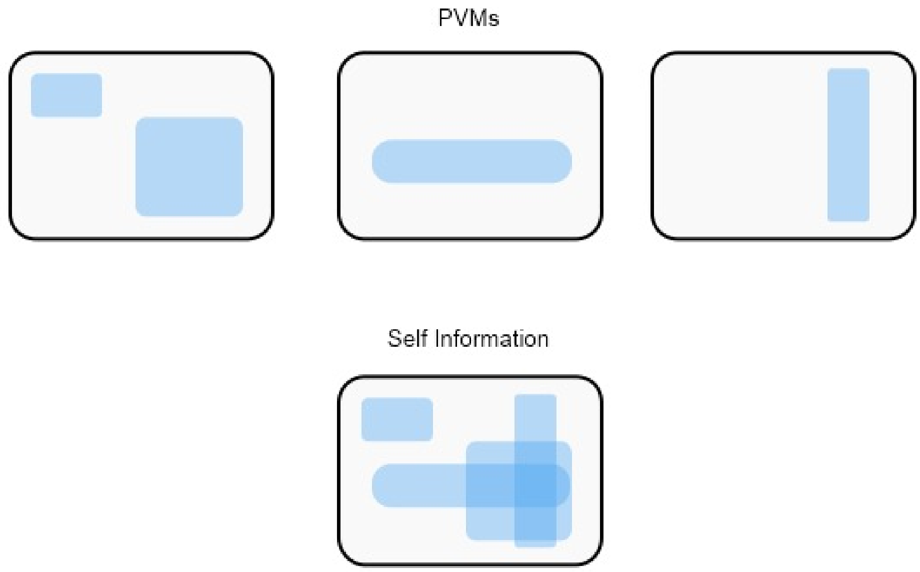

Figure 10.1: Each box on the top row represents an n qubit Hilbert space, with the shaded rectangles being the subspaces of the PVM projectors. Thus there are three PVMs. The self-information majorizes these subspaces, inversely weighted by the PVM's complexity.

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. 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 34, 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 8. 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 Corollary 7, 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 10.1). So this fact, and Theorem 29, 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 mea-

sured 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 31, 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.

10.3 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,

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

This represents the shared signal strength between $|1\rangle$ and $|2\rangle$ when the states[2] were created from a unitary transform U of $|\psi\rangle$ tensored with an ancillia state $|0\rangle$. Note that by Theorems 32 and 34, cloneable information is less than self information, with

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

The question is, given an initial distribution over density operators with low expected $\mathbf{I}_{\mathrm{Clone}}$, 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 29 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 29,

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

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.

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

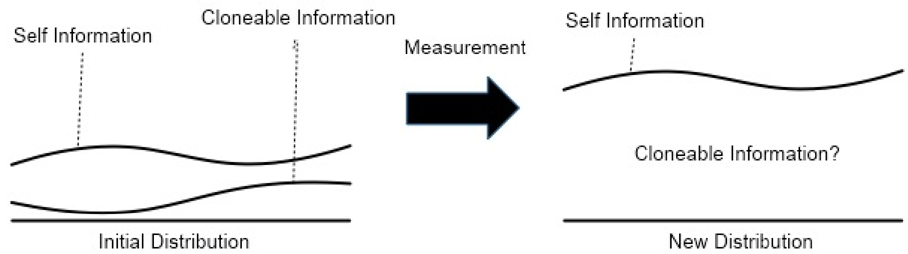

Figure 10.2: The intial distribution has low self information and cloneable information. A measurement increases the self information and potentially increases the cloneable information.

How do you create a distribution with high expected $\mathbf{I}_{\mathrm{Clone}}$, where most states can have cloneable information? Any transform that increases cloneable information must increase self-information. However Theorem 31, 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 34 and 37 show, often times cause an uptake in self-information (see Figure 10.2). This is also discussed in the quotes of Section 10.2. 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 29 and 32, we have that

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

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 $\mathbf{I}_{\mathrm{Clone}}(|i\rangle) = \mathbf{I}_{\mathrm{Prob}}(E|i\rangle : F|i\rangle) = \mathbf{K}(i)$. Since there are $2^{n-O(1)}$ basis states $|i\rangle$ where $n <^+ \mathbf{K}(i)$, we have the following uptake in cloneable information.

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

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

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 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 $A \subseteq \mathcal{X}$.

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 paper, we introduce a modified version of coarse grained entropy. 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

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

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

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

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 the fluctuation, the lesser its measure. The thermodynamics section of this manuscript is arranged as follows.

- **Chapter 11:** 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 12:** 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 [Ǵ21] that algorithmic fine grain entropy is conformant to the addition inequality analagous to that of string algorithmic information theory.
- **Chapter 13:** 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 15:** 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 14:** 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 11

Computable Measure Spaces

The results in the thermodynamics uses computable metric spaces and computable measure sapaces. We use the constructs from[HR09], generalizing from probability measures to arbitrary nonnegative measures and from binary representations to dual binary representations.

Definition 14 *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 5 *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.*

11.1 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 6 *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 $x = \sup\{p_i : i \in E\}$. $\square$

Definition 15 *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 7 *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 $f(y)$. $\square$

Definition 16 *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*

$$f_{(i,j)}(x) = \begin{cases} p_j & \text{if } x \in B_i \\ \perp & \text{otherwise.} \end{cases}$$

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 5 *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 17 *A subset A of X is semi-decidable if it is an r.e. open set.*

Proposition 8 *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.* $\square$

11.2 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 18 (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 18). 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. From this definition, the results of Theorem **??** apply directly to arbitrary measures $\mu \in \mathfrak{M}(X)$.

Proposition 9 *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 $(s_{n_1}, q_{m_1}), \dots, (s_{n_k}, q_{m_k})$ 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 $\mu_i(B_{i_1} \cup \dots B_{i_k})$ is lower-computable and $\mu_i(\overline{\overline{B}}_{i_1} \cup \dots \overline{\overline{B}}_{i_k})$ is upper semi-computable, uniformly in i and $(i_1, \dots, i_k)$.

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 $\pi(\mu_i,\mu_j) = \sup\{\epsilon \in \mathbb{Q} : \exists A \subset S_i, \mu_i(A) > \mu_j(A^{\overline{\overline{\epsilon}}}))\}$, with $A^{\overline{\overline{\epsilon}}} = \{x : d(x,A) \leq \epsilon\}$, and using the upper semi-computability of $\mu_j(A^{\overline{\overline{\epsilon}}})$.

In addition, it easy to see that $d_\mathbb{R}(v_i, v_j)$ is computable. Thus the following term is computable.

$$d_\mathfrak{M}((\mu_i,v_i),(\mu_j,v_j)) = \max\{\pi(\mu_i,\mu_j), d_R(v_i,v_j))\}.$$

□

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 10 *The valuation operator v is bi-lower computable, in the second argument.*

Proof. Since $v((\mu, m), \cdot)$ is Scott-continuous, due to Proposition 7, 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 9, $\mu_i(B_{i_1} \cup \cdots \cup B_{i_k})$ 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 $(i, i_1, \ldots, i_k)$. $\square$

Proposition 11 *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 5. 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 5 can be applied. Thus one can construct a fast sequence of ideal measures converging to (μ, m). $\square$

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 11 *If a measure $(\mu, m) \in \mathfrak{M}(\{0, 1\}^\infty)$ is computable then the cylinders are uniformly computable.*

Proposition 12 *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 $m \int f_{(i,j)} d\mu = mq_j \mu(B_i)$ and

$$m \int \sup \left\{f_{(i_1,j_1)}, \ldots, f_{(i_k,j_k)}\right\} d\mu = mq_{j_z}\mu(B_{i_1} \cup \cdots \cup B_{i_k}) + m \int \sup \left\{f_{(i_1,j_1')}, \ldots, f_{(i_k,j_k')}\right\} d\mu$$

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 10, 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 7, the opertor is bi-lower computable. $\square$

Corollary 12 *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 $m\int f_i d\mu = m\int(f+M_i)d\mu - mM_i = mM_i - m\int(M_i - f_i)d\mu$ is lower and upper computable by Proposition 12.

11.3 Computable Measure Space

Definition 19

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 $f :\subset (\mathcal{X},(\mu,mxf)) \to \mathcal{Y}$ 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.*

The follow proof of existence of an almost decidable set is from [GHR11].

Lemma 11 *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 13 *Let $(\mathcal{X}, (\mu, m), (\nu, m))$ be a dual measure space nd $(f_i)_i$ 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 11, (λ, m) and (γ, m) are computable measures so by Lemma 11 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 14 *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 13 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.

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 14. 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 48 *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 $x \in \overline{\Gamma(\omega[0\dots i])}$, we have $\Gamma(\omega[0\dots i]) \neq \emptyset$ 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 $\overline{\Gamma[0\dots i_n]} \subseteq \overline{\overline{B}}_{i_n}$ with radius going to zero, the sequence $\overline{\Gamma(\omega[0\dots n])}$ 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 $\omega = \omega[0]\omega[1]\dots$ such that for all i, $x \in \overline{\Gamma(\omega[0\dots i])}$. Let $i \geq 0$ andd suppose that $\omega[0\dots i-1]$ has been constructed. Since $B_i \cup C_i$ is open and dense and $\Gamma(\omega 0\dots i-1)$ is open, $\overline{\Gamma(\omega_{0\dots i-1})} = \overline{\Gamma(\omega_{0\dots i-1}) \cap (B_i \cup C_i)} = \overline{\Gamma\omega_{0\dots i-1}0} \cup \overline{\Gamma\omega_{0\dots i-1}1}$, so for some $\omega[i] \in \{0,1\}$, has $x \in \overline{\Gamma(\omega_{0\dots i})}$. So $x \in \cap_i \overline{\Gamma(\omega_{0\dots i-1})}$. Since $(B_i)_i$ is a basis and $\omega_i = 1$ whenever $x \in B_i$, ω is an expansion of x.

11.4 Randomnesss

Definition 20 *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 [Gá21]), 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 6, 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 12 *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 $\int mH_i((\mu, m))d\mu < 1$ 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 $Y = \mathfrak{M}(X), \mathscr{C}(X, \overline{R}^+)$ endowed with the canonical computable metric structure. By "curryfication" it provides functions $h_n \in \mathfrak{M}(X), \mathscr{C}(X, \overline{R}^+))$ 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 12, the integration operator $\int : \mathfrak{M}(X) \times \mathbb{N} \to [0, 1]$ which maps $((\mu, m)\langle i_1, \dots, i_k\rangle)$

to $m\int \sup\{h_{i_1}((\mu,m)),\dots,h_{i_k}((\mu,m))\}d\mu$ 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 49 *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 12, 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 [G2́1] with the proofs adapted to uniform tests. Let $F : \mathcal{Y}\times\mathfrak{M}(X) \to \mathbb{Z}\cup\{-\infty,\infty\}$ be lower computable, where $\mathcal{Y}$ is a computable metric space. An F uniform randomness test R is a computable function from $\mathcal{Y}\times\mathfrak{M}(X)$ to $\mathscr{C}(X,\overline{R}^+)$ such that $m\int R^{(y,(\mu,m))}d\mu \leq 2^{-F(y,(\mu,m))}$.

