Programming Research Group

ENTROPY AS A FIXED POINT

Keye Martin

PRG-RR-03-05

Oxford University Computing Laboratory
Wolfson Building, Parks Road, Oxford OX1 3QD

Abstract

We study complexity and information and introduce the idea that while complexity is relative to a given class of processes, information is process independent: Information is complexity relative to the class of all conceivable processes. In essence, the idea is that information is an extension of the concept algorithmic complexity from a class of desirable and concrete processes, such as those represented by binary decision trees, to a class more general that can only in pragmatic terms be regarded as existing in the conception. It is then precisely the fact that information is defined relative to such a large class of processes that it becomes an effective tool for analyzing phenomena in a wide range of disciplines.

We test these ideas on the complexity of classical states. A domain is used to specify the class of processes, and both qualitative and quantitative notions of complexity for classical states emerge. The resulting theory is used to give new proofs of fundamental results from classical information theory, to give a new characterization of entropy in quantum mechanics, to establish a rigorous connection between entanglement transformation and computation, and to derive lower bounds on algorithmic complexity. All of this is a consequence of the setting which gives rise to the *fixed point theorem*: The least fixed point of the copying operator above complexity is information.

Introduction

Is it possible that information is more than merely an object of the intuition? Though information is usually regarded as being *subjective* in nature, can it to a large extent be understood as a consequence of objective things? And if so, what can it teach us about the very things from which it is derived? Can information be used to solve *concrete* problems, can we use it to derive *objective* conclusions about certain types of phenomena in the sciences? These are pragmatic questions, and the process of trying to answer them leads to mutually beneficial connections between classical information theory, domain theory, quantum mechanics, and computer science.

Let us assume that words like 'complexity' and 'information' are just that – words – and begin talking about them as though we knew what they meant. We might say:

- The complexity of a *secret* is the amount of work required to *guess* it.
- The complexity of a *problem* is the amount of work required to *solve* it.
- The complexity of a *rocket* is the amount of work required to *escape gravity.*
- The complexity of a *probabilistic state* is the amount of work required to *resolve* it.

In all cases, there is a task we want to accomplish, and a way of measuring the work done by a process that *actually achieves* the task; such a process belongs to a prespecified class of processes which themselves are the stuff that science is meant to discover, study and understand. Then there are two points not to miss about complexity:

(i) It is relative to a prespecified class of processes,

(ii) The use of the word 'required' necessitates the *minimization* of quantities like work over the class of processes.

Complexity is *process dependent.* Now, what is information in such a setting?

Information, in seeming stark contrast to complexity, is *process independent.* Here is what we mean: *Information is complexity relative to the class of all conceivable processes.* For instance, suppose we wish to measure the complexity of an object x with respect to several different classes $P_1, \ldots, P_n$ of processes. Then the complexity of x varies with the notion of process: It will have complexities $c_1(x), \ldots, c_n(x)$, where c_i is calculated with respect to the class P_i. However, because information is complexity relative to the class of *all* conceivable processes, the information in an object like x will *not* vary. That is what we mean when we say information is process independent: It is an element present in *all* notions of complexity. So we expect

$$\text{complexity} \;\geq\; \text{information}$$

if only in terms of the mathematics implied by the discussion above. For example, this might allow us to *prove* that the amount of work you expect to do in solving a problem always exceeds the a priori uncertainty you have about its solution: The less you know about the solution, the more work you should expect to do. The inequality above is a valuable pragmatic entity. It can be used for instance to derive lower bounds on algorithmic complexity, which definitely qualifies as using information to do something 'real' and 'concrete.'

To test these ideas, we study the complexity of classical states relative to a class of processes. A class of processes will be derived from a domain (D, μ) with a measurement μ that supports a new notion called *orthogonality.* Write $c_D(x)$ for the complexity of a classical state x relative to (D, μ). Then we will see for example that

$$\inf_{D \in \Sigma} c_D = \sigma \tag{1}$$

where σ is Shannon entropy and Σ is the class of domains (D, μ). This equation provides a setting where it is clear that information in the sense of the discussion above is σ, and that the class of all conceivable processes is Σ.

From the point of view of the author trying to write a brief introduction, it would be nice if that were the end of the story, but the truth is that it is less than half of the beginning. Another limit also exists

$$\bigcap_{D \in \Sigma} \leq_D \; = \; \leq \tag{2}$$

where $\leq_D$ is a relation on classical states which means $x \leq_D y$ iff for all processes p on (D, μ), it takes more work for p to resolve x than y. This is *qualitative complexity,* and the value of the intersection above $\leq$ just happens to be a relation called *majorization.* The limits (1) and (2) comprise what we call *the universal limit.*

Now, the universal limit is taken over the class of *all* domains. We only understand its true nature when we discover that $\leq$ and σ can also be arrived at on a *fixed* domain (D, μ) provided that one has the ability to *copy* processes. The mathematics of copying necessitates the addition of algebraic structure $\otimes$ to domains (D, μ) already supporting orthogonality. It is from this setting, which also benefits from basic techniques in the study of codes, the equilibrium state in thermodynamics, and so on, that the *fixed point theorem* springs forth: Just as is the case with recursive programs, the semantics of *information* can also be specified by a least fixed point:

$$\mathrm{fix}(\Phi) = \bigsqcup_{n\geq 0} \Phi^n(\perp) = \sigma$$

where Φ is copying and $\perp$ is the complexity c_D. A number of unexpected consequences emerge along the way:

- Majorization, discovered by Muirhead in 1903 [5], a relation which over the last 100 years has found impressive applications in areas such as economics, computer science, physics and pure mathematics [2][4], is a continuous dcpo on the set of classical states that determine complexity and information.
- The identification of the essential mathematical structure required to execute classical information theory [11] over the class of semantic domains. The two fundamental notions on (D, μ) are

  (a) Orthogonality $\perp$, and
  (b) Tensor $\otimes : D^2 \to D$

  The domain theoretic tensor can be mapped homomorphically onto the tensor of quantum states in such a way that (a) implies orthogonality in Hilbert space.
- Indisputable proof of the pragmatic and theoretical relevance of continuous domains to quantum mechanics, including but by no means restricted to:

  (a) The derivation and characterization of von Neumann entropy,
  (b) Entanglement transformation [10],
  (c) The classification theorem for ensembles [9][13], a domain theoretic result which answers a fundamental question about quantum states originally considered by Schrödinger [12].
- The quantitative notion c_D can be used to derive lower bounds on algorithmic complexity, such as searching and sorting; the qualitative notion $\leq_D$ is used to establish that a well-known form of entanglement transformation in quantum mechanics, called *local operations and classical communication,* can be characterized precisely as an event which, from the perspective of a computer scientist, reduces the average case complexity of all binary decision trees.

The measurement formalism [6] – a theory about information originally formulated in the context of the semantics of computation – extends the applicability of domains at a fundamental level: Fixed point theorems including nonmonotonic functions, an informatic derivative, distance from content, unified approaches to the continuous and discrete, a first order view of recursion $\varphi = \delta + \varphi \circ r$ which models iteration in its natural state. The present work adds to this list the complexity $(c_D, \leq_D)$ as a new technique for the analysis of informatic phenomena.

1 Classical states

We begin with the objects whose complexity we wish to study. These are the classical states.

Definition 1.1 The set of *classical n-states* is

$$\Delta^n := \{x \in [0,1]^n : \sum_{i=1}^{n} x_i = 1\}$$

The set of *monotone decreasing n-states* is

$$\Lambda^n := \{x \in \Delta^n : (\forall i < n)\, x_i \geq x_{i+1}\}$$

for $n \geq 2$.

In 1903, Muirhead [5] discovered an important relation on classical states called *majorization.*

Definition 1.2 For $x, y \in \Lambda^n$, it is

$$x \leq y \equiv (\forall k)\, s^k x \leq s^k y,$$

where

$$s^k x := \sum_{i=1}^{k} x_i$$

for all $k \in \{0, \dots, n\}$. Note that $s^0 x = 0$ for all $x \in \Lambda^n$.

In the last one hundred years, majorization has arisen in a number of contexts, including economics, computer science, physics and mathematics [2] [4]. It is a domain.

Theorem 1.3 *$(\Lambda^n, \leq)$ is a continuous dcpo with least element $\perp = (1/n, \dots, 1/n)$.*

(i) *If (x_i) is an increasing sequence in Λ^n, then*

$$\bigsqcup_{i \geq 1} x_i = \lim_{i \to \infty} x_i$$

where the limit is in the Euclidean topology on Λ^n.

(ii) *For all $t < 1$, $\pi_{\perp x}(t) \ll x$, where $\pi_{\perp x}$ is the straight line path from $\perp$ to x.*

The proof is straightforward and given in the appendix; a separate appendix reviews basic domain theoretic ideas. This order for us will play a dual role both as a device for understanding complexity from a relational viewpoint, and as a general mathematical technique for proving inequalities.

Lemma 1.4 *The order on Λ^n is defined inductively by requiring the maps*

- $\Lambda^n \to [0,1] :: x \mapsto x^+ = x_1$
- $\Lambda^{n+1} \to \Lambda^n :: x \mapsto (x_1 + x_2, \ldots, x_{n+1})$

monotone for all $n \geq 1$.

The second mapping in the above lemma we call a *projection* and denote it by $p : \Lambda^{n+1} \to \Lambda^n$. With the benefit of this inductive principle, we can give a new, direct proof of the following result, which is already known in the literature. We write

$$\langle x|y\rangle := \sum_{i=1}^{n} x_i \cdot y_i$$

for the standard inner product on $\mathbb{R}^n$.

Lemma 1.5 *For $x, y \in \Lambda^n$, we have $x \leq y$ iff for all increasing $a : \{1, \ldots, n\} \to [0, \infty)$, $\langle a|x\rangle \geq \langle a|y\rangle$.*

Proof. The direction ($\Leftarrow$) is simple. For the other, assume the result for $n \geq 2$, and let $a : \{1, \ldots, n+1\} \to [0, \infty)$ be increasing. Define a new *increasing* observable $\varepsilon : \{1, \ldots, n+1\} \to [0, \infty)$ by $\varepsilon_i := a_i - a_1$. Then

$$\langle \varepsilon(2 \ldots n+1)|px\rangle = x_1(a_2 - a_1) + \langle a|x\rangle - a_1,$$

where $\varepsilon(2 \ldots n+1)$ is the vector ε with its first component removed, and $p : \Lambda^{n+1} \to \Lambda^n$ is the natural projection. Since $x \leq y$, we have $px \leq py$ by Lemma 1.4; now the inductive hypothesis kicks in:

$$\langle \varepsilon(2 \ldots n+1)|px\rangle \geq \langle \varepsilon(2 \ldots n+1)|py\rangle.$$

When simplified we get

$$\langle a|x\rangle - \langle a|y\rangle \;\geq\; (y_1 - x_1)(a_2 - a_1) \;\geq\; 0,$$

where the inequality on the right uses that a increases and $x_1 \leq y_1$ ($\Leftarrow x \leq y$). $\square$

2 Processes from the order on a domain

In order to study processes which may result in one of several different outcomes, we have to know what 'different' means. This is what *orthogonality* does: It provides an order theoretic definition of 'distinct.' Let (D, μ) be a continuous dcpo with a measurement μ and least element $\perp$.

Definition 2.1 A pair of elements $x, y \in D$ are *orthogonal* if $\mu(\uparrow\!x \cap \uparrow\!y) \subseteq \{0\}$. This is written $x \perp y$.

Definition 2.2 By a *domain* (D, μ), we will mean a continuous dcpo D whose measurement $\mu \to \sigma_D$ satisfies $\mu\bot = 1$ and

$$\mu(\bigwedge F) \geq \sum_{x \in F} \mu x$$

for each finite set $F \subseteq D$ of pairwise orthogonal elements.

Notice that replacing μ with $\mu/\mu\bot$ if necessary, we can always assume $\mu(\bot) = 1$. Also, the inequality for pairwise orthogonal sets is worth comparing to its "opposite": That $\mu(x \sqcap y) \leq \mu x + \mu y$ if x and y are consistent. The latter is what allows one to derive metrics on $\ker \mu$ [7].

Lemma 2.3 *$(\mathbf{I}[0,1], \mu)$ is a domain.*

The following results give techniques for proving (D, μ) is a domain. They also provide abstract explanations for *why* Kraft's inequality holds (which we will see soon).

Lemma 2.4 *Let $\phi : (D, \mu) \to (E, \mu)$ be a monotone map with $\mu\phi = \mu$ which preserves orthogonality. If (E, μ) is domain, then (D, μ) is also a domain.*

Proof. Given an orthogonal set $F \subseteq D$, $\phi(F)$ is orthogonal in E, so the monotonicity of ϕ and the definition of infimum gives

$$\phi(\bigwedge F) \sqsubseteq \bigwedge \phi(F),$$

and we get

$$\mu(\bigwedge F) = \mu(\phi(\bigwedge F)) \geq \mu(\bigwedge \phi(F)) \geq \sum_{x \in F} \mu(\phi x) = \sum_{x \in F} \mu x$$

where we use the monotonicity of μ into $[0, \infty)^*$, the fact that (E, μ) is a domain, and that $\mu\phi = \mu$. This proves (D, μ) is also a domain in the special sense of this paper. $\square$

Here is one reason ϕ might preserve orthogonality:

Proposition 2.5 *Let $\phi : (D, \mu) \to (E, \mu)$ be an order embedding with $\mu\phi = \mu$ whose image is dense in the Scott topology. If no compact element of D has measure zero, and each $x \in E$ with $\mu x > 0$ has $\Uparrow x \neq \emptyset$, then*

$$x \perp y \Rightarrow \phi x \perp \phi y$$

for all $x, y \in D$. Thus, if (E, μ) is a domain, then so is (D, μ).

Proof. Let $x \perp y$. Suppose ϕx and ϕy are not orthogonal. Then there is $z \in E$ with $\mu z > 0$ and $\phi x, \phi y \sqsubseteq z$. Since $\mu z > 0$, we have by assumption that $\Uparrow z \neq 0$, and since this set is Scott open, there is $a \in D$ with $\phi(a) \in \uparrow\!\!\uparrow z \cap \mathrm{Im}(\phi) \neq \emptyset$. The map ϕ is monotone and has continuous measure $\mu\phi = \mu$ so it is Scott continuous. Then

$$z \ll \phi(a) = \phi(\bigsqcup \mathord{\downarrow\!\!\downarrow} a) = \bigsqcup \phi(\mathord{\downarrow\!\!\downarrow} a),$$

so for some $b \ll a$ we get $\phi x, \phi y \sqsubseteq z \sqsubseteq \phi(b)$, and since ϕ reflects the order,

$$x, y \sqsubseteq b \ll a.$$

But $x \perp y$ so $\mu b = 0$, and since μ is a measurement, $b \in \max(D)$, which means $b = a$, and hence that $a = b \ll a$, i.e., a is a compact member of $\ker \mu$ on D, and that is a contradiction. $\square$

Here is an important example.

Example 2.6 Let $x \in \Delta^n$ be a classical state with all $x_i > 0$ and Σ^∞ the streams over the alphabet $\Sigma = \{1, \ldots, n\}$. Define $\mu : \Sigma^\infty \to [0, \infty)^*$ by $\mu\perp = 1$ and $\mu i = x_i$, and then extend it homomorphically by

$$\mu(s \cdot t) = \mu s \cdot \mu t$$

where the inner dot is concatenation of finite strings. The unique Scott continuous extension, which we call μ, yields a *domain* (D, μ).

We first embed (Σ^∞, μ) into $\mathbf{I}[0, 1]$. Later the algebraic structure of $\mathbf{I}[0, 1]$ will provide a simpler way, but for now, visualize an interval $x \in \mathbf{I}[0, 1]$ as a line segment partitioned into n consecutive line segments having lengths $x_i \cdot \mu x$ for $1 \leq i \leq n$. Let $\phi_i(x)$ be the i^{th} such interval. The map $\phi : \Sigma^\infty \to \mathbf{I}[0, 1]$ is defined by

$$\phi(x) = \begin{cases} \perp & \text{if } x = \perp \\ \phi_i(s) & \text{if } x = s \cdot i \end{cases}$$

Notice that this is a simple extension of the map in [6], where $\phi_0 = $ left and $\phi_1 = $ right.

Having defined ϕ on finite strings, we take its unique Scott continuous extension, and call this ϕ. It is an order embedding whose image is dense in the Scott topology because all $x_i > 0$. Now Prop. 2.5 applies.

An immediate corollary of this example is the case when $x = (1/2, 1/2) \in \Delta^2$, the binary streams with the usual measurement: $(2^\infty, 1/2^{|\cdot|})$ is a domain. This is the basis for the study of binary codes. Another corollary is the vital *Kraft inequality* from classical information theory.

Theorem 2.7 (Kraft) *We can find a finite antichain of Σ^∞ which has finite word lengths $a_1, a_2, \ldots, a_n$ iff*

$$\sum_{i=1}^{n} \frac{1}{|\Sigma|^{a_i}} \leq 1.$$

Finite antichains of finite words are sometimes also called *instantaneous codes.* The inequality in Kraft's result can be derived as follows:

Example 2.8 *The Kraft inequality.* We apply the last example with

$$x = (1/|\Sigma|, \ldots, 1/|\Sigma|) \in \Delta^{|\Sigma|}.$$

A finite subset of $\Sigma^{<\infty}$ is pairwise orthogonal iff it is an antichain. Thus, we may write

$$\mu(\bigwedge F) \geq \sum_{x \in F} \mu x.$$

In particular, $1 = \mu\bot \geq \mu(\bigwedge F)$, using the monotonicity of μ. Notice that the bound we derive on the sum of the measures is more precise than the one given in the Kraft inequality. We call μ the *standard measurement* and assume *it* when writing (Σ^∞, μ), unless otherwise specified.

Finally, the order theoretic structure of a domain (D, μ) gives rise to a notion of *process*: A set of outcomes which are (a) different, and (b) achievable in finite time.

Definition 2.9 A *process* on (D, μ) is a function $p : \{1, \ldots, n\} \to D$ such that $p_i \perp p_j$ for $i \neq j$ and $\mu p > 0$. $P^n(D)$ denotes the set of all such processes.

It is interesting to notice that $\mathbf{I}[0,1]$, like Σ^∞, also satisfies the converse to the Kraft inequality, i.e., the direction we did not prove. This direction permits us to characterize the vectors representable by processes on each of these domains.

Example 2.10 *Processes on binary streams.* The function $-\log \mu : P^n(D) \to (0, \infty)^n$ that takes a process $p \in P^n(D)$ to the vector

$$-\log \mu p = (-\log \mu p_1, \ldots, -\log \mu p_n)$$

produces positive vectors $a = -\log \mu p$ which by the orthogonality of $\mathrm{Im}(p)$ satisfy

$$\sum_{i=1}^{n} \frac{1}{2^{a_i}} \leq 1.$$

In the case of streams, a will also be *integer valued.* However, using the converse to the Kraft inequality, we can say that these vectors are *exactly* the image of $-\log \mu$. That is, any such integer valued vector a can be represented by a process on the domain of binary streams. For $\mathbf{I}[0,1]$ we get all positive vectors obeying the Kraft inequality.