Corollary 15 *There exists a universal* F *uniform test* $\mathbf{r}$ *such that* $m\int \mathbf{r}^{(y,(\mu,m))}d\mu \leq 2^{-F(y,(\mu,m))}$ *and for every* F *uniform test* R*,* $\mathbf{m}(R|\langle\overrightarrow{y}\rangle)R_y \overset{*}{<} \boldsymbol{r}_y$.

Proof. The proof follows analogously to that of Lemma 12, except $T(i,y,(\mu,m)) = \sup\{H_i^k(y,(\mu,m)) : m\int H_i^k(y,(\mu,m)) \leq 2^{-F(y,(\mu,m))}\}$ where $H_i^k = \sup_{n<k} h_{\phi(i,n,\langle\overrightarrow{y}\rangle)}$. The term $\phi(i,n,\langle\overrightarrow{y}\rangle)$ is the partial recursive function being given the numbers i and n, and an encoding of a fast Cauchy sequence for $y\in\mathcal{Y}$.

Corollary 16 *Let* $\overline{Z} = \mathbb{Z}\cup\{-\infty,\infty\}$*. For computable metric spaces* $\mathcal{X}$ *and* $\mathcal{Y}$*, consider computable functions* $\mathcal{Y}\to\mathscr{C}(X,\overline{\mathbb{R}}^+)$*. There is a modification to a computable* $w'\mathfrak{M}(X)\times\mathcal{Y}\times\overline{Z} \to\to \mathscr{C}(X,\overline{\mathbb{R}}^+)$ *such that for all* $y\in\mathcal{Y}$*,* $(\nu,m)\in\mathfrak{M}(X)$*,* $m\in\overline{Z}$*,* $n\int w'((n,\nu),y,m,x)d\nu \leq 2^{-m+1}$ *and if* $n\int w((n,\nu),y,x)d\nu \leq 2^{-m+1}$.

Proof. For every w, there is an i such that $w((\nu,n),x,y) = \sup_k H_k((\nu,n),x,y)$, where $H_k = \sup_{j<k} h_{\phi(i,j,\langle\overline{y}\rangle)}$, where h is defined in the proof of Lemma 12. The transformed function is

$$w'((\nu,n),y,m,x) = \sup_k\{H_k((\nu,n),x,y) : n\int H_k((\nu,n),x,y)d\nu \leq 2^{-m}\}.$$

□

Claim 1 *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 [G2́1],* $2^{\mathbf{D}}$ *is a lower-computable* (P,p)*-test, in that*

$$p\int_{\{0,1\}^\infty} 2^{\mathbf{D}(\alpha|(P,p))}dP(\alpha) = O(1).$$

Thus since $\mathbf{t}_{(P,p)}$ *is a universal uniform test that takes the computable point* P *as a parameter, by Corollary 11,* $\mathbf{t}_{(P,p)}$ *can enumerate lower* (P,p)*-tests and give them an oracle can compute the measure cylinders* $p{\cdot}P(x)$ *for* $x \in \{0,1\}^*$*. Thus* $\mathbf{t}_{(P,p)}(\alpha) \overset{*}{>} 2^{\mathbf{D}(\alpha|(P,p))}$*.*

Furthermore since the test $\mathbf{D}$ *is given as oracle a program that compute the measure cylinders,* $p = P(\emptyset)$ *can be computed so by Proposition 11, the test* $\mathbf{D}$ *has access to a fast Cauchy sequence converging to the measure* (P,p)*. Thus there exists a test that can compute* $\mathbf{t}_{(P,p)}$ *given access to this converging sequence. Thus* $\mathbf{t}_{(P,p)}(\alpha) \overset{*}{<}^{\mathbf{D}(\alpha|(P,p))}$*.*

Proposition 13 *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 $\{(i,k) : \exists j, (i,j) \in E q_k < q_j\}$. 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 13 *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* $\mathbf{t}_{(\nu,n)}$*. 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* $\mathbf{t}_{(\nu,m)}(Q(x)) \overset{*}{<} c\mathbf{t}_{(\mu,m)}(x)$*.*

Proof. Assuming $\mathbf{t}_{(\mu,m)}(x) < \infty$, then x is a random point then $x \in D$, because due to Propostion 13, 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 $\mathbf{t}_{(\nu,m)} \circ Q \overset{*}{<} c\mathbf{t}_{(\mu,m)}$. □

Chapter 12

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 canonica properties of $\mathbf{H}_\mu$, originating from [G21], with modifications to the proofs as needed to be compatible with Chapter 11. In this chapter and in subsequent ones, we represent (not necessarily probabilistic) measures as μ, dropping the (μ, m) notation.

Definition 21 *Given a measure space $(\mathcal{X}, \mu)$, its corresponding algorithmic fine grained entropy is $\mathbf{H}_\mu(x) = -\log \mathbf{t}_\mu(x)$.*

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-nullset then $\mathbf{H}_\mu(x) = -\infty$.

Proposition 14 *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 $\mu^x \nu^y 2^{-\mathbf{H}_\mu(y)} \leq \mu^x 1 = 1$. □

Definition 22 (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 $G^t x$.*

12.1 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 23 (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 15 is due to [Gac94], with usage of algorithmic fine grained entropy. Put another way,

> *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 15 $-\mathbf{I}(\alpha;t) <^+ \mathbf{H}_\mu(G^t\alpha) - \mathbf{H}_\mu(\alpha) <^+ \mathbf{I}(G^t\alpha;t)$.

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 16 (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 $2^{-\mathbf{H}_\mu((G^t\times\mathbf{Id})(\alpha,\beta))}$ is a $\mu\times\nu$ test. So $\mathbf{H}_\mu(\alpha) >^+ \mathbf{H}(G^t\alpha)$ and $\mathbf{H}_{\mu,\nu}(\alpha,\beta) <^+ \mathbf{H}(\mu,\nu)(G^t\alpha,\beta)$. So

$$\begin{aligned}\mathbf{I}(G^t\alpha:\beta) &= \mathbf{H}_\mu(G^t\alpha:\beta) + \mathbf{H}_n(\beta) - \mathbf{H}_{\mu,\nu}(G^t\alpha,\beta)\\ &<^+ \mathbf{H}_\mu(\alpha:\beta) + \mathbf{H}_n(\beta) - \mathbf{H}_{\mu,\nu}(\alpha,\beta).\end{aligned}$$

□

12.2 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 14 $\Delta\mathbf{H}_\mu(\alpha) + \Delta\mathbf{H}_\nu(\beta) >^+ \mathbf{I}(\alpha_t:\beta_t) - \mathbf{I}(\alpha:\beta) - \mathbf{I}((\alpha,\beta);t)$.

Proof. Due to Proposition 15 applied to (α,β), $\Delta\mathbf{H}_{\mu\times\nu}(\alpha,\beta) >^+ -\mathbf{I}((\alpha,\beta);t)$. So

$$\begin{aligned}\mathbf{H}_\mu(\alpha_t)+\mathbf{H}_\nu(\beta_t) =&\mathbf{H}_{\mu\times\nu}(\alpha_t,\beta_t)+\mathbf{I}((\alpha_t,\beta_t);t)\\ >^+&\mathbf{H}_{\mu\times\nu}(\alpha,\beta)-\mathbf{I}((\alpha,\beta);t)+\mathbf{I}(\alpha_t:\beta_t)\\ =^+&\mathbf{H}_\mu(\alpha)+\mathbf{H}_\nu(\beta)+\mathbf{I}(\alpha_t:\beta_t)-\mathbf{I}(\alpha:\beta)-\mathbf{I}((\alpha:\beta);t).\end{aligned}$$

□

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.

12.3 Maxwell's Demon

We revisit Maxwell's demon, providing yet another interpretation. This is done by reworking Lemma 14 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 Claim 1, $\mathbf{H}_\lambda(\alpha) =^+ -\mathbf{D}(\alpha|\lambda)$, 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 $\alpha\in\{0,1\}^\infty$, 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 $\mathbf{I}(\alpha:\beta)\approx 0$. The marginal states are assumed independent, with $\mathbf{I}(\alpha:\beta)\approx 0$. Joint dynamics are applied to get $(\alpha^t,\beta^t)=G(\alpha,\beta)$. By Lemma 14,

$$\mathbf{H}_\mu(\alpha^t)-\mathbf{H}_\mu(\alpha) >^+ \mathbf{D}(\beta|\lambda)-\mathbf{D}(\beta^t|\lambda)-\mathbf{I}((\alpha,\beta);t).$$

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.

12.4 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.

Proposition 17 *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 50 *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 17,

$$\begin{aligned}
\nu(\mu) &= -\mu^x \log\frac{\mu(dx)}{\nu(dx)} \\
-\mu^x\mathbf{H}_\nu(x) &= \mu^x\log\frac{\delta(dx)}{\nu * dx)} = -\mu^x\log\frac{\nu(dx)}{\delta(dx)} \\
\mathcal{H}_\nu(\mu) - \mu^x\mathbf{H}_\nu(x) &= -\mu^x\log\frac{\mu(dx)}{\delta(dx)} \leq -\mu(X)\log\frac{\mu(X)}{\delta(X)} \leq 0.
\end{aligned}$$

□

12.5 Addition Equality

For a computable measures μ and ν over a computable metrix space $\mathcal{X}$, the term $\mathbf{H}_\mu(x|\nu)$ is equal to $\mathbf{H}_\mu(x)$ when the universal Turing machine is given access to a fast Cauchy sequence to ν in the measure space $\mathfrak{M}(X)$. By Proposition 10, this means algorithms can lower compute the ν measure of effectively open sets.

Proposition 18 $\mathbf{H}_\mu(x|\nu) <^+ -\log\nu^y 2^{-\mathbf{H}_{\mu,\nu}(x,y)}$.

Proof. Let $f(x,\mu,\nu) = -\log \nu^y 2^{-\mathbf{H}_{\mu,\nu}(x,y)}$. The function f is upper computable and has $\mu^x 2^{-f(x,\mu,\nu)} \leq 1$. Due to the universal properties of $\mathbf{t}_\mu$ and thus minimum property of $\mathbf{H}_\mu$, the inequality is proven. $\square$

Proposition 19 *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

$$g_\mu(x,y) = \sum_z 2^{-\mathbf{H}_\mu(x|f(y,z)) - \mathbf{K}(z)},$$

is lower computable and $\mu^x g_\mu(x,y) \leq \sum_z 2^{-\mathbf{K}(z)} \leq 1$. So $g_\mu(x,y) \overset{*}{<} 2^{-\mathbf{H}_\mu(x|y)}$. The left hand side is a summation, so the inequality holds for each element of the sum, proving the proposition. $\square$

Algorithmic fine grained entropy is bounded from above. It can take arbitrarily low negative values, and even $-\infty$.