We will now use this notion of process to define the complexity of classical states. Two notions arise: A quantitative measure, called h_D, and a qualitative measure, $\leq_D$, which takes the form of a relation on classical states Λ^n.

3 Complexity (quantitative)

By considering processes on $(2^\infty, \mu)$, it is clear that the expected work done by an algorithm which takes one of n different computational paths $p : \{1, \ldots, n\} \to D$ is $\langle -\log \mu p | x\rangle$. Thus, the complexity of a state $c : \Delta^n \to [0,\infty)^*$ is

$$c(x) := \inf\{\langle -\log \mu p | x\rangle : p \in P^n(D)\}.$$

The function sort^+ reorders the components of a vector so that they increase; its dual sort^- reorders them so that they decrease. We now show that c is essentially a map defined on monotone states in disguise. Here is the crucial result:

Lemma 3.1 *If $a : \{1, \ldots, n\} \to [0,\infty)$ is increasing, then*

$$\langle a|x\rangle \geq \langle a|\mathrm{sort}^-(x)\rangle$$

for all $x \in \Delta^n$.

Proof. By induction, for $x \in \Delta^{n+1}$ with $n \geq 1$, we have

$$\mathrm{sort}^-(px) \leq p(\mathrm{sort}^-(x))$$

where p is the projection from the discussion on majorization $(\Lambda^n, \leq)$.

Assume the claim for $n \geq 1$. To prove it for $n+1$, let $a : \{1, \ldots, n+1\} \to [0,\infty)$ be increasing. Again we define the increasing $\varepsilon : \{1, \ldots, n+1\} \to [0,\infty)$ as before and for ease of exposition set $b := \varepsilon(2 \ldots n+1)$. Recall from the proof of Lemma 1.5 that

$$\langle b|p(x)\rangle = x_1(a_2 - a_1) + \langle a|x\rangle - a_1.$$

By the inductive hypothesis,

$$\langle b|p(x)\rangle \geq \langle b|\mathrm{sort}^-(px)\rangle,$$

and since $\mathrm{sort}^-(px) \leq p(\mathrm{sort}^- x)$,

$$\langle b|\mathrm{sort}^-(px)\rangle \geq \langle b|p(\mathrm{sort}^- x)\rangle$$

using the monotonicity of $\langle b|\cdot\rangle : \Lambda^n \to [0,\infty)^*$ from Lemma 1.5. Putting these two together and rearranging terms gives

$$\langle a|\mathrm{sort}^-(x)\rangle - \langle a|x\rangle \;\leq\; x_1(a_2 - a_1) - (\mathrm{sort}^- x)_1(a_2 - a_1) \;\leq\; 0,$$

done. $\square$

And now the fun starts: The complexity of a classical state does not depend on the order of the probabilities within it.

Proposition 3.2 *For all $x \in \Delta^n$,*

$$c(x) = \inf\{\langle \mathrm{sort}^+(-\log \mu p) | \mathrm{sort}^-(x)\rangle : p \in P^n(D)\}.$$

In particular, the function c is symmetric.

Proof. Throughout this one, we fix $x \in \Delta^n$ and denote the infimum on the right by h. For $c(x) \geq h$, let $p \in P^n(D)$. Then

$$\begin{array}{rcll} \langle -\log \mu p | x\rangle & = & \langle \mathrm{sort}^+(-\log \mu p) | x \cdot \sigma\rangle & \text{(for some permutation } \sigma \in S(n)) \\ & \geq & \langle \mathrm{sort}^+(-\log \mu p) | \mathrm{sort}^-(x)\rangle & \text{(Lemma 3.1)} \\ & \geq & h. & \end{array}$$

Since $c(x)$ is the infimum of such terms, $c(x) \geq h$.

For $h \geq c(x)$, let $p \in P^n(D)$. Let $\sigma \in S(n)$ be a permutation with $x \cdot \sigma = \mathrm{sort}^-(x)$. By rearranging p, there is a process $q \in P^n(D)$ such that $(-\log \mu q) \cdot \sigma = \mathrm{sort}^+(-\log \mu p)$. We have

$$\begin{array}{rcl} \langle \mathrm{sort}^+(-\log \mu p) | \mathrm{sort}^-(x)\rangle & = & \langle \mathrm{sort}^+(-\log \mu p) \cdot \sigma^{-1} | \mathrm{sort}^-(x) \cdot \sigma^{-1}\rangle \\ & = & \langle -\log \mu q | x\rangle \\ & \geq & c(x) \end{array}$$

Since h is the infimum of such terms, $h \geq c(x)$. $\square$

As a consequence, we restrict our attention to monotone decreasing states Λ^n.

Definition 3.3 The *expectation* of a process $p \in P^n(D)$ is $\langle p\rangle : \Lambda^n \to [0, \infty)^*$ given by

$$\langle p\rangle x = \langle \mathrm{sort}^+(-\log \mu p) | x\rangle.$$

If the outcomes of process p are distributed as $x \in \Lambda^n$, then the work we expect p will do when taking one such computational path is $\langle p\rangle x$. And finally, complexity.

Definition 3.4 The *complexity* of a state $h : \Lambda^n \to [0, \infty)^*$ is

$$h(x) = \inf\{\langle p\rangle x : p \in P^n(D)\}.$$

Thus, the relation of h to c is that the diagram

$$\begin{tikzcd} \Lambda^n \arrow[r, "h"] & {[0,\infty)} \\ \Delta^n \arrow[u, "\mathrm{sort}^-"] \arrow[ur, "c"'] & \end{tikzcd}$$

commutes: $(\forall x \in \Delta^n)\ c(x) = h(\mathrm{sort}^-(x))$. The Shannon entropy $\sigma : \Delta^n \to [0, \infty)$ is

$$\sigma x := -\sum_{i=1}^{n} x_i \log x_i$$

We also view it as a map on Λ^n, and as a map on *all* monotone states. Its type will be clear from the context.

Lemma 3.5 *If $a : \{1, \ldots, n\} \to (0, \infty)$ is a vector, there is a unique classical state $y \in \Delta^n$ such that*

$$\langle a|y\rangle - \sigma y = \inf\{\langle a|x\rangle - \sigma x : x \in \Delta^n\}.$$

The state y is given pointwise by $y_i = 2^{-a_i}/Za$ and satisfies

$$\langle a|y\rangle - \sigma y = -\log Za$$

where

$$Za := \sum_{i=1}^{n} \frac{1}{2^{a_i}}.$$

In addition, if a is increasing, then $y \in \Lambda^n$.

Proof. First, arithmetic gives $\langle a|y\rangle - \sigma y = -\log Za$. Next, it is the minimum value of $f(x) = \langle a|x\rangle - \sigma x$ on Δ^n:

$$\begin{aligned}
f(x) &= f(x) + \log Za - \log Za \\
&= -\sum_{x_i > 0} \log\left(\frac{y_i}{x_i}\right) x_i - \log Za \\
&\geq \sum_{x_i > 0} \left(1 - \frac{y_i}{x_i}\right) x_i - \log Za \quad (\text{using } \log x \leq x - 1 \text{ for } x > 0) \\
&= \left(1 - \sum_{x_i > 0} y_i\right) - \log Za \\
&\geq -\log Za
\end{aligned}$$

Finally, y is the *unique state* where f takes it minimum: If $f(x) = -\log Za$, then the string of inequalities above implies

$$-\sum_{x_i > 0} \log\left(\frac{y_i}{x_i}\right) x_i = \sum_{x_i > 0} \left(1 - \frac{y_i}{x_i}\right) x_i$$

which can be rewritten as

$$\sum_{x_i > 0} (t_i - 1 - \log t_i) x_i = 0$$

where $t_i = y_i/x_i$. Because $\log x \leq x - 1$ for $x > 0$, this is a sum of nonnegative terms which results in zero. Then each term must be zero, so $t_i = 1$ which means $x_i = y_i$ whenever $x_i > 0$. However, since $\sum y_i = 1$ and each $y_i > 0$, we must have $x_i > 0$ for *all* $i \in \{1, \ldots, n\}$. Then $x = y$. $\Box$

The last lemma is extremely valuable to us. It asserts the existence and uniqueness of the equilibrium state associated to (energy) observable a from thermodynamics. Notice that we have given a purely elementary proof (no derivatives!)

Proposition 3.6 *If (D, μ) is a domain, then the complexity $h_D : (\Lambda^n, \leq) \to [0, \infty)^*$ is Scott continuous and satisfies $h_D \geq \sigma$ where σ is entropy.*

Proof. First we prove h is Scott continuous. By Lemma 1.5, $\langle p \rangle : \Lambda^n \to [0, \infty)^*$ is monotone when $p \in P^n(D)$, so h is monotone as the sup of such maps. For its continuity, if $h(x) < \varepsilon$, then $\langle p \rangle x < \varepsilon$ for some $p \in P^n(D)$. But $\langle p \rangle \circ \pi_{\perp x} : [0, 1] \to [0, \infty)$ is Euclidean continuous, so

$$(\exists \delta > 0)(\forall t \in (1 - \delta, 1]) \ ((\langle p \rangle \circ \pi_{\perp x})(t) < \varepsilon$$

Thus, $h(a) \leq \langle p \rangle a < \varepsilon$, where $a = \pi_{\perp x}(t)$ for some $t < 1$. But $a \ll x$. By the monotonicity of h, $x \in \twoheaduparrow a \subseteq h^{-1}[0, \varepsilon)$, so h is Scott continuous.