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. For (1), We use the μ-test $t_\mu(x) = 1/\lceil \mu(X) \rceil$, where $\int t_\mu d\mu \leq 1$. Thus $-\log \mu(X) <^+ \log t_\mu(x) <^+ \log \mathbf{t}_\mu(x|\lceil \mu(X) \rceil)$. For (2), we use Proposition 19. $\square$

Proposition 21 *If $i < j$, then*

$$i + \mathbf{H}_\mu(x|i) <^+ j + \mathbf{H}_\mu(x|j).$$

Proof. Using Proposition 19, 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 [G2́1], where Corollary 16 has been introduced, leveraging the results in Chapter 11

Proposition 22 *Let $F : \mathcal{Y} \times \mathfrak{M} \to \mathbb{Z} \cup \{-\infty, \infty\}$. By Corollary 15, among F uniform tests $g_\nu(x,y)$ with $\nu^x g_\nu(x,y) \leq 2^{-F_\nu(y)}$ there is a maximal F uniform test f within a multiplicative constant. For all x with $F_\nu(y) > -\infty$,*

$$f_\nu(x,y) \overset{*}{=} 2^{-F_\nu(y)} \mathbf{t}_\nu(x|y, F_\nu(y)).$$

Proof. To prove the inequality $\overset{*}{>}$, let $g_\nu(x,y,m) = \max_{i \geq m} 2^{-i} \mathbf{t}_\nu(x|y,i)$. This function is lower computable, and decreasing in m. Let $g_\nu(x,y) = g_\nu(x,y,F_\nu(y))$ is lower semicomputable since F is upper semi-computable. The multiplicative form of Propisition 21 implies

$$g_\nu(x,y,m) \overset{*}{=} 2^{-m} \mathbf{t}_\nu(x|y,m)$$
$$g_\nu(x,y) \overset{*}{=} 2^{-F_\nu(y)} \mathbf{t}_\nu(x|y, F_\nu(y)).$$

Since $\mathbf{t}_\nu$ is a test:

$$\nu^x 2^{-m}\mathbf{t}_\nu(x|y,m) \le 2^{-m}$$
$$\nu^x g_\nu(x,y) \overset{*}{<} 2^{-F_\nu(y)},$$

which impliies $g_\nu(x,y) \overset{*}{<} f_\nu(x,y)$ by the optimality of $f_\nu(x,y)$. For the upper bound, consider all lower semicomputable functions $\phi(x,y,m,v)$. By Corollary 16, $f'_\nu(x,y,m)$ be the modification of f, which is a lower computable function such that $\nu^x f'_\nu(x,y,m) \le 2^{-m+1}$ and if $\nu^x f_\nu(x,y) \le 2^{-m}$ then $f'_\nu(x,y,m) = f_\nu(x,y)$. The function $2^{m-1}f'_\nu(x,y,m)$ is a uniform test conditioned on y,m so it has $\overset{*}{<} \mathbf{t}_\nu(x|y,m)$. Substituting $F_\nu(y)$ for m, we have that $\nu^x f_\nu(x,y) \le 2^{-m}$ and so

$$f_\nu(x,y) = f'_\nu(x,y,F_\nu(y)) \overset{*}{<} 2^{-F_\nu(y)+1}\mathbf{t}_\nu(x|y,F_\nu(y)).$$

□

Theorem 51

$$\mathbf{H}_{\mu\times\nu} =^+ \mathbf{H}_\mu(x|\nu) + \mathbf{H}_\nu(y|x,\lceil \mathbf{H}_\mu(x|\nu)\rceil,\mu).$$

Proof. We first prove the $<^+$ inequality. Let $G_{\mu,\nu}(x,y,m) = \min_{i\ge m} i + \mathbf{H}_\nu(y|x,i,\mu)$, which is upper computable and increasing in m. So the function

$$G_{\mu,\nu}(x,y) = G_{\mu,\nu}(x,y,\lceil \mathbf{H}_\mu(x|\nu)\rceil).$$

which is also upper computable because m is replaced with an upper computable function $\lceil \mathbf{H}_\mu(x|\nu)\rceil$. Proposition 19 implies

$$G_{\mu,\nu}(x,y,m) =^+ m + \mathbf{H}_\nu(y|x,m,\mu),$$
$$G_{\mu,\nu}(x,y,m) =^+ \mathbf{H}_\mu(x|\nu) + \mathbf{H}_\nu(y|x,\mathbf{H}_\mu(x|\nu),\mu).$$

So

$$\nu^y 2^{-m-\mathbf{H}_\nu(y|x,m,\mu)} \le 2^{-m}$$
$$\nu^y 2^{-G_{\mu,\nu}(x,y)} \overset{*}{<} 2^{-\mathbf{H}_\mu(x|\nu)}.$$

Integrating over x gives $\mu^x\nu^y 2^{-G_{\mu,\nu}(x,y)} \overset{*}{<} 1$, implying $\mathbf{H}_{\mu,\nu}(x,y) <^+ G_{\mu,\nu}(x,y)$.

To prove the inequality $>^+$, let $f_\nu(x,y,\mu) = 2^{-\mathbf{H}_{\mu,\nu}(x,y)}$. Proposition 18 implies there exists $c \in \mathbb{N}$ with $\nu^y f_\nu(x,y,\mu) \le 2^{-\mathbf{H}_\mu(x|\nu)+c}$. Let $F_\nu(x,\mu) = \lceil \mathbf{H}_\mu(x|\nu)\rceil$. Note that if h is a lower computable function such that $\nu^y h(x,y,\mu) \overset{*}{<} 2^{-\mathbf{H}_\mu(x|\nu)}$, then $\mu^x\nu^y h(x,y,\mu) \overset{*}{<} \mu^x \mathbf{t}_\mu(x|\nu) \overset{*}{<} 1$, so $h \overset{*}{<} f$. Proposition 22 (substituting y for x and (x,μ) for y) gives

$$\mathbf{H}_{\mu,\nu}(x,y) = -\log f_\nu(x,y,\mu) >^+ F_v(x,u) + \mathbf{H}_\nu(y|x,F_\nu(x,u),\mu).$$

□

Chapter 13

Oscillation of Algorithmic Fine Grained Entropy

13.1 On Exotic Sets of Natural Numbers

Lemma 15 *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 $cd2^{-s}$, where $c \in \mathbb{N}$ is chosen later. $\mathbf{E}[p(F)] = \mathbf{E}[q(F)] \leq cd2^{-s}$. Furthermore

$$\mathbf{E}[Q(\{G : |G| = 2^s, G \cap F = \emptyset\})] \leq \sum_G Q(G)(1 - cd2^{-s})^{2^s} < e^{-cd}.$$

Thus finite $W \subset \mathbb{N}$ can be chosen such that $p(W) \leq 4cd2^{-s}$, $q(W) \leq 4cd2^{-s}$, and $Q(\{G : |G| = 2^s, G\cap W = \emptyset\}) \leq e^{2-cd}$. $D\cap W \neq \emptyset$, otherwise, using the Q-test, $t(G) = e^{cd-1}$ if $(|G| = 2^s, G\cap W = \emptyset)$ and $t(G) = 0$ otherwise, we have

$$\begin{aligned}
\mathbf{K}(D|Q, d, c) &<^+ -\log Q(D) - (\log e)cd \\
(\log e)cd &<^+ -\log Q(D) - \mathbf{K}(D|Q) + \mathbf{K}(d, c) \\
(\log e)cd &<^+ d + \mathbf{K}(d, c),
\end{aligned}$$

which is a contradiction for large enough c. 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 6 results in

$$\begin{aligned}
s &<^+ \min\{\mathbf{d}(a|p), \mathbf{d}(a|q)\} + \mathbf{Ks}(D) + O(\mathbf{K}(s, p, q)) \\
s &< \max_{a\in D}\min\{\mathbf{d}(a|p), \mathbf{d}(a|q)\} + \mathbf{Ks}(D) + O(\mathbf{K}(s, p, q)) \\
s &< \max_{a\in D}\min\{\mathbf{d}(a|p), \mathbf{d}(a|q)\} + \mathbf{I}(D;\mathcal{H}) + O(\mathbf{K}(\mathbf{I}(D;\mathcal{H}), s, p, q))).\square
\end{aligned}$$

13.2 On Exotic Sets of Reals

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 16 *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. $\square$

The following lemma, while lengthy, is a series of straightforward application of inequalities.

Lemma 17 *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 s, 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 $k = \mathbf{bb}(n)$. Let p and q be probabilities over $\{0,1\}^*$, where $p(x) = [\|x\| = k]P(x)$ and $\langle p\rangle = \langle k, P\rangle$ and let $q(x) = [\|x\| = k]Q(x)$ and $\langle q\rangle = \langle k, P\rangle$. Using $D = Z_k$, Lemma 15, 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 $\mathbf{I}(Z_k;\mathcal{H}|k) + \mathbf{K}(k) < \mathbf{I}(Z_k;\mathcal{H}) + O(\log n)$,

$$s < \max_{\alpha\in Z}\min\{\mathbf{D}(\alpha|P),\mathbf{D}(\alpha|Q)\} + \mathbf{I}(Z_k;\mathcal{H}) + O(\log n + \mathbf{K}(P,Q)) + O(\log \mathbf{I}(Z_k;\mathcal{H})).$$

By Lemma 16, $\mathbf{K}(\Omega[n]|Z_k) <^+ \mathbf{K}(n)$ so by Lemma 4,

$$n <^{\log} \mathbf{I}(\Omega[n];\mathcal{H}) <^{\log} \mathbf{I}(Z_k;\mathcal{H}) + \mathbf{K}(n) <^{\log} \mathbf{I}(Z_k;\mathcal{H}).$$

The above equation used the common fact that the first n bits of Ω has $n - O(\log n)$ bits of mutual information with $\mathcal{H}$. So

$$s < \max_{\alpha\in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{I}(Z_k;\mathcal{H}) + O(\mathbf{K}(P,Q) + \log \mathbf{I}(Z_k;\mathcal{H})).$$

By the definition of mutual information $\mathbf{I}$ between infinite sequences

$$\mathbf{I}(Z_k;\mathcal{H}) <^+ \mathbf{I}(Z:\mathcal{H}) + \mathbf{K}(Z_k|Z) <^{\log} \mathbf{I}(Z:\mathcal{H}) + \mathbf{K}(k|Z).$$