For $h \geq \sigma$, given a process $p \in P^n(D)$, the vector

$$a = \text{sort}^+(-\log \mu p) : \{1, \ldots, n\} \to (0, \infty)$$

satisfies

$$Za = \sum_{i=1}^{n} \mu p_i \leq \mu(\bigwedge \text{Im}(p)) \leq \mu \perp = 1$$

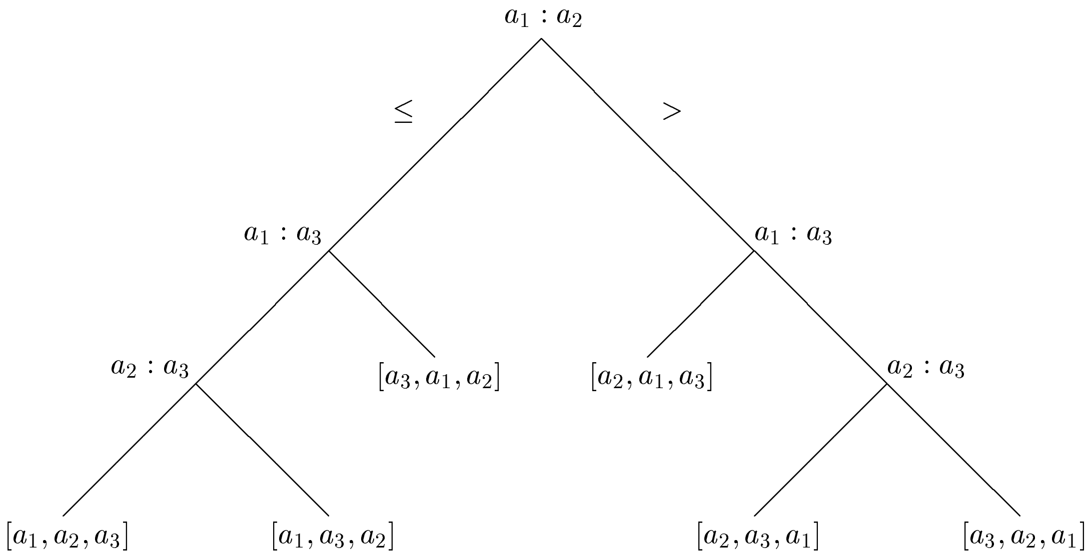
where we appeal to the pairwise orthogonality of $\text{Im}(p)$. Then by Lemma 3.5, using $-\log Z(a) \geq 0$,

$$\langle p \rangle x = \langle a | x \rangle \geq \sigma x,$$

and since $h_D(x)$ is the infimum of such terms, $h_D(x) \geq \sigma x$. Thus, $h_D \geq \sigma$. $\Box$

We have now *proven* the following: The amount of work we expect to do when solving a problem exceeds our a priori uncertainty about the solution. That is, the less you know about the solution, the more work you should expect to do. We mean this *literally* as the following example shows.

Example 3.7 *Lower bounds on algorithmic complexity.* We consider the problem of sorting lists of n objects by comparisons. Any such algorithm for achieving this has a binary decision tree. For example, for lists with three elements a_1, a_2 and a_3, it is



where a move left corresponds to a decision $\leq$, while a move right corresponds to a decision $>$. The leaves of this tree, which are labelled with lists representing potential outcomes of the algorithm, form an antichain of $n!$-many finite words in 2^∞ using the correspondence $\leq \mapsto 0$ and $> \mapsto 1$. This defines a process $p : \{1, \ldots, n!\} \to 2^\infty$. If our knowledge about the answer is $x \in \Lambda^{n!}$, then

$$\begin{aligned} \text{avg. comparisons} &= \langle -\log \mu p | x \rangle \\ &\geq \langle p \rangle (\text{sort}^- x) \\ &\geq h(\text{sort}^- x) \\ &\geq \sigma x. \end{aligned}$$

Assuming complete uncertainty about the answer, $x = \perp$, we get

$$\text{avg. comparisons} \geq \sigma \perp = \log n! \approx n \log n.$$

In addition, we can derive an entirely *objective conclusion.* In the *worst case,* we must do at least

$$\max(-\log \mu p) \geq \langle p \rangle \perp \geq \sigma \perp \approx n \log n$$

comparisons. Thus, sorting by comparisons is in general at least $O(n \log n)$. A similar analysis shows that searching by comparison is at least $O(\log n)$.

Different orders can give rise to different complexity classes, for the simple reason that changing the order changes the notion of process. An example of this is the subdomain $(L, \mu) \subseteq (2^\infty, \mu)$ that models linear search (its complexity is given in Example 6.1).

4 Complexity (qualitative)

Each domain (D, μ), because it implicitly defines a notion of process, provides an intuitive notion of what it means for one classical state to be more complex than another: x is more complex than y iff for all processes $p \in P^n(D)$, the work that p does in resolving x exceeds the work it does in resolving y. This is *qualitative complexity.*

Definition 4.1 For $x, y \in \Lambda^n$, the relation $\leq_D$ is

$$x \leq_D y \equiv (\forall p \in P^n(D)) \, \langle p \rangle x \geq \langle p \rangle y.$$

Only one thing is clear about $\leq_D$: The qualitative analogue of Prop. 3.6.

Lemma 4.2 *For each domain (D, μ), $\leq \, \subseteq \, \leq_D$.*

Proof. Let $x \leq y$. Given a process $p \in P^n(D)$, $\langle p \rangle : \Lambda^n \to [0, \infty)$ is the expectation of the increasing vector $\text{sort}^+(-\log \mu p)$, so by Lemma 1.5 we have $\langle p \rangle x \geq \langle p \rangle y$, which is $x \leq_D y$. $\square$

The calculation of $\leq_D$ requires knowing more about the structure of D. We consider domains whose orders allow for the simultaneous description of *orthogonality* and *composition.* In the simplest of terms: These domains allow us to say what *different* outcomes are, and they allow us to form *composite* outcomes from pairs of outcomes.

Definition 4.3 A domain (D, μ) is *symbolic* when it has an associative operation $\otimes : D^2 \to D$ such that $\mu(x \otimes y) = \mu x \cdot \mu y$ and

$$x \perp u \text{ or } (x = u \;\&\; y \perp v) \Rightarrow x \otimes y \perp u \otimes v$$

for all $x, y, u, v \in D$.

Notice that $\otimes$ has a qualitative axiom and a quantitative axiom.

Example 4.4 The product on $\mathbf{I}[0, 1]$ is

$$[a, b] \otimes [y_1, y_2] = [a + y_1 \cdot (b - a), a + y_2 \cdot (b - a)].$$

It has many neat properties. For example,

$$\perp \otimes x = x \otimes \perp = x,$$

so it is a *monoid* and the measurement μ is now a homomorphism! You can also calculate zeroes of real-valued functions by repeatedly multiplying $\text{left}(\perp) = [0, 1/2]$ and $\text{right}(\perp) = [1/2, 1]$, i.e., the bisection method. Try it, its fun.

We can tensor processes too.

Lemma 4.5 *If $p : \{1, \dots, n\} \to D$ and $q : \{1, \dots, m\} \to D$ are processes, then $p \otimes q : \{1, \dots, nm\} \to D$ is a process.*

If p and q are processes, then $p \otimes q$ is the process whose possible actions are $p_i \otimes q_j$, where p_i is any possible action of p, and q_j is any possible action of q. The exact indices assigned to these composite actions for our purposes is immaterial.

Lemma 4.6 *For $x, y \in \Lambda^n$, we have $x \leq y$ if and only if*

$$\sum_{i=1}^{n} k_i x_i \geq \sum_{i=1}^{n} k_i y_i$$

for every increasing sequence $(k_i)_{i=1}^n$ of positive integers.

Proof. The direction ($\Rightarrow$) is obvious (Lemma 1.5). What makes ($\Leftarrow$) counterintuitive is its use of *positive* integers. We need to prove $s^j x \leq s^j y$ for each $1 \leq j \leq n$. For fixed j, let $k : \{1, \dots, n\} \to \mathbb{N} \setminus \{0\}$ be the vector defined by

$$k_i := \begin{cases} 1 & \text{if } i \leq j; \\ 2 & \text{otherwise.} \end{cases}$$

Then k is an increasing sequence of positive integers so

$$\sum_{i=1}^{j} x_i + 2 \sum_{i=j+1}^{n} x_i \geq \sum_{i=1}^{j} y_i + 2 \sum_{i=j+1}^{n} y_i.$$

This can be written

$$\left(\sum_{i=1}^{j} x_i + \sum_{i=j+1}^{n} x_i \right) + \sum_{i=j+1}^{n} x_i \geq \left(\sum_{i=1}^{j} y_i + \sum_{i=j+1}^{n} y_i \right) + \sum_{i=j+1}^{n} y_i$$

and since the terms grouped in parentheses sum to one, they cancel out leaving

$$\sum_{i=j+1}^{n} x_i \geq \sum_{i=j+1}^{n} y_i$$

which is exactly $s^j x \leq s^j y$. $\square$

Theorem 4.7 *Let $(D, \otimes, \mu)$ be a symbolic domain. If there is a binary process $p : \{1, 2\} \to D$, then the relation $\leq_D = \leq$.*

Proof. By replacing p with the binary process q defined by $q_1 = p_1 \otimes p_2$ and $q_2 := p_2 \otimes p_1$ if necessary, we can assume $\mu p_1 = \mu p_2 > 0$. First we prove $\leq_D \subseteq \leq$.