Now m is simple relative to Z and by Lemma 16, $\Omega[n]$ is simple relative to m and n. Furthermore k is simple relative to $\Omega[n]$. Therefore $\mathbf{K}(Z_k|Z) <^+ \mathbf{K}(n)$. So

$$s < \max_{\alpha\in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{I}(Z:\mathcal{H}) + O(\log n) + O(\mathbf{K}(P,Q) + \log \mathbf{I}(Z;\mathcal{H}))$$
$$s < \max_{\alpha\in Z} \min\{\mathbf{D}(\alpha|P), \mathbf{D}(\alpha|Q)\} + \mathbf{I}(Z:\mathcal{H}) + O(\mathbf{K}(s,P,Q) + \log \mathbf{I}(Z;\mathcal{H}))).$$

□

13.3 Oscillations Occur

Definition 24 (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 $\langle\overrightarrow{x}\rangle = \langle x_1\rangle\langle x_2\rangle\dots$. Each $x_i \in \mathcal{S}$ is an ideal point, and $\langle x_i\rangle$ is its order in the enumeration of $\mathcal{S}$. Each point $x \in \mathcal{X}$ has a certain mutual information with the halting sequence $\mathbf{I}(x:\mathcal{H}) = \inf\{\mathbf{I}(\langle\overrightarrow{x}\rangle:\mathcal{H}) : \langle\overrightarrow{x}\rangle$ is a fast Cauchy sequence for $x\}$.*

Theorem 52 ([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*

$$P_\rho\{\gamma : \mathbf{I}(\langle\gamma,\rho\rangle:\mathcal{H}) > m\} < 2^{\mathbf{I}(\rho:\mathcal{H})-m+c_T}.$$

Theorem 53 (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 48) 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 19), for the computable measure space $(\mathcal{X}, \mu)$. Thus μ_δ is a computable (not necessarily probability) measure over $\{0,1\}^\infty$. By Lemma 13, 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 Claim 1, (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 52, with the universal Turing machine relativized to n^*, and folding the constants together,

$$\begin{aligned}
&\Gamma^{n+c}(\{\gamma : \mathbf{I}(\gamma:\mathcal{H}|n^*) > m\})\\
<&\Gamma^{n+c}(\{\gamma : \mathbf{I}(\langle\gamma,\eta\rangle:\mathcal{H}|n^*) >^+ m\})\\
\stackrel{*}{<}&2^{-m+\mathbf{I}(\eta:\mathcal{H}|n^*)+c_T}\\
\stackrel{*}{<}&2^{-m+\mathbf{K}(n,\mathbf{K}(n),c|n^*)+\mathbf{I}(\xi:\mathcal{H}|n^*)+c_T}\\
\stackrel{*}{<}&2^{-m+\mathbf{K}(c)}.
\end{aligned}$$

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 17, relativized to n^*, on the set $D' = f(D,\xi)$ and probability $\overline{\mu}_\delta$, there exists constants $d, f \in \mathbb{N}$ where

$$\begin{aligned}
m = \log|D| &< \max_{\beta\in D'} \mathbf{D}(\beta|\overline{\mu}_\delta, n^*) + 2\mathbf{I}(D':\mathcal{H}|n^*) + d\mathbf{K}(m|v) + f\mathbf{K}(\overline{\mu}_\delta|n^*)\\
m &< \max_{\beta\in D'} \mathbf{D}(\beta|\overline{\mu}_\delta) + \mathbf{K}(n) + 2\mathbf{I}(D':\mathcal{H}|n^*) + d\mathbf{K}(m|n^*) + f\mathbf{K}(\mu_\delta|n^*)\\
&<^+ \max_{\beta\in D'} \mathbf{D}(\beta|\overline{\mu}_\delta) + \mathbf{K}(n) + 2\mathbf{K}(c) + d\mathbf{K}(m|v) + f\mathbf{K}(\mu_\delta|n^*)\\
&<^+ n + \mathbf{K}(n) + d\mathbf{K}(m|v) + 2\mathbf{K}(c) + (f+1)\mathbf{K}(\mu_\delta). \qquad (13.1)
\end{aligned}$$

Therefore:

$$\begin{aligned}
m &= n + \mathbf{K}(n) + c - 1\\
\mathbf{K}(m|n^*) &<^+ \mathbf{K}(c). \qquad (13.2)
\end{aligned}$$

Plugging Equation 13.2 back into Equation 13.1 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. □

13.4 Oscillations are Rare

Lemma 18 *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

$$\int_{\mathcal{X}}\int_{[0,1]} 2^{-\mathbf{H}_{\mu\times L}(\alpha,t)}dL(t)d\mu(\alpha) = \int_{\mathcal{X}}\int_{[0,1]} \mathbf{t}_{\mu\times L}(\alpha,t)dL(t)d\mu(\alpha) \le 1,$$

the function $f(\alpha) = \int_{[0,1]} 2^{-\mathbf{H}_{\mu\times L}(\alpha,t)}dL(t)$ is a μ-test. So

$$\int_{[0,1]} 2^{-\mathbf{H}_{\mu\times L}(\alpha,t)}dt = f(\alpha) \overset{*}{<} \mathbf{t}_\mu(\alpha) \overset{*}{=} 2^{-\mathbf{H}_\mu(\alpha)}.$$

So

$$\{t \in [0,1] : 2^{-\mathbf{H}_{\mu\times L}(\alpha,t)} > 2^{m-\mathbf{H}_\mu(\alpha)}\} \overset{*}{<} 2^{-m}$$
$$\{t \in [0,1] : \mathbf{H}_{\mu\times L}(\alpha,t) < \mathbf{H}_\mu(\alpha) - m\} \overset{*}{<} 2^{-m}$$

$\mathbf{H}_{\mu\times L}(\alpha,t) <^+ \mathbf{H}_\mu(G^t\alpha)$ because

$$\begin{aligned}
&\int_{[0,1]}\int_{\mathcal{X}} \mathbf{t}_\mu(G^t\alpha)d\mu(\alpha)dL(t)\\
=&\int_{[0,1]}\int_{\mathcal{X}} \mathbf{t}_\mu(\alpha)d\mu(G^{-t}\alpha)dL(t)\\
=&\int_{[0,1]}\int_{\mathcal{X}} \mathbf{t}_\mu(\alpha)d\mu(\alpha)dL(t),\\
=&\int_{[0,1]} 1dL(t)\\
\le&1,
\end{aligned}$$

which means $\mathbf{t}_\mu(G^t\alpha) \overset{*}{<} \mathbf{t}_{\mu\times L}(\alpha,t)$ and thus $2^{-\mathbf{H}_\mu(G^t\alpha)} \overset{*}{<} 2^{-\mathbf{H}_{\mu\times L}(\alpha,t)}$. Thus

$$\{t \in [0,1] : \mathbf{H}_\mu(G^t\alpha) < \mathbf{H}_\mu(\alpha) - m\} \overset{*}{<} 2^{-m}.$$

□

Corollary 17 *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 are constants c_1 and c_2 with*

$$2^{-n-\mathbf{K}(n)-c_1} < L\{t \in [0,1] : \mathbf{H}_\mu(G^t\alpha) < \log\mu(\mathcal{X}) - n\} < 2^{-n+c_2}.$$

Chapter 14

Discrete Dynamics

14.1 Synchronized Oscillations

Discrete dynamics is modeled by a transform group G^t from Definition 22, but with $t \in \mathbb{Z}$, being an integer. We assume there no $\alpha \in \mathcal{X}$ with a finite orbit. Discrete dynamics will visit states with ever increasing $\mathbf{t}_\mu$ and $\mathbf{t}_\nu$ score. 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 $\mathbf{I}(D : \mathcal{H}) = \inf_{\overrightarrow{\alpha_1},\dots,\overrightarrow{\alpha_n}} \mathbf{I}(\langle \overrightarrow{\alpha_1},\dots,\overrightarrow{\alpha_n}\rangle : \mathcal{H})$, which is the infimum over all encoded fast Cauchy sequences to members of D.

Lemma 19 *Given dual computable measure space $(\mathcal{X},\mu,\nu)$, there is a constant $c_{\mathcal{X},\mu,\nu}$, with universal uniform tests $\mathbf{t}_\mu$ and $\mathbf{t}_\nu$, such that $U = \mathbf{t}_\mu(\mathcal{X}) = \mathbf{t}_\nu(\mathcal{X})$, for a finite set $Z \subset \mathcal{X}$ with $n = \lceil \log |Z| \rceil$,*

$$n < \log \max_{\alpha \in Z} \min\{\mathbf{t}_\mu(\alpha), \mathbf{t}_\nu(\alpha)\} + U + \mathbf{I}(\langle Z\rangle : \mathcal{H}) + O(\log \mathbf{I}(\langle Z\rangle : \mathcal{H}) + \mathbf{K}(n) + c_{\mathcal{X},\mu,\nu}).$$

Proof. 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 $t_\nu(\alpha) < \infty$, so by Lemma 13, $\delta^{-1}(\alpha)$ is defined. Let $\overline{\mu_\delta} = \mu_\delta/U$ and $\overline{\nu_\delta} = \nu_\delta/U$. Let $W = \delta^{-1}(Z) \subset \{0,1\}^\infty$. By Theorem 17 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 Claim 1,

$$n < \max_{\alpha\in W} \min\{\log \mathbf{t}_{\overline{\mu_\delta}}(\alpha), \log \mathbf{t}_{\overline{\nu_\delta}}(\alpha)\} + \mathbf{I}(W : \mathcal{H}) + O(\log \mathbf{I}(W : \mathcal{H}) + \mathbf{K}(n) + c_{\mathcal{X},\mu}),$$

$$n < \max_{\alpha\in W} \min\{\log \mathbf{t}_{\mu_\delta}(\alpha), \log \mathbf{t}_{\nu_\delta}(\alpha)\} + U + \mathbf{I}(W : \mathcal{H}) + O(\log \mathbf{I}(W : \mathcal{H}) + \mathbf{K}(n) + c_{\mathcal{X},\mu}).$$

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 13,

$$n < \max_{\alpha\in Z} \min\{\log \mathbf{t}_{\mu}(\alpha), \log \mathbf{t}_{\nu}(\alpha)\} + U + \mathbf{I}(W : \mathcal{H}) + O(\log \mathbf{I}(W : \mathcal{H}) + \mathbf{K}(n) + c_{\mathcal{X},\mu}).$$

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

$$n < \max_{\alpha\in Z} \min\{\log \mathbf{t}_{\mu}(\alpha), \log \mathbf{t}_{\nu}(\alpha)\} + U + \mathbf{I}(Z : \mathcal{H}) + O(\log \mathbf{I}(Z : \mathcal{H}) + \mathbf{K}(n) + c_{\mathcal{X},\mu}).$$

□

Theorem 54 *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*

$$\max_{\gamma\in G^{\{1,\dots,2^n\}}\alpha} \max\{\mathbf{H}_\mu(\gamma),\mathbf{H}_\nu(\gamma)\} < U - n + O(\mathbf{K}(n)) + c.$$

Proof. Let $Z_n = G^{\{1,\dots,2^n\}}\alpha$. Lemma 19, applied to $(\mathcal{X},\mu,\nu)$ and Z_n, results in $\gamma \in Z_n$ such that

$$n < \min\{\log \mathbf{t}_\mu(\gamma), \log \mathbf{t}_\nu(\gamma)\} + U + \mathbf{I}(Z_n:\mathcal{H}) + O(\log \mathbf{I}(Z_n:\mathcal{H}) + \mathbf{K}(n) + c_{\mathcal{X},\mu,\nu,\alpha}).$$

Since $\mathbf{I}(Z_n:\mathcal{H}) <^+ \mathbf{I}(\{\alpha\}:\mathcal{H}) + \mathbf{K}(n)$,

$$n < \min\{\log \mathbf{t}_\mu(\gamma), \log \mathbf{t}_\nu(\gamma)\} + U + \mathbf{I}(\alpha:\mathcal{H}) + O(\log \mathbf{I}(\alpha:\mathcal{H}) + \mathbf{K}(n) + c_{\mathcal{X},\mu,\nu,\alpha,G}).$$

The theorem is proven by noting $\mathbf{I}(\alpha:\mathcal{H}) < \infty$. □

14.2 Ergodic Dynamics

For a measure space $(\mathcal{X},\mu)$ a discrete transformation group G^t is ergodic if all the invariant sets have measure 0 or $\mu(X)$. The following theorem adapts Theorem y 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 7 is that overlapping open balls are used instead of cylinders.

Definition 25 *A set D is an ad-set if it is a finite union of almost decidable balls, with $D = B_{i_1} \cup \dots B_{i_k}$. We have $\overline{\overline{D}} = \overline{\overline{B}}_{i_1} \cup \dots \overline{\overline{B}}_{i_k}$, which may differ than the closure of D if there are isolated points.*

Proposition 23 *For computable measure space $(\mathcal{X},\mu)$ and ad-set D, $\mu(D)$ is computable.*

Proof. This follows from Proposition 11, 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.

14.2.1 Single Points

Theorem 55 *Let $(\mathcal{X},\mu)$ be a computable measure space such that $\mu(X)$ is computable. Let G^t be an ergodic discrete transformation group. 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 $G^i(x) \in A$ 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 20. We recall that from Corollary 14, 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 $\mathcal{X}$. 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 $\nu(B_j \cap \bigcap_{i\leq n} G^{-i}(A)) \leq r\nu(B)$. Note that it could be that $B_j \cap B_k \neq \emptyset$ for $j \neq k$. This gives an effective open cover of $A^* \cap I$ 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 $\nu(B\cap\bigcap_{i\leq n} G^{-i}(A))$, we note that it does not exceed $\min i \leq n\nu(B\cap G^{-i}(A))$ which does not exceed $\frac{1}{n+1}\sum_{i\leq n}\nu(B\cap G^{-i}(A))$. This average,

$$\frac{1}{n+1}\left[\nu(B\cap A)+\nu(B\cap G^{-1}(A)+\cdots+\nu(B\cap G^{-n}(A))\right] \tag{14.1}$$

is equal to

$$\frac{1}{n+1}\left[\nu(G^{-n}(B)\cap G^{-n}(A))+\nu(G^{-(n-1))}(B)\cap G^{-n}(A)+\cdots+\nu(B\cap G^{-n}(A))\right],$$

because G is measure preserving. The latter expression is the scalar product of the indicator function of $G^{-n}(A)$ and the average $a_n = (\mathbf{1}_0+\cdots+\mathbf{1}_n)/(n+1)$, where $\mathbf{1}_i$ is the indicator of $G^{-i}(B)$.

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-Schwars 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 $G^{-i}(B)$ is an effectively open set of measure $\nu(B)$, and, since G is measure preserving, $\nu(B)$ is computable. There for any i and $\epsilon>0$, one can unformly approximate $G^{-i}(B)$ by its ad-set subset U, where $\nu(G^{-i}(B)\setminus U)<\epsilon$ can be computed, due to Proposition 23. This means that the L_2-distance 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 14.1) is less than $r\nu(B)$. We then have $\nu(B\cap\bigcap_{i\leq n} G^{-i}(A))<r\nu(B)$. □

The above theorem has implications for algorithmic coarse grain entropy and in particular Theorem 59 which says that if a state travels through enough partitions (effective open sets) then oscillations will occcur. Theorem 55 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.

14.2.2 Indicator Functions

The following theorem adapts Theorem 8 from [BDH+12] to computable measure space using the proof sketch in Theorem 12.

Theorem 56 *Let $(\mathcal{X},\mu)$ be a computable measure space with computable $\mu(X)$. Let G^t be an ergodic transformation group. Let U be an effectively open set. If $\mathbf{H}_\mu(\omega)\neq-\infty$ then*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\mathbf{1}_U(G^k(\omega)) = \mu(U)/\mu(X).$$

This also applies to effectively closed sets.

Proof. Let $\nu(x)=\mu(x)/\mu(X)$ be a computable probability measure over $\mathcal{X}$. Let $g_n(\omega)=\frac{1}{n}\sum_{k=0}^{n-1}\mathbf{1}_U(G^k(\omega))$ be the frequency of U elements among the first n iterations of ω. We first prove $\limsup g_n(\omega)\leq\nu(U)$. We then prove $\liminf g_n(\omega)\geq\nu(U)$.

Let $r>\nu(U)$ be some rational number and $G_N=\{\omega : (\exists n\geq N)g_n(\omega)>r\}$ 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 55 to effectively open set G_n we get that for $\omega \in U$ with $\mathbf{H}_\mu(\omega) \neq -\infty$ and k such that $G^k(\omega) \notin G_N$. So $\limsup_n g_n(G^k(\omega)) \leq r$ Since finite number of iterations does not change lim sup, we have $g_n(\omega) \leq r$. Since r was an arbitray rational number greater than $\nu(U)$ so $\limsup g_n(\omega) \geq \nu(U)$.

(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 24 *For computable non-atomic measure space $(\mathcal{X}, \mu)$,*

1. *Every set E with $\mu(E) > 0$ contains sets of arbitrarily small positive measure.*
2. *For any $\delta \in [0, \mu(X)]$, there exists an open 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 $\sup_{C \in \mathcal{C}} \mu(C)$ □

Proposition 25 *Given non-atomic computable measurable space $(\mathcal{X}, \mu)$ with computable $\mu(X)$, 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 Proposition 24, 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 $n \neq m$. 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 Proposition 24, 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 $t \overset{*}{<} \mathbf{t}_\mu$. 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 $\mu(X)2^{-n-\mathbf{K}(n)-c} < \mu(\{x : \mathbf{H}_\mu(x) < \log \mu(X) - n\}) < \mu(X)2^{-n}$. □

The following shows that during the course of discrete 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 53 in its inequalities.

Theorem 57 (Discrete Oscillations) *Let $(\mathcal{X}, \mu)$ be a non-atomic computable measure space with computable $\mu(X)$. There is a $c \in \mathbb{N}$ with the following properties. Let G^t be a ergodic transformation group, and $U_n = \{x : \mathbf{H}_\mu(x) < \log \mu(X) - n\}$. If $\omega \in X$ has $\mathbf{H}_\mu(\omega) \neq -\infty$,*

$$2^{-n-\mathbf{K}(n)-c} < \lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} \mathbf{1}_{U_n}(G^t(\omega)) < 2^{-n}.$$

Proof. By Proposition 25, there is a c where $\mu(X)2^{-n-\mathbf{K}(n)-c} < \mu(U_n) < \mu(X)2^{-n}$. By Theorem 56, $\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_{U_n}(G^k(\omega)) = \mu(U_n)/\mu(X)$. So

$$2^{-n-\mathbf{K}(n)-c} < \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_{U_n}(G^k(\omega)) < 2^{-n}.$$

14.2.3 Lower Computable Functions

Theorem 58 *Let $(\mathcal{X}, \mu)$ be a computable measure space with computable $\mu(X)$. Let G^t be an ergodic transformation group. Let $f : X \to \overline{R}^+ \cup \infty$ be lower computable. If $\mathbf{H}_\mu(\omega) \neq -\infty$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n=1} f(G^t(\omega)) = \frac{1}{\mu(X)} \int f d\mu.$$

Proof. Let $\nu = \mu/\mu(\mathcal{X})$ be a computable measure over $\mathcal{X}$. Let f be a lower computable function with a finite integral. Let $f_n = \frac{1}{n}(f + \cdots + f \cdot G^{n-1}$. Let $r > \int f d\nu$ be a ration number and $G_N = \{\omega : (\exists n \geq N) f_n(\omega) > r\}$. The set G_N is effectively open and $\nu(\bigcup_N G_N) = 0$ 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 55, if $\mathbf{H}_\nu(\omega) \neq -\infty$, then there exists k such that $G^k(\omega) \neq G_n$. So $\limsup f_n(G^k(\omega)) \leq r$ and $\limsup f_n(\omega) = \limsup f_n(G^k(\omega)) \leq r$. 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 $\hat{f}(\omega) = \sum_{i=1}^n c_n c_n \mathbf{1}_{B_n}(\omega)$, 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 56, and their integrals can be arbitrarily close to $\int f \nu = \frac{1}{\mu(X)} \int f d\mu$. □

Corollary 18 *Let $(\mathcal{X}, \mu)$ be a computable measure space with computable $\mu(X)$ and G be an ergodic transformation group. For $\omega \in X$ with $\mathbf{H}_\mu(\omega) \neq -\infty$, then*

$$\lim_{n\to\infty} \mu(X) \sum_{t=0}^{n-1} 2^{-\mathbf{H}_\mu(G^t\omega)} < 1.$$

Proof. This follows from Theorem 58 and the fact that $2^{-\mathbf{H}_\mu(\omega)} = \mathbf{t}_\mu(\omega)$ is lower computable and $\int \mathbf{t} d\mu < 1$. $\square$

Chapter 15

Algorithmic 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 11. Let $\Pi(\cdot)$ be a set of disjoint uniformly enumerable open sets in the computable metric space $\mathcal{X}$.

Definition 26 (Algorithmic Coarse Grained Entropy) $\mathbf{H}_\mu(\Pi_i) = \mathbf{K}(i|\mu) + \log \mu(\Pi_i)$.

Coarse grained entropy is an excellent approximation of fine grained entropy, as shown by the following two results.

Proposition 26 *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 $t(\alpha) = [\alpha \in \Pi_i]\mathbf{m}(\Pi_i)/\mu(\Pi_i)$. $t \in \mathcal{F}$ 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 $\mathbf{t}_\mu(\alpha) \overset{*}{>} t(\alpha)$.