Suppose we have $x, y \in \Lambda^n$ with $x \leq_D y$. To show $x \leq y$, we use Lemma 4.6. To this end, let k be an increasing vector of positive integers. First we construct n orthogonal elements with the same measure: Let $a : \{1, \ldots, n\} \to D$ be the process defined by restricting $\bigotimes_{i=1}^{n} p$ to $\{1, \ldots, n\}$; we have $\mu a_i = (\mu p_1)^n$ for all $1 \leq i \leq n$.

Now we exponentiate each a_i exactly k_i many times; formally, define a process $b : \{1, \ldots, n\} \to D$ given by

$$b_i := \left(\bigotimes_{1}^{k_i} a_i \right) \otimes \left(\bigotimes_{k_i+1}^{k_n} \perp \right).$$

That is, *each* b_i is the product of k_n different elements: The first $k_i \geq 1$ elements are copies of a_i, the others are copies of $\perp$. It is important to check that b is in fact a process.

First, because $\otimes$ is associative, we may write each b_i in the form

$$b_i := a_i \otimes (\ldots)$$

where crucially this 'picture' requires each $k_i \geq 1$. Now suppose $i \neq j$. Then because $a_i \perp a_j$, the qualitative property of $\otimes$ gives $b_i \perp b_j$. Thus, the image of b is a pairwise orthogonal set. Calculating $\mu b > 0$ benefits from $\mu \perp = 1$:

$$\mu b_i = (\mu a_i)^{k_i} \cdot 1 = (\mu p_1)^{n k_i} > 0.$$

Then b is a process. Since $x \leq_D y$, the definition of $\leq_D$ implies $\langle b \rangle x \geq \langle b \rangle y$.

Recall that $\langle b \rangle x = \langle \mathrm{sort}^+(-\log \mu b) | x \rangle$. But the vector $-\log \mu b$ is already sorted into increasing order: Because the k_i's increase, the vector μb decreases (using $\mu p_1 \leq \mu \perp = 1$). Thus the vector $-\log \mu b$ increases and we get

$$(-\log \mu p_1) n \sum_{i=1}^{n} k_i x_i = \langle b \rangle x \geq \langle b \rangle y = (-\log \mu p_1) n \sum_{i=1}^{n} k_i y_i.$$

Notice that $\mu p_1 < 1$ (or else strict monotonicity of μ gives $p_1 = \perp$ which is impossible since $p_1 \perp p_2$ and $\mu p_2 > 0$). Thus, $-\log \mu p_1 > 0$, so we may divide through by it to get

$$\sum_{i=1}^{n} k_i x_i \geq \sum_{i=1}^{n} k_i y_i.$$

Since this holds for any increasing sequence of positive integers (k_i), Lemma 4.6 implies $x \leq y$. Putting this together with $\leq \, \subseteq \, \leq_D$ from Lemma 4.2 gives $\leq \, = \, \leq_D$. $\square$

5 The universal limit

We now prove that $\leq$ and σ are two sides of the same coin: The former is a qualitative limit; the latter is a quantitative limit. Each is taken over the class of domains.

Theorem 5.1 *Let $\sigma : \Lambda^n \to [0, \infty)^*$ denote Shannon entropy and Σ denote the class of domains. Then*

$$\inf_{D \in \Sigma} h_D = \sigma$$

and

$$\bigcap_{D \in \Sigma} \leq_D = \leq$$

where the relation $\leq$ on Λ^n is majorization.

Proof. The inequality $\inf h_D \geq \sigma$ follows immediately from Prop. 3.6. The proof of the other inequality is all domain theory. Consider $(D, \mu) = (\mathbf{I}[0,1], \mu)$. By Prop. 3.6, its complexity h is Scott continuous. Further, for any state $x \in \Lambda^n$ with $x_i > 0$ for all i,

$$p_i := [s^{i-1}x, s^i x]$$

defines a process $p \in P^n(\mathbf{I}[0,1])$ such that

$$\langle p \rangle x = \sigma x,$$

which means $h(x) = \sigma x$ on a *basis* of the domain $(\Lambda^n, \leq)$. Now given *any* $x \in \Lambda^n$, let $(y_n) \in \Lambda^n$ be the increasing sequence $y_n = \pi_{\perp x}(t - 1/n)$. Each term of this sequence is a *positive* vector. We have

$$h(x) = \bigsqcup_{n \geq 1} h(y_n) = \bigsqcup_{n \geq 1} \sigma(y_n) = \lim_{n \to \infty} \sigma(y_n) = \sigma\left(\lim_{n \to \infty} y_n\right) = \sigma x$$

using only the Scott continuity of h, the Euclidean continuity of σ, the fact that the two maps agree on positive states, and the characterization of suprema in $(\Lambda^n, \leq)$. This proves $h = \sigma$ in all dimensions when $D = \mathbf{I}[0,1]$ which gives $\sigma \geq \inf h_D$.

By Lemma 4.2, the intersection of these relations contains $\leq$. If we take $D = \Sigma^\infty$, the last theorem gives $\leq_D = \leq$, which implies the intersection of the $\leq_D$ contains $\leq$. $\square$

Corollary 5.2 *Shannon entropy $\sigma : (\Lambda^n, \leq) \to [0, \infty)^*$ is Scott continuous.*

Proof. In the last result we saw that entropy is an *example* of complexity, and hence Scott continuous by Prop. 3.6. $\square$

Thus, by Theorem 5.1, the optimum value of $(h_D, \leq_D)$ is $(\sigma, \leq)$. But when does a domain have a value of $(h_D, \leq_D)$ that is close to $(\sigma, \leq)$? Though it is subtle, if we look at the case when $\leq_D$ achieves $\leq$ in the proof of Theorem 4.7, we see that a strongly contributing factor is the ability to *copy* processes – we made use of this idea when we formed the process $\bigotimes_{i=1}^n p$. We will now see that the ability to copy on a given domain *also guarantees* that h is close to σ.

6 Inequalities relating complexity to entropy

We begin with some long overdue examples of complexity. It is convenient on a given domain (D, μ) to denote the complexity in dimension n by $h_n : \Lambda^n \to [0, \infty)$.

Example 6.1 Examples of h.

(i) On the lazy naturals $(L, \mu) \subseteq (2^\infty, \mu)$, where the L is for linear,

$$h_n(x) = x_1 + 2x_2 + \ldots + (n-1)x_{n-1} + (n-1)x_n$$

which is the average number of comparisons required to find an object among n using linear search.

(ii) On the domain of binary streams $(2^\infty, \mu)$,

$$h_2(x) \equiv 1$$

$$h_3(x) = x_1 + 2x_2 + 2x_3 = 2 - x_1$$

$$h_4(x) = \min\{2, x_1 + 2x_2 + 3x_3 + 3x_4\} = \min\{2, 3 - 2x_1 - x_2\}$$

In general, $h_n(x)$ is the average word length of an optimal code for transmitting n symbols distributed according to x.

(iii) On $(\mathbf{I}[0,1], \mu)$,

$$h_n(x) = -\sum_{i=1}^{n} x_i \log x_i,$$

which is the entropy σ.

These examples do little to help us understand the relation of h to σ. What we need is some math. For each integer $k \geq 2$, let

$$c(k) := \inf\{\max(-\log \mu p) : p \in P^k(D)\}.$$

Intuitively, over the class $P^k(D)$ of algorithms with k outputs, $c(k)$ is the worst case complexity of the algorithm whose worst case complexity is *least.*

Theorem 6.2 *Let $(D, \otimes, \mu)$ be a symbolic domain with a process $p \in P^k(D)$. Then*

$$\sigma \leq h \leq \frac{c(k)}{\log k} \cdot (\log k + \sigma)$$

where h and σ can be taken in any dimension.

Proof. Let (Σ^∞, λ) be the domain of streams over an alphabet $\Sigma = \{a_1, \ldots, a_k\}$ with k letters and standard measurement $\lambda a_i = 1/k$. Define a partial map $\phi : \Sigma^\infty \rightharpoonup D$ by $\phi(a_i) = p_i$ and then extend it homomorphically. Let $x \in \Lambda^n$ with all $x_i > 0$. Let m_i be the positive integer such that

$$-\frac{\log x_i}{\log k} \leq m_i < 1 - \frac{\log x_i}{\log k}.$$

Then $(m_i)_{i=1}^n$ is an increasing sequence of positive integers. Because

$$\sum_{i=1}^{n} \frac{1}{k^{m_i}} \leq 1,$$

the converse of the Kraft inequality applied to (Σ^∞, λ) gives a process $q \in P^n(\Sigma^\infty)$ such that $|q_i| = m_i$. Now we use ϕ to map q into a process $r \in P^n(D)$ given by

$$r_i := \phi(q_i) \otimes \left(\bigotimes_{m_i+1}^{m_n} \perp \right)$$

As before, we send each q_i to the tensor of m_n many elements; the first m_i are $\phi(q_i)$, the other $m_n - m_i$ are copies of $\perp$. The fact that r is a process follows from the fact that q is a process and that each r_i is a product of exactly m_n elements. By Lemma 3.1,

$$h(x) \leq \langle r \rangle x \leq \langle \mathrm{sort}^+(-\log \mu r) | x \cdot \sigma \rangle$$

for *any* permutation $\sigma \in S(n)$. Thus, for an appropriate choice of σ, we get

$$h(x) \leq \langle r \rangle x \leq -\sum_{i=1}^{n} x_i \cdot \log \mu r_i$$

and since each μr_i is a product of m_i many components of the vector μp,

$$\mu r_i \geq \min(\mu p)^{m_i}$$

so

$$hx \leq \langle r \rangle x \leq (-\log \min(\mu p)) \cdot \left(1 + \frac{\sigma x}{\log k}\right)$$

and hence

$$h \leq \frac{c(k)}{\log k}(\log k + \sigma)$$

on a *basis* of the domain $(\Lambda^n, \leq)$. But both of these functions are Scott continuous, so the inequality holds on *all* of Λ^n. $\square$

Since any symbolic domain with a process must have a binary process:

Corollary 6.3 *Let $(D, \otimes, \mu)$ be a symbolic domain with a binary process $p \in P^2(D)$. Then $\sigma \leq h \leq c \cdot (1+\sigma)$ where the constant $c \geq 1$ depends only on (D, μ).*

Thus, the mere existence of a process on a *symbolic* domain (D, μ) means not only that $\leq_D = \leq$ but also that h and σ are of the same order. Without the ability to copy, h and σ can be very different: Searching costs $O(n)$ on L, so h_L and σ are not of the same order. However, the distinction 'of the same order' is not precise enough for what we have in mind.

Definition 6.4 If $(D, \otimes, \mu)$ is a symbolic domain, then assuming it exists, the integer

$$\inf\{k \geq 2 : c(k) = \log k\}$$

is called the *algebraic index* of (D, μ) when it exists.

Notice that we always have $c(k) \geq \log k$ by orthogonality, so to calculate the algebraic index we need only prove $c(k) \leq \log k$.

Corollary 6.5 *If $(D, \otimes, \mu)$ is a symbolic domain with algebraic index $k \geq 2$, then*

$$\sigma \leq h \leq \log k + \sigma$$

where h and σ can be taken in any dimension.

For instance, the algebraic index of $\mathbf{I}[0,1]$ is 2. But *why* should such an integer exist? To take a closer look, recall that a process p for solving a problem always does an amount of work which exceeds our ignorance about its solution: $\langle p \rangle \geq \sigma$. Thus, the amount of work not due to our a priori uncertainty about the solution is $\langle p \rangle - \sigma$. When is this amount *minimized?* The answer is the equilibrium state $y \in \Lambda^n$ from thermodynamics: The *unique* state y such that

$$\langle p \rangle y - \sigma y = \inf\{\langle p \rangle x - \sigma x : x \in \Lambda^n\},$$

which is found to be

$$y := \frac{\mathrm{sort}^-(\mu p)}{Z(-\log \mu p)}$$

by applying Lemma 3.5 to the positive increasing vector $-\log \mu p$. Thus, the answer to our question implicitly defines a *function* $g : P^n(D) \to \Lambda^n$ for each n that is called the *Gibbs map*.

Definition 6.6 The map $g : P^n(D) \to \Lambda^n$ is defined by

$$g(p) := \frac{\mathrm{sort}^-(\mu p)}{Z(p)}$$

where $Z(p) := \sum_{i=1}^n \mu p_i$ is called the *partition function.* We call $g(p)$ the *Gibbs state* associated with process p.

Equilibrium can be expressed as follows: Given $w(\cdot) = \langle\cdot\rangle + \log Z(\cdot)$ and the entropy $\sigma : \Lambda \to [0,\infty)$, there is a unique $g : P(D) \to \Lambda$ which makes

$$\begin{tikzcd}
P(D) \arrow[r, "\Delta"] \arrow[d, "g"'] & P(D) \times P(D) \arrow[r, "w \times g"] & {[\Lambda \to [0,\infty)] \times \Lambda} \arrow[d, "ev"] \\
\Lambda \arrow[rr, "\sigma"'] & & {[0,\infty)}
\end{tikzcd}$$

commute. It is the Gibbs map.

Definition 6.7 The *optimal processes* on (D, μ) are

$$S^n(D) := \{p \in P^n(D) : Zp = 1\}.$$

Notice that any $p \in P^n(D)$ has $Z(p) \leq 1$, by the pairwise orthogonality of $\mathrm{Im}(p)$. For $p \in S^n(D)$, we have $\bigwedge \mathrm{Im}(p) = \perp$. Another sense in which p is optimal is that it resolves the state $x = g(p)$ faster than any other process q:

$$\langle q\rangle x \geq hx \geq \sigma x = \langle p\rangle x$$

It seems to the author that optimal codes (from classical information theory) also yield optimal processes in the sense defined above.

Mathematically, the Gibbs map gives us a natural way to formalize the idea that a classical state can be represented by a process. For instance, suppose that a basis of $(\Lambda^n, \leq)$ can be represented by a subset of $S^n(D)$, as is the case with the interval domain.

Proposition 6.8 *Let (D, μ) be any domain. If the image of the Gibbs map g restricted to $S^n(D)$ is a basis for the domain Λ^n, then $h = \sigma$ in dimension n.*

Proof. For each $p \in S^n(D)$,

$$\langle p\rangle g(p) - \sigma g(p) = -\log Z(p) = 0$$

which means that $h = \sigma$ on the image of $g|_{S^n(D)}$. But this image is a basis for the domain Λ^n and both maps are Scott continuous; hence $h = \sigma$. $\Box$

Continuing, the above result has a generalization which is subtle but significant for the work at hand: If the image of the Gibbs map g is a basis for Λ^n, then it must contain $\perp \in \Lambda^n$. The reason is that any basis for a domain must contain the least element. Even with this far weaker hypothesis, we can prove the following:

Proposition 6.9 *Let $(D, \otimes, \mu)$ be a symbolic domain. If the image of the Gibbs map $g : S^k(D) \to \Lambda^k$ contains $\perp \in \Lambda^k$ for some k, then*

$$\sigma \leq h \leq \log k + \sigma$$

where h and σ can be taken in any dimension.

Proof. If $g(p) = \perp$ with $p \in S^k(D)$, then $\mu p = \perp$, which means $c(k) = \log k$. Then the algebraic index m exists and we get

$$\sigma \leq h \leq \log m + \sigma \leq \log k + \sigma$$

since m is the least integer for which $c(m) = \log m$. $\square$

This result identifies a simple and concrete way to calculate the algebraic index of a symbolic domain: Find the least dimension where $\perp$ can be represented by an optimal process on D. Notice too that if the algebraic index of D is k, then D basically contains an algebraic copy of the free monoid on k letters. This copy can be made order theoretic by imposing $x \sqsubseteq x \otimes y$.

7 The fixed point theorem

Let Λ be the set of all monotone decreasing states and $P(D)$ be the set of all processes on (D, μ). If we now regard the Gibbs map as a function $g : P(D) \to \Lambda$ we notice that

$$g(p \otimes q) = g(p) \otimes g(q) \quad \text{and} \quad Z(p \otimes q) = Zp \cdot Zq,$$

where $\otimes : \Lambda \times \Lambda \to \Lambda$ is defined by

$$x \otimes y := \text{sort}^-(x_1 y, \ldots, x_n y)$$

That is, given $x \in \Lambda^n$ and $y \in \Lambda^m$, we multiply any x_i by any y_j and use these nm different products to build a vector in Λ^{nm}. Thus, the product $\otimes$ on $P(D)$ *causes* a tensor $\otimes$ on Λ.

Definition 7.1 Let X be a set with a tensor $\otimes$. The copying operator $! : X \to X$ is

$$!x := x \otimes x$$

for all $x \in X$.

What is copying?

Example 7.2 *Ensembles of molecules.* As is well-known, the states of a molecule are represented by a vector $p : \{1, \ldots, n\} \to D$, usually using letters in an alphabet. Thus, p is a process in the formal sense. Each possible state p_i of the molecule has a probability x_i associated to it. Thus, to $p \in P^n(D)$ there is $x \in \Lambda^n$.

Now consider two identical molecules as a single system. This defines a new process $!p := p \otimes p$ whose possible states are $p_i \otimes p_j$. In addition, we usually assume the molecules interact lightly, which means the distribution for the state of the joint system is $!x := x \otimes x$ (independence). If we consider several such molecules, say n, then

$$\bigotimes_{i=1}^{n} p$$

describes the state of a *gas*. Its distribution is x^n.

If after this example it seems that $x^2 = 2x$, it's true:

Example 7.3 Let $f, g : \Lambda \to \Lambda$ be functions with informatic derivatives at $p \in \Lambda$. Then

$$d(f \otimes g)_\sigma(p) = df_\sigma(p) + dg_\sigma(p).$$

In particular, for $f = g = \mathrm{id}_\Lambda$, we get $d(!)_\sigma = 2$.

Let's take a different viewpoint now. Suppose we want to solve a problem, like sorting or searching a list, and for this we have a process $p : \{1, \ldots, n\} \to D$. No matter what our a priori knowledge $x \in \Lambda^n$ is about the solution, in the worst case $x = \perp$, we will have to do an amount of work $\langle p \rangle x \geq h(x)$ where h is the complexity from (D, μ). But suppose we try to solve our problem as follows:

- Copy the process p to obtain a new process $!p$, this results in two uninteracting copies of p.
- When p wants to solve a problem, we solve it "indirectly" by executing $!p$. This results in an outcome $p_i \otimes p_j$.
- We extract the answer to p from the outcome provided by $!p$.

It is not at all clear that such ideas make computational sense. We need an example.

Example 7.4 Consider the binary process on $\mathbf{I}[0, 1]$,

$$p = ([0, 1/2], [1/2, 1])$$

It arises when we want to compute a zero of a real valued function $f : [0, 1] \to \mathbb{R}$ using the bisection method. Now consider

$$!p := ([0, 1/4], [1/4, 1/2], [1/2, 3/4], [3/4, 1]).$$

To solve the problem p, we form $!p$, execute it, and then using its outcome $p_i \otimes p_j$, extract the solution to p. Notice that $!p$ produces an outcome by executing p twice.