Lemma 20 *For computable measure space $(\mathcal{X}, \mu)$, for lower computable function $f \in \mathcal{F}$, 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 10, 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

$$\int_U \sup\{f_{i_1,j_1}, \dots f_{i_k,j_k}\} d\mu = q_{j_m} \mu\left((B_{i_1} \cup \dots \cup B_{i_k}) \cap U\right) + \int_U \sup\{f_{i_1,j_1'}, \dots f_{i_k,j_k'}\} d\mu,$$

where q_{j_m} is minimal among $\{q_{j_1}, \dots, q_{j_k}\}$ and $q_{j_1'} = q_{j_1} - q_{i_m}, \dots, q_{j_k'} = q_{j_k} - q_{i_k}$. 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 21 *For computable measure space $(\mathcal{X}, \mu)$, $\mu\{\alpha \in \Pi_i : \mathbf{H}_\mu(\alpha) < \mathbf{H}_\mu(\Pi_i) - \mathbf{K}(\Pi) - m\} \overset{*}{<} 2^{-m}\mu(\Pi_i)$.*

Proof. Let $f(i) = \int_{\Pi_i} \mathbf{t}_\mu(\alpha)d\mu(\alpha)$. By Lemma 20, the function $f(i)$ is lower computable, and $\sum_i f(i) \leq 1$. Thus $f(i) \stackrel{*}{<} \mathbf{m}(i)/\mathbf{m}(\Pi)$. 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 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 19 *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 59 *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*

$$\min_{i\in\{1,\ldots,2^n\}} \mathbf{H}_\mu(i) <^{\log} \max_{i\in\{1,\ldots,2^n\}} \mathbf{H}_\mu(i) - n + \mathbf{I}(\alpha : \mathcal{H}).$$

Proof. Let $f(i) = \lceil \log \mu(\mathbf{i}) \rceil$. Let $D \subset \mathbb{N}$, $|D| = 2^n$ be a set of partitions that α travels through, so $\mathbf{K}(D|\alpha) <^+ \mathbf{K}(n)$. Theorem 61, on $f : D \to \mathbb{N}$, produces $x \in D$, where

$$\begin{aligned}
f(x) + \mathbf{K}(x) &<^{\log} -\log \sum_{a\in D} \mathbf{m}(a)2^{-f(a)} + \mathbf{I}(f, D; \mathcal{H}) \\
f(x) + \mathbf{K}(x) + n &<^{\log} \max_{a\in D} f(a) + \mathbf{K}(a) + \mathbf{I}(f, D; \mathcal{H}) \\
\mathbf{H}_\mu(x) + n &<^{\log} \max_{a\in D} \mathbf{H}_\mu(a) + \mathbf{I}(\alpha : \mathcal{H}).
\end{aligned}$$

Bibliography

[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.

[Bho21] T. Bhojraj. Algorithmic randomness and kolmogorov complexity for qubits. *CoRR*, abs/2106.14280, 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.

[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.

[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.

[Eps22] S. Epstein. The outlier theorem revisited. *CoRR*, abs/2203.08733, 2022.

[Eps23a] S. Epstein. On the algorithmic information between probabilities. *CoRR*, abs/2303.07296, 2023.

[Eps23b] S. Epstein. A quantum el theorem. *CoRR*, abs/2301.08348, 2023.

[Eps23c] 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.

[G6̈1] 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.

[G0́1] P. Gács. Quantum Algorithmic Entropy. *Journal of Physics A Mathematical General*, 34(35), 2001.

[G2́1] 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.

[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.

[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.

[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.

[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.

[Mul08] M. Muller. Strongly Universal Quantum Turing Machines and Invariance of Kolmogorov Complexity. *IEEE Transactions on Information Theory*, 54(2), 2008.

[NS19] A. Nies and V. Scholz. Martin-Löf Random Quantum States. *Journal of Mathematical Physics*, 60(9), 2019.

[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.

[She15] M. Shell. *Thermodynamics and Statistical Mechanics: An Integrated Approach*. Cambridge Series in Chemical Engineering. Cambridge University Press, 2015.

[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.

[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.

Part IV

Appendix

Appendix A

Information Between Basis States

Special information inequalities and conservation inequalities can be achieved for orthogonal sequences of pure basis states $|1\rangle, |2\rangle, |3\rangle, \ldots$ We use Theorem 8 that $\mathbf{K}(i|n) =^+ \mathbf{Hg}(|i\rangle)$. Let $S(\rho)$ be the von Neumann entropy of σ.

Theorem 60 *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. $\mathbf{d}(|i\rangle \,|\, |j\rangle) = \infty$ *for* $i \neq j$.
2. $\mathbf{Hg}(\sigma \otimes |i\rangle\langle i|) =^+ \mathbf{Hg}(|i\rangle) + \mathbf{Hg}(\sigma| \,|i\rangle, \mathbf{Hg}(|i\rangle))$.
3. $\mathbf{I}(|k\rangle : |i\rangle) <^+ \mathbf{I}(|j\rangle : |i\rangle) + \mathbf{K}(k|j, N)$.
4. $\mathbf{I}(i : j|N) <^{\log} \mathbf{I}(|i\rangle : |j\rangle)$.
5. $\mathbf{I}(|i\rangle : |j\rangle) <^+ \mathbf{I}(i : j|N) + \mathbf{I}(i, j : \mathcal{H}|N)$.
6. $\mathbf{K}(i|N) <^+ \mathbf{I}(|i\rangle : |i\rangle) <^+ \mathbf{K}(i|N) + \mathbf{I}(i : \mathcal{H}|N)$.
7. $\mathbf{I}(|i\rangle : |i\rangle) <^+ 4n|3$.
8. $\mathbf{I}(|i\rangle : |j\rangle) <^+ \mathbf{I}(|i\rangle : |i\rangle) + \mathbf{I}(i, j : \mathcal{H}|n)$.

Proof.

(1) This is due to the fact that $\mathfrak{T}_{|i\rangle\langle i|} \overset{*}{>} \sum_n \mathbf{m}(n) n\, |j\rangle\langle j|$. Thus $\log \mathrm{Tr}\mathfrak{T}_{|i\rangle\langle i|}|j\rangle\langle j| = \infty$.

(2) We use the lower semicomputable matrix $\rho = \boldsymbol{\mu}_{(|i\rangle, \mathbf{Hg}(|i\rangle))} \otimes |i\rangle\langle i|$. so we have that $\boldsymbol{\mu}_{2n} \overset{*}{>} \mathbf{m}(|i\rangle, \mathbf{Hg}(|i\rangle)|2^{2n})\rho \overset{*}{=} \mathbf{m}(i, \mathbf{K}(i|n)|2^{2n})\rho \overset{*}{=} \mathbf{m}(i|2^n)\rho \overset{*}{=} 2^{-\mathbf{Hg}(|i\rangle)}\rho$. So $\mathbf{Hg}(\sigma \otimes |i\rangle\langle i|) <^+ \mathbf{Hg}(|i\rangle) - \log \mathrm{Tr}\rho(\sigma \otimes |i\rangle\langle i|) <^+ \mathbf{Hg}(|i\rangle) + \mathbf{Hg}(\sigma| \,|i\rangle, \mathbf{Hg}(|i\rangle))$. The other direction is given by Theorem 12.

(3) We let $T \in \mathcal{T}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}$ vary over $\boldsymbol{\mu} \otimes \boldsymbol{\mu}$ tests. Since $\boldsymbol{\mu} \overset{*}{>} \sum_i |i\rangle \mathbf{m}(i|N)\langle i|$ it must be that $1 > \mathrm{Tr}T\boldsymbol{\mu} \otimes \boldsymbol{\mu} \overset{*}{>} \mathrm{Tr}T \sum_{i,j} \mathbf{m}(i|N)\mathbf{m}(j|N)\, |i\rangle|j\rangle\langle i|\langle j| \overset{*}{>} \sum_{i,j} \mathbf{m}(i|N)\mathbf{m}(j|N)\langle i|\langle j|\, T\, |i\rangle|j\rangle$.

From each T, let $T' = \sum_{i,j}\left(\sum_k \mathbf{m}(k|j,N)\langle i|\langle k|T|i\rangle|k\rangle\right)|i\rangle|j\rangle\langle i|\langle j|$. Since $\mathbf{m}$ and T are lower computable, then so is T'. In addition, $O(1)T' \in \mathcal{T}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}$, because

$$\begin{aligned}
\mathrm{Tr}T'\boldsymbol{\mu}\otimes\boldsymbol{\mu} &= \sum_{i,j}\langle i|\langle j|\boldsymbol{\mu}\otimes\boldsymbol{\mu}|i\rangle|j\rangle\sum_k \mathbf{m}(k|j,N)(\langle i|\langle k|T|i\rangle|k\rangle)\\
&\stackrel{*}{=}\sum_{i,j}\mathbf{m}(i|N)\mathbf{m}(j|N)\sum_k \mathbf{m}(k|j,N)(\langle i|\langle k|T|i\rangle|k\rangle)\\
&\stackrel{*}{<}\sum_{i,j}\mathbf{m}(i|N)\sum_k \mathbf{m}(j,k|N)(\langle i|\langle k|T|i\rangle|k\rangle)\\
&\stackrel{*}{=}\sum_{i,j}\mathbf{m}(i|N)\sum_k \mathbf{m}(k|N)\mathbf{m}(j|k,\mathbf{K}(k),N)(\langle i|\langle k|T|i\rangle|k\rangle)\\
&\stackrel{*}{=}\sum_{i,k}\mathbf{m}(i|N)\mathbf{m}(k|N)\sum_j \mathbf{m}(j|k,\mathbf{K}(k),N)\langle i|\langle k|T|i\rangle|k\rangle\\
&\stackrel{*}{<}\sum_{i,k}\mathbf{m}(i|N)\mathbf{m}(k|N)(\langle i|\langle k|T|i\rangle|k\rangle)\\
&\stackrel{*}{<}\mathrm{Tr}T\boldsymbol{\mu}\otimes\boldsymbol{\mu}\\
&< O(1).
\end{aligned}$$

So

$$\begin{aligned}
\mathbf{I}(|i\rangle:|j\rangle) &> \log\mathrm{Tr}\sum_T \mathbf{m}(O(1)T'|N^2)O(1)T'|i\rangle|j\rangle\langle i|\langle j|\\
&>^+ \log\mathrm{Tr}\sum_T \mathbf{m}(T|N^2)T'|i\rangle|j\rangle\langle i|\langle j|\\
&>^+ \log\sum_T \mathbf{m}(T|N^2)\sum_k \mathbf{m}(k|j,N)\langle k|\langle i|T|k\rangle|i\rangle\\
&>^+ \log\sum_T \mathbf{m}(T|N^2)\mathbf{m}(k|j,N)\langle k|\langle i|T|k\rangle|i\rangle\\
&=^+ \mathbf{I}(|k\rangle:|i\rangle)-\mathbf{K}(k|j).
\end{aligned}$$

(4) This follow as a special case of Theorem 34.

(5) Let $s(i,j) = \mathbf{m}(i|N)\mathbf{m}(j|N)2^{\mathbf{I}(|i\rangle:|j\rangle)}$. The function s is lower semicomputable relative to $\mathcal{H}$ because $\mathbf{m}$ and $\mathfrak{T}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}$ are lower computable relative to $\mathcal{H}$. Furthermore we have that

$$\begin{aligned}
\sum_{i,j}s(i,j) &= \sum_{i,j}\mathbf{m}(i|N)\mathbf{m}(j|N)\mathrm{Tr}\mathfrak{T}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}|i\rangle\langle i|\otimes|j\rangle\langle j|\\
&=\mathrm{Tr}\mathfrak{T}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}\sum_{i,j}\mathbf{m}(i|N))|i\rangle\langle i|\otimes\mathbf{m}(j|N)|j\rangle\langle j|\\
&<O(1)\mathrm{Tr}\mathfrak{T}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}\boldsymbol{\mu}\otimes\boldsymbol{\mu} < O(1).
\end{aligned}$$

Therefore $s(i,j) \stackrel{*}{<} \mathbf{m}(i,j|N,\mathcal{H})$ and so $\mathbf{I}(|i\rangle:|j\rangle) <^+ \log\mathbf{m}(i,j|N,\mathcal{H})|(\mathbf{m}(i|N)\mathbf{m}(j|N)) =^+ \mathbf{I}(i:j|N)+\mathbf{I}(i,j:\mathcal{H}|N)$.

(6) For $\mathbf{K}(i|N) <^+ \mathbf{I}(|i\rangle:|i\rangle)$, we prove the stronger statement: for elementary ρ, $2\mathbf{Hg}(\rho)-\mathbf{K}(\rho,\mathbf{Hg}(\rho))-2S(\rho) <^+ \mathbf{I}(\rho:\rho)$. Let $\nu = 2^{2\mathbf{Hg}(\rho)-2}(\rho\otimes\rho)$. The matrix $\nu \in \mathcal{T}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}$ because

$\mathrm{Tr}(\boldsymbol{\mu}\otimes\boldsymbol{\mu})\nu \leq 1$. Therefore

$$\begin{aligned}\mathbf{I}(\rho:\rho) &= \log \mathrm{Tr}\mathfrak{T}_{\boldsymbol{\mu}\otimes\boldsymbol{\mu}}(\rho\otimes\rho)\\ &\geq \log \mathbf{m}(\nu)\mathrm{Tr}\nu(\rho\otimes\rho)\\ &>^{+} \log \mathbf{m}(\nu)2^{2\mathbf{Hg}(\rho)}\mathrm{Tr}(\rho\rho\otimes\rho\rho)\\ &>^{+} 2\mathbf{Hg}(\rho)-\mathbf{K}(\rho,\mathbf{Hg}(\rho))+2\log\sum_i \lambda_i^2,\end{aligned}$$

where λ_i are the eigenvalues of ρ. Due to concavity $-2S(\rho) \leq 2\log\sum_i \lambda_i^2$. So $\mathbf{I}(\rho:\rho) >^{+} 2\mathbf{Hg}(\rho)-\mathbf{K}(\rho,\mathbf{Hg}(\rho))-2S(\rho)$. The inequality follows from $\mathbf{Hg}(|i\rangle) =^{+} \mathbf{K}(i)$ and $S(|i\rangle\langle i|)=0$. For $\mathbf{I}(|i\rangle:|i\rangle <^{+} \mathbf{K}(i)+\mathbf{I}(i:\mathcal{H}|n)$, we note that it is a special case of (5).

(7) If $T \in \mathcal{T}_{\boldsymbol{\mu}_n\otimes\boldsymbol{\mu}_n}$, then $\mathrm{Tr}T \overset{*}{<} 2^{2n}$. This is because $1 \geq \mathrm{Tr}T(\boldsymbol{\mu}_n\otimes\boldsymbol{\mu}_n) \overset{*}{>} \mathrm{Tr}T(2^{-n}I\otimes 2^{-n}I)$. Since the set of lower computable matrices of trace not more than 2^{2n} is enumerable, $2^{-2n}\mathfrak{T}_{\boldsymbol{\mu}_n\otimes\boldsymbol{\mu}_n} \overset{*}{<} \boldsymbol{\mu}_{2n}$. Assume $\mathbf{I}(|i\rangle:|i\rangle) = \log\mathrm{Tr}\mathfrak{T}_{\boldsymbol{\mu}_n\otimes\boldsymbol{\mu}_n}|ii\rangle\langle ii| = 2n-c$. Then $-\log\mathrm{Tr}u_{2n}|ii\rangle\langle ii| <^{+} c$. This means that $\mathbf{K}(ii|2^{2n}) <^{+} c$. So $\mathbf{K}(i|2^n) <^{+} c$. So $1 \geq \mathrm{Tr}\mathfrak{T}_{\boldsymbol{\mu}_n\otimes\boldsymbol{\mu}_n}\boldsymbol{\mu}\otimes\boldsymbol{\mu} \overset{*}{>} \mathbf{m}(i|2^n)^2\mathrm{Tr}\mathfrak{T}_{\boldsymbol{\mu}_n\otimes\boldsymbol{\mu}_n}|ii\rangle\langle ii| \overset{*}{>} 2^{2n-3c}$. Thus $c >^{+} 2n/3$. This implies that $\mathbf{I}(|i\rangle:|i\rangle) = 2n-c <^{+} 4n/3$.

(8) This follows from (5) and (6).

Corollary 20 *For elementary ρ,* $2\mathbf{Hg}(\rho)-\mathbf{K}(\rho,\mathbf{Hg}(\rho))-2S(\rho) <^{+} \mathbf{I}(\rho:\rho)$.

Appendix B

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.

B.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 B.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

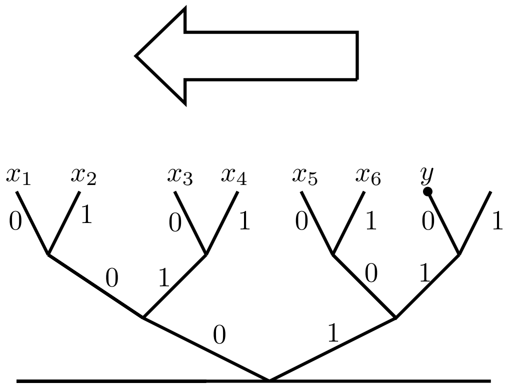

Figure B.1: The above diagram represents the domain of a left total machine T with the 0 bits branching to the left and the 1 bits branching to the right. For $i \in \{1..5\}$, $x_i \lhd x_{i+1}$ and $x_i \lhd y$. Assuming $T(y)$ halts, each x_i is total. This also implies each x_i^- is total as well.

[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'.

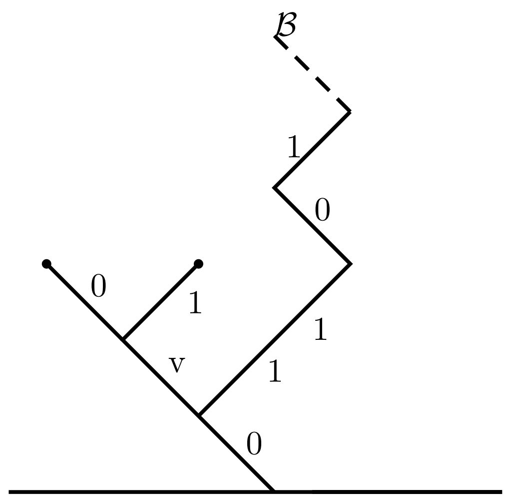

Figure B.2: The above diagram represents the domain of the universal left-total algorithm U, with the 0 bits branching to the left and the 1 bits branching to the right. The strings in the above diagram, $0v0$ and $0v1$, are halting inputs to U with $U(0v0) \neq \perp$ and $U(0v1) \neq \perp$. So $0v$ 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 B.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.

B.2 Extended Coding Theorem

Lemma 22 *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 $g_{\le n} = \cup_{i=0}^n g_i$, for $n \in \mathbb{W} \cup \{\infty\}$. Let $s = \lceil -\log \sum_{a \in f_{\le\infty}} m(a)2^{-f(a)} \rceil$. Using the reasoning of Markov's inequality,

$$\sum_{a \in f_{\le\infty}} m(a)2^{-f(a)} \ge 2^{-s}, \tag{B.1}$$

$$\sum_{a \in f_{\le\infty}\setminus f_{\le s}} m(a)2^{-f(a)} \le \sum_{a \in f_{\le\infty}\setminus f_{\le s}} m(a)2^{-s-1} \le 2^{-s-1}, \tag{B.2}$$

$$\sum_{a \in f_{\le s}} m(a)2^{-f(a)} \ge 2^{-s-1}. \tag{B.3}$$

Equation (B.1) follows from the definition of s and Equation (B.3) follows from Equations (B.1) and (B.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 $\sum_{a \in g_{\le s}} m(a)2^{-g(a)} \ge 2^{-s-1}$.

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 $\sum_{a \in g_{\le s}} m(a)2^{-g(a)} \ge 2^{-s-1}$. Thus $Q(\langle g \rangle) = [g \in S]Q'(g)/Q'(S)$, where $S \subset \{0,1\}^*$, the support of Q, is defined as $S = \{\langle g \rangle : g \in \mathrm{Supp}(Q'), \sum_{a \in g_{\le s}} m(a)2^{-g(a)} \ge 2^{-s-1}\}$. 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

$$\begin{aligned}
\|v\| &<^+ \|v'\| + \mathbf{K}(s),\\
\|v\| + 3\log d &<^+ \|v'\| + \mathbf{K}(s) + 3\log d\\
&<^+ \|v'\| + \mathbf{K}(s) + 3\log(\max\{-\log Q(f) - \mathbf{K}(f|v),1\})\\
&<^+ \|v'\| + \mathbf{K}(s) + 3\log(\max\{-\log Q'(f) - \mathbf{K}(f|v),1\}) && \text{(B.4)}\\
&<^+ \|v'\| + \mathbf{K}(s) + 3\log(\max\{-\log Q'(f) - \mathbf{K}(f|v') + \mathbf{K}(v|v'),1\}) && \text{(B.5)}\\
&<^+ \|v'\| + \mathbf{K}(s) + 3\log(\max\{-\log Q'(f) - \mathbf{K}(f|v') + \mathbf{K}(s),1\}) && \text{(B.6)}\\
&<^{\log} \|v'\| + \mathbf{K}(s) + 3\log(\max\{-\log Q'(f) - \mathbf{K}(f|v'),1\}),\\
\|v\| + 3\log d &<^{\log} \mathbf{Ks}(f) + \mathbf{K}(s). && \text{(B.7)}
\end{aligned}$$