The next step is most important. You will notice that various costs arise when we want to solve p using $!p$. For example, it takes work to extract the answer to p from $!p$. These costs, and any others you can imagine, we want to *ignore.* This is entirely *unrealistic* and that is the point: Under wholeheartedly ridiculous assumptions about how the world works, we will see that entropy is *still* a lower bound on the amount of work required to solve a problem. This is another illustration of the idea we mentioned in the beginning: 'information' is complexity relative to the class of all conceivable processes.

In general now, if our a priori knowledge about the result of p is x, then our a priori knowledge about the result of $!p$ is $!x$. Thus, the fastest that $!p$ can execute is $h(!x)$. Since $!p$ produces two outcomes of p, the fastest that $!p$ can indirectly solve the problem p is about $h(!x)/2$. If we now repeat this copying process *forever,* successively making use of processes p^{2^n}, we find that the amount of work $h(x^{2^n})/2^n$ required to solve the original problem of p *always* exceeds σx. And in the limit,

$$\lim_{n\to\infty} \frac{h(x^{2^n})}{2^n} = \sigma x.$$

Thus, even assuming this unrealistic appeal to the conception, we *still* have to do an amount of work no less than σx: Another example of the process independent nature of information.

Of course, this is nothing more than a computational interpretation of Shannon's idea from coding theory. The advantage of this new interpretation, though, is two fold: First, the author understands this one; and second, one realizes that *entropy is an ideal object.* The situation between complexity and entropy is no different than the way we calculate $\sqrt{2}$: Beginning with a rational initial value x_0, we successively iterate a map I_f which produces a sequence $I_f^n(x_0)$ of constructively generated rationals with $\sqrt{2} = \lim_{n\to\infty} I_f^n(x_0)$. In our case, the role of the initial guess x_0 is played by the complexity h, the limit $\sqrt{2}$ is replaced with entropy σ, and the analogue of the Newton iterate I_f is the copying operator Φ given in the following theorem.

Theorem 7.5 *Let $(D, \otimes, \mu)$ be a symbolic domain whose algebraic index is $k \geq 2$. Then the least fixed point of the Scott continuous operator*

$$\Phi : [\Lambda \to [0,\infty)^*] \to [\Lambda \to [0,\infty)^*]$$

$$\Phi(f) = \frac{f!}{2}$$

on the set $\uparrow (h + \log k)$ *is*

$$\mathrm{fix}(\Phi) = \bigsqcup_{n\geq 0} \Phi^n(h + \log k) = \sigma,$$

where $h : \Lambda \to [0,\infty)$ is the complexity on all states.

Proof. Let $k \geq 2$ satisfy $\sigma \leq h \leq \log k + \sigma$. First, we prove that $h + \log k \sqsubseteq \Phi(h + \log k)$. Note that

$$\sigma \leq \frac{h!}{2} \leq \frac{\log k}{2} + \sigma$$

using $\sigma = \sigma!/2$. Then

$$\left(h + \frac{\log k}{2}\right) - \left(\frac{h!}{2}\right) \;\geq\; \left(\sigma + \frac{\log k}{2}\right) - \left(\frac{\log k}{2} + \sigma\right) \;=\; 0$$

and this is exactly the statement that $h + \log k \sqsubseteq \Phi(h + \log k)$. Thus, the least fixed point of Φ on $\uparrow (h + \log k)$ is

$$\mathrm{fix}(\Phi) = \bigsqcup_{n\geq 0} \Phi^n(h + \log k).$$

Since $\sigma = \Phi(\sigma) \in \uparrow(h + \log k)$, we must have $\mathrm{fix}(\Phi) \sqsubseteq \sigma$. However, we also have

$$\sigma + \frac{\log k}{2^n} \sqsubseteq \Phi^n(h + \log k) \sqsubseteq \mathrm{fix}(\Phi)$$

and since this holds for all n, we get $\sigma \sqsubseteq \mathrm{fix}(\Phi)$. This proves $\mathrm{fix}(\Phi) = \sigma$. $\square$

There are a few interesting points to be made about this result. First, Φ has many fixed points above σ: Consider $c \cdot \sigma$ for $c < 1$. Thus, Φ cannot be a contraction on any subset containing $\uparrow h$. But Φ also has fixed points *below* σ: The map $f(x) = \log \dim(x) = \sigma\bot_{\dim(x)}$ is one such example. This iterative process is very sensitive to where one begins, so much so that a far reaching truth must be near: Information is the least fixed point of the copying operator that is above complexity.

Example 7.6 *Shannon's first theorem.* This follows from the fixed point theorem. Let $D = \Sigma^\infty$ be the domain of streams over the alphabet $\Sigma = \{0, \ldots, k-1\}$ whose natural measurement is $\mu s = 1/k^{|s|}$. Then the algebraic index of D is k so

$$\bigsqcup_{n \geq 0} \Phi^n(h + \log k) = \sigma.$$

Now, it is a general domain theoretic fact, though admittedly obscure, that *because σ is the least fixed point of Φ on $\uparrow(h + \log k)$* we must have

$$\bigsqcup_{n \geq 0} \Phi^n(f) = \mathrm{fix}(\Phi)$$

for any f with $h + \log k \sqsubseteq f \sqsubseteq \mathrm{fix}(\Phi)$, even though the sequence $(\Phi^n(f))$ need not be increasing. This is the attractive nature of least fixed points discussed in [8]. If we take f to be the complexity h, we get

$$\bigsqcup_{n \geq 0} \Phi^n(h) = \sigma.$$

Recalling that in the case of $(D, \mu) = (\Sigma^\infty, \mu)$, $h(x)$ is the minimum achievable average word length for transmitting n symbols distributed as x, we have Shannon's first noiseless coding theorem. It is true *for any* symbolic domain with algebraic index k.

In the proof of the fixed point theorem, we can regard Λ a continuous dcpo by viewing it as a disjoint union of domains. But we don't have to, we could just view it as a set. And if we do, the function space is still a dcpo, the theorem remains valid, and we obtain a new characterization of entropy:

Corollary 7.7 *Let $(D, \otimes, \mu)$ be a symbolic domain with algebraic index $k \geq 2$. Then there is a greatest function $f : \Lambda \to [0, \infty)$ which satisfies $h \geq f$ and $f(x \otimes x) \geq f(x) + f(x)$. It is Shannon entropy.*

The question then, "Does h approximate σ, or is it σ which approximates h" is capable of providing one with hours of entertainment.

8 Entropy in quantum mechanics

Let $\mathcal{H}$ be a one dimensional complex Hilbert space with inner product $\langle\cdot|\cdot\rangle$. The inner product in quantum mechanics is taken as conjugate linear in its *first argument.* The tendency in mathematical settings is conjugate linearity in its *second argument.* This is worth keeping in mind; otherwise, standard results from functional analysis have a way of appearing slightly off.

The *Cartesian product* (or direct sum) of n copies of $\mathcal{H}$ is written $\mathcal{H}^n$. It will model the state space of an n dimensional quantum system as follows. Each unit vector $|\psi\rangle \in \mathcal{H}^n$ defines a *pure state*

$$|\psi\rangle\langle\psi| : \mathcal{H}^n \to \mathcal{H}^n$$

which is the operator that takes vector $|\phi\rangle$ to a scalar multiple of $|\psi\rangle$, given by $|\psi\rangle \cdot \langle\psi|\phi\rangle$. Two normalized states $|\psi\rangle$ and $|\phi\rangle$ describe the same pure state iff $|\psi\rangle = e^{i\theta}|\phi\rangle$ for some $\theta \in \mathbb{R}$, i.e., they are equal to within a phase factor.

To *know* the state of a quantum system is to be in possession of a pure state describing it. Otherwise, when we don't know the state of the system exactly, our knowledge is represented by a *mixed state* ρ: A linear, self-adjoint, positive $\rho : \mathcal{H}^n \to \mathcal{H}^n$ with $\mathrm{tr}(\rho) = 1$ called a *density operator.* Like any self-adjoint operator, a density operator has a spectral decomposition into a sum of pure states

$$\rho = \sum_{i=1}^{n} \lambda_i |\psi_i\rangle\langle\psi_i|$$

where the unit vectors $\{|\psi_i\rangle\}$ are pairwise orthogonal, and the λ_i are *real numbers.* However, by positivity, each $\lambda_i \geq 0$, while $\mathrm{tr}(\rho) = 1$ gives

$$\sum_{i=1}^{n} \lambda_i = 1.$$

Thus, a natural map taking density operators to monotone decreasing classical states exists, called the *spectrum.*

Definition 8.1 Let Ω^n be the set of density operators on $\mathcal{H}^n$. Define

$$\mathrm{spec} : \Omega^n \to \Lambda^n$$

$$\mathrm{spec}(\rho) = \mathrm{sort}^-(\lambda_1, \ldots, \lambda_n)$$

using the eigenvalues of ρ. These maps give rise to

$$\mathrm{spec} : \Omega \to \Lambda$$

in the natural way, where $\Omega := \bigcup_{n\geq 2} \Omega^n$.

Now we want to define complexity for quantum states. For this, we have to think about what a quantum state is made of.

Definition 8.2 An *n-ensemble* for $\rho \in \Omega^n$ is a pair $(x, |\cdot\rangle)$ consisting of a classical state $x \in \Lambda^n$ and a list $|\cdot\rangle : \{1, \ldots, n\} \to \mathcal{H}^n$ such that $|i\rangle$ is a unit vector for all $1 \leq i \leq n$ and

$$\rho = \sum_{i=1}^{n} x_i \cdot |i\rangle\langle i|$$

To illustrate, if we take $x = \mathrm{spec}(\rho)$ and $|i\rangle := |\psi_i\rangle$ in a spectral decomposition of ρ, we get an n-ensemble which gives rise to ρ. The vectors in this particular ensemble are orthogonal, but in general, the vectors in an ensemble need not be.