Equation (B.4) follows from $Q(f) = Q'(f)/Q'(\mathrm{Supp}(Q))$, and thus $-\log Q(f) \leq -\log Q'(f)$. Equation (B.5) follows from the inequality $\mathbf{K}(f|v') <^+ \mathbf{K}(f|v) + \mathbf{K}(v|v')$. Equation (B.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

$$\begin{aligned}
\mathbf{E}_{g\sim Q}\mathbf{E}_{A\sim w}[\mathbf{1}(g,A)] &= \sum_g Q(g)\prod_{n=0}^s (1-m(g_n))^{|A^n|}\\
&\leq \sum_g Q(g)\prod_{n=0}^s \exp\{-|A^n|m(g_n)\} && \text{(B.8)}\\
&= \sum_g Q(g)\exp\left\{-\sum_{n=0}^s |A^n|m(g_n)\right\}\\
&= \sum_g Q(g)\exp\left\{-\sum_{n=0}^s cd2^{s+1-n}m(g_n)\right\}\\
&= \sum_g Q(g)\exp\left\{-cd2^{s+1}\sum_{n=0}^s m(g_n)2^{-n}\right\}\\
\mathbf{E}_{g\sim Q}\mathbf{E}_{A\sim\lambda}[\mathbf{1}(g,A)] &\leq \sum_g Q(g)\exp\{-cd\} = \exp\{-cd\}. && \text{(B.9)}
\end{aligned}$$

Equation (B.8) follows from the inequality $(1-a)\leq e^{-a}$ over $a \in [0,1]$. Equation (B.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 (B.9). So

$$\mathbf{K}(A) <^+ \mathbf{K}(c, d, v). \tag{B.10}$$

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,

$$\begin{aligned}
\mathbf{d}(g|Q, v) &= \lfloor -\log Q(g)\rfloor - \mathbf{K}(g|v)\\
&> -\log Q(g) - (-\log \mathbf{1}(g, A)\lfloor e^{cd}\rfloor Q(g) + \mathbf{K}(\mathbf{1}(\cdot, A)\lfloor e^{cd}\rfloor Q(\cdot)|v)) - O(1) \qquad \text{(B.11)}\\
&> cd\log e - \mathbf{K}(\mathbf{1}(\cdot, A)\lfloor e^{cd}\rfloor Q(\cdot)|v) - O(1)\\
&> cd\log -\mathbf{K}(A, c, d|v) - O(1)\\
&> cd\log e - \mathbf{K}(c, d) > d. \qquad \text{(B.12)}
\end{aligned}$$

With c being chosen, it is removed from consideration for the rest of the proof, with $c \in \O(1)$. Equation B.11 is due to the fact that for any elementary semimeasure P, $\mathbf{K}(x) <^+ \mathbf{K}(P) - \log P(x)$. Equation B.12 is due to Equation B.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

$$\begin{aligned}
\mathbf{K}(a) &<^+ \log|A^n| + \mathbf{K}(A^n)\\
&<^+ \log|A^n| + \mathbf{K}(A) + \mathbf{K}(A^n|A)\\
&<^+ (\log d + s - n) + \mathbf{K}(d, v) + \mathbf{K}(n) \qquad \text{(B.13)}\\
&=^+ \log d + s - f(a) + \mathbf{K}(d, v) + \mathbf{K}(f(a))\\
\mathbf{K}(a) + f(a) &<^+ \log d + s + \mathbf{K}(v) + \mathbf{K}(d) + \mathbf{K}(f(a))\\
\mathbf{K}(a) + f(a) &<^{\log} s + \|v\| + 3\log d \qquad \text{(B.14)}\\
\mathbf{K}(a) + f(a) &<^{\log} s + \mathcal{H}(f) \qquad \text{(B.15)}\\
\min_{a\in f_{\leq\infty}} \mathbf{K}(a) + f(a) &<^{\log} -\log \sum_{a\in f_{\leq\infty}} m(a)2^{-f(a)} + \mathcal{H}(f). \qquad \text{(B.16)}
\end{aligned}$$

Equation (B.13) follows from Equation (B.10), and from $c \in O(1)$. Equation (B.14) follows from $\mathbf{K}(x) <^{\log} \|x\|$ for $x \in \{0,1\}^* \cup \mathbb{W}$. Equation (B.15) follows directly from Equation (B.7). Equation (B.16) follows from the definition of s and its form proves the theorem.

Proposition 27 *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 28 *If $b \in \{0,1\}^*$ is total and b^- is not total, then b^- is a border prefix, with $b^- \sqsubset \mathcal{B}$.*

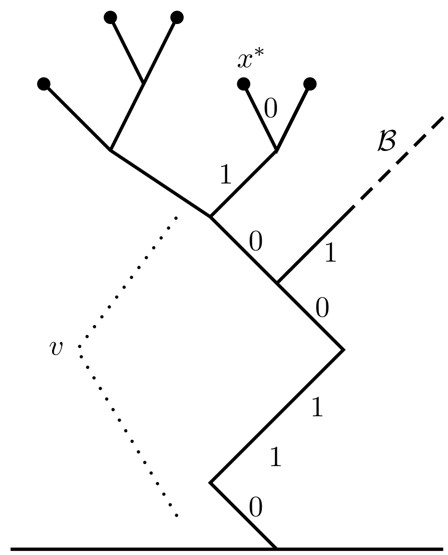

Figure B.3: The above figure shows an example of the domain of left-total U with the terms used in Lemma 23. $x^* = 0110010$ and $v = 01100$. Since v is total and v^- is not, v^- is a prefix of the border sequence $\mathcal{B}$. In the above example, assuming all halting extensions of v produce a unique 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 23 *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

$$\begin{aligned}
\mathbf{d}(x|Q,v) &= \lfloor -\log Q(x) \rfloor - \mathbf{K}(x|v) \\
&\leq \mathbf{K}(x) - \|v\| - \mathbf{K}(x|v) \\
&<^+ (\mathbf{K}(v) + \mathbf{K}(x|v)) - \|v\| - \mathbf{K}(x|v) \\
&<^+ (\|v\| + \mathbf{K}(\|v\|) + \mathbf{K}(x|v)) - \|v\| - \mathbf{K}(x|v), \\
\mathbf{d}(x|Q,v) &<^+ \mathbf{K}(\|v\|). \qquad \text{(B.17)}
\end{aligned}$$

Since v is total and v^- is not total, by proposition (28), v^- is a prefix of the border sequence

$\mathcal{B}$ (see Figure B.3). In addition, Q is computable from v. Therefore

$$\begin{aligned}
\mathbf{K}(x|\mathcal{H}) &<^+ \mathbf{K}(x|Q) + \mathbf{K}(Q|\mathcal{H}) \\
&<^+ \mathbf{K}(x|Q) + \mathbf{K}(v|\mathcal{H}) \\
&<^+ -\log Q(x) + \mathbf{K}(\|v\|) \qquad \text{(B.18)} \\
&<^+ \mathbf{K}(x) - \|v\| + \mathbf{K}(\|v\|), \\
\|v\| &<^+ \mathbf{K}(x) - \mathbf{K}(x|\mathcal{H}) + \mathbf{K}(\|v\|), \\
\|v\| &<^{\log} \mathbf{I}(x;\mathcal{H}). \qquad \text{(B.19)}
\end{aligned}$$

Equation (B.18) is due to Proposition (27). 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 B.17, one gets $\mathbf{Ks}(x) \le \|v\| + O(\mathbf{K}(\|v\|)) <^{\log} \|v\|$. Due to Equation B.19, one gets $\mathbf{Ks}(x) <^{\log} \mathbf{I}(x;\mathcal{H})$.

Theorem 61 *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 $S(z^-)\le S(z)$. 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 (28), $b^-\sqsubseteq\mathcal{B}$ is a prefix of border. Lemma 22, 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 (B.20). Lemma (23), conditional on b, results in Equation (B.21), with

$$\mathbf{K}(a|b) + f(a) <^{\log} s + \mathbf{Ks}(f|b), \qquad \text{(B.20)}$$

$$\mathbf{K}(a|b) + f(a) <^{\log} s + \mathbf{I}(f;\mathcal{H}|b), \qquad \text{(B.21)}$$

$$\mathbf{K}(a|b) + f(a) <^{\log} s + \mathbf{K}(f|b) - \mathbf{K}(f|b,\mathcal{H}). \qquad \text{(B.22)}$$

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 (B.22), we get Equation (B.23). Equation (B.24) is due to the chain rule $\mathbf{K}(b)+\mathbf{K}(f|b) <^{\log} \mathbf{K}(f)+\mathbf{K}(b|f)$. Equation (B.25) follows from the inequality $\mathbf{K}(f|\mathcal{H}) <^+ \mathbf{K}(f|b,\mathcal{H}) + \mathbf{K}(b|\mathcal{H})$.

$$\mathbf{K}(a) + f(a) <^{\log} s + \mathbf{K}(b) + \mathbf{K}(f|b) - \mathbf{K}(f|b,\mathcal{H}), \qquad \text{(B.23)}$$

$$\mathbf{K}(a) + f(a) <^{\log} s + \mathbf{K}(f) + \mathbf{K}(b|f) - \mathbf{K}(f|b,\mathcal{H}), \qquad \text{(B.24)}$$

$$\mathbf{K}(a) + f(a) <^{\log} s + \mathbf{K}(f) + \mathbf{K}(b|f) - \mathbf{K}(f|\mathcal{H}) + \mathbf{K}(b|\mathcal{H}), \qquad \text{(B.25)}$$

$$\mathbf{K}(a) + f(a) <^{\log} s + \mathbf{I}(f;\mathcal{H}) + (\mathbf{K}(b|f) + \mathbf{K}(b|\mathcal{H})). \qquad \text{(B.26)}$$

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 (B.23) being larger than $s+\mathbf{K}(b)$ (up to a logarithmic factor). Since b is a prefix of border, due to proposition (27), one gets that $\mathbf{K}(b|\mathcal{H}) < O(\mathbf{K}(\|b\|)) < O(\log\|b\|) < O(\log\mathbf{K}(b))$. Thus combined with Equation (B.26) and also Equation (B.23), one gets

$$\mathbf{K}(a) + f(a) <^{\log} s + \mathbf{I}(f;\mathcal{H}) + \mathbf{K}(b|f). \qquad \text{(B.27)}$$

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 (B.28) is again due to the right hand side of Equation (B.23).

$$\begin{aligned}
\mathbf{K}(a) + f(a) &<^{\log} s + \mathbf{I}(f;\mathcal{H}) + \mathbf{K}(s, \|b\|), \\
\mathbf{K}(a) + f(a) &<^{\log} s + \mathbf{I}(f;\mathcal{H}), \\
\min_{a \in \mathrm{Dom}(f)} \mathbf{K}(a) + f(a) &<^{\log} -\log \sum_{a \in \mathrm{Dom}(f)} \mathbf{m}(a) 2^{-f(a)} + \mathbf{I}(f;\mathcal{H}).
\end{aligned} \tag{B.28}$$