So remembering that by complexity of a state we mean the amount of work *required* to resolve it, the *complexity* of a quantum state ρ is

$$h(\rho) = \inf\{h(x) : (x, |\cdot\rangle) \text{ is an } n\text{-ensemble for } \rho\}.$$

Now we come to a special case of a remarkable result: *The classification theorem for ensembles.* The direction ($\Rightarrow$) is due to Uhlmann [13] while ($\Leftarrow$) is due to Nielsen [9].

Theorem 8.3 *For all $x \in \Lambda^n$, we can find normalized vectors $\{|\psi_i\rangle\}_{i=1}^n$ such that*

$$\rho = \sum_{i=1}^{n} x_i \cdot |\psi_i\rangle\langle\psi_i|$$

if and only if

$$x \leq \mathrm{spec}(\rho)$$

where the continuous dcpo $(\Lambda^n, \leq)$ is majorization.

For domain theorists: In passing from density operators to the classical fragment of their generating ensembles we obtain exactly the *irreducible* Scott closed subsets of $(\Lambda^n, \leq)$. Continuing with complexity now, the Scott continuity of h on Λ^n gives

$$\begin{aligned} h(\rho) &= \inf\{h(x) : (x, |\cdot\rangle) \text{ is an } n\text{-ensemble for } \rho\} \\ &= \inf\{h(x) : x \in \downarrow\mathrm{spec}(\rho)\} \\ &= h\left(\bigsqcup \downarrow\mathrm{spec}(\rho)\right) \\ &= h(\mathrm{spec}(\rho)) \end{aligned}$$

We have thus derived the following

Definition 8.4 The *complexity* of a quantum state $h : \Omega \to [0, \infty)$ is the map that causes

$$\begin{tikzcd} \Lambda \arrow[r, "h"] & {[0,\infty)} \\ \Omega \arrow[u, "\mathrm{spec}"] \arrow[ur, "h"'] & \end{tikzcd}$$

to commute.

In the classical setting we had an operation $\otimes$ on processes and states; now we consider $\otimes$ in the quantum setting. The *tensor product* of $\mathcal{H}^n$ and $\mathcal{H}^m$ is the Hilbert space

$$\mathcal{H}^n \otimes \mathcal{H}^m := \mathcal{H}^{nm}.$$

In the appendix we give a clear account of 'what' and 'why.' If $\{\psi_i\}_{i=1}^n$ is an orthonormal basis for $\mathcal{H}^n$ and $\{\phi_j\}_{j=1}^m$ is an orthonormal basis for $\mathcal{H}^m$, then $\{\psi_i \otimes \phi_j\}$ is an orthonormal basis for $\mathcal{H}^n \otimes \mathcal{H}^m$, so any $\psi \in \mathcal{H}^n \otimes \mathcal{H}^m$ is a linear combination

$$\psi := \sum_{i,j} c_{ij}(\psi_i \otimes \phi_j)$$

where the c_{ij} are complex. If $\rho : \mathcal{H}^n \to \mathcal{H}^n$ is the density operator for a system called 1, and $\sigma : \mathcal{H}^m \to \mathcal{H}^m$ is the density operator for a system called 2, then the density operator for these systems when considered together as forming a new single system is

$$\rho \otimes \sigma : \mathcal{H}^n \otimes \mathcal{H}^m \to \mathcal{H}^n \otimes \mathcal{H}^m$$

which is defined by

$$(\rho \otimes \sigma)(x \otimes y) = \rho x \otimes \sigma y$$

and then extended linearly.

Lemma 8.5 *For $\rho, \sigma \in \Omega$, we have* $\mathrm{spec}(\rho \otimes \sigma) = \mathrm{spec}(\rho) \otimes \mathrm{spec}(\sigma)$.

Proof. Using the spectral decompositions of ρ and σ, we can write the operator $\rho \otimes \sigma$ as

$$\begin{aligned} \rho \otimes \sigma &= \left(\sum_{i=1}^n \lambda_i |\psi_i\rangle\langle\psi_i|\right) \otimes \left(\sum_{j=1}^m \mu_j |\phi_j\rangle\langle\phi_j|\right) \\ &= \sum_{i=1}^n \sum_{j=1}^m \lambda_i \mu_j \left(|\psi_i\rangle\langle\psi_i| \otimes |\phi_j\rangle\langle\phi_j|\right) \end{aligned}$$

using bilinearity. Now we plug in $|\psi_i\rangle \otimes |\phi_j\rangle$ and get

$$(\rho \otimes \sigma)(|\psi_i\rangle \otimes |\phi_j\rangle) = \lambda_i \mu_j \cdot (|\psi_i\rangle \otimes |\phi_j\rangle)$$

using the pairwise orthogonality of the sets $\{|\psi_i\rangle\}$ and $\{|\phi_j\rangle\}$. Since $\{|\psi_i\rangle \otimes |\phi_j\rangle\}$ is an orthonormal basis for $\mathcal{H}^n \otimes \mathcal{H}^m$, we produce all nm eigenvalues of $\rho \otimes \sigma$ in this manner, which must be the set $\{\lambda_i \mu_j\}$. $\square$

To copy $\rho \in \Omega$ means to consider two identically prepared but physically distinct copies of the same system:

$$!\rho := \rho \otimes \rho.$$

In programming terms: On machine A, we write a program p; on machine B, physically different from A, we write program p. The two considered together yield a process that behaves like $!p$. The commutativity of

$$\begin{array}{ccc} \Omega & \xrightarrow{\text{spec}} & \Lambda \\ \Big\downarrow{!} & & \Big\downarrow{!} \\ \Omega & \xrightarrow[\text{spec}]{} & \Lambda \end{array}$$

implied by $\operatorname{spec}(\rho \otimes \sigma) = \operatorname{spec}(\rho) \otimes \operatorname{spec}(\sigma)$ makes proving the next result simple.

Proposition 8.6 *Let $(D, \otimes, \mu)$ be a symbolic domain with algebraic index $k \geq 2$. Then the least fixed point of the Scott continuous operator*

$$\Phi : [\Omega \to [0, \infty)^*] \to [\Omega \to [0, \infty)^*]$$

$$\Phi(f) = \frac{f!}{2}$$

on the set $\uparrow(h + \log k)$ is

$$\operatorname{fix}(\Phi) = \bigsqcup_{n \geq 0} \Phi^n(h + \log k) = \sigma$$

where $\sigma\rho = -\operatorname{tr}(\rho \log \rho)$ is the von Neumann entropy.

Notice that if one accepts the definition of complexity $h(\rho)$ in terms of ensembles, then exactly two steps allow us to derive von Neumann entropy: (1) The classification theorem for ensembles, and (2) The calculation of a least fixed point. Each step is domain theoretic.

Corollary 8.7 *Let $(D, \otimes, \mu)$ be a symbolic domain with algebraic index $k \geq 2$. Then there is a greatest function $f : \Omega \to [0, \infty)$ which satisfies $h \geq f$ and $f(\rho \otimes \rho) \geq f(\rho) + f(\rho)$. It is von Neumann entropy.*

What we want to look closely at now is the *meaning* of this iteration, and we can do so using the ideas from the classical setting. What makes this possible is the connection between orthogonality in the domain theoretic sense, and orthogonality in Hilbert space. For this, we construct a map $\bar{f} : P(D) \to \Omega$ whose canonical nature will be clear.

Let $\Sigma = \{1, \ldots, n\}$ be an alphabet of n symbols, $\lambda \in \Lambda^n$ and $(D, \mu) = (\Sigma^\infty, \mu)$ the domain of streams with $\mu i = \lambda_i > 0$ extended homomorphically. Let $\Sigma \to \Omega$ be an injective map onto an orthonormal basis of $\mathcal{H}^n$. We extend it to $f : \Sigma_\perp \to \mathcal{H}^n$ by choosing for $\perp$ any other unit vector in $\mathcal{H}^n$. For a process $p \in P(D)$,

- The *dimension* of p is

$$\dim(p) := |\mathrm{dom}(p)|$$

  the cardinality of its domain.

- The composition $|\cdot| \circ p$, where $|\cdot|$ is the length function on strings, is a vector of positive integers we denote by $|p|$. The largest integer in $|p|$ is

$$|p|^+ := \max_{1 \leq i \leq \dim(p)} |p_i|$$

  while the smallest integer in this vector is

$$|p|^- := \min_{1 \leq i \leq \dim(p)} |p_i|.$$

- We define

$$d(p) := n^{|p|^+} = |\Sigma|^{|p|^+}$$

  which will be the dimension of the space that operator $\bar{f}(p)$ is defined on.

Here is an important and revealing relationship between these quantities.

Lemma 8.8 *For $p \in P(D)$, we always have $d(p) \geq \dim(p)$.*

Proof. This is *not* a property of vectors of strings; it is a characteristic of *processes.* The usual measurement λ on Σ^∞, given by $\lambda i = 1/|\Sigma|$, yields the same notion of process as μ because $\ker \mu = \ker \lambda$. But using λ and the orthogonality of $\mathrm{Im}(p)$ we get

$$\sum_{i=1}^{\dim(p)} \frac{1}{|\Sigma|^{|p|^+}} \leq \sum_{i=1}^{\dim(p)} \frac{1}{|\Sigma|^{|p_i|}} \leq 1$$

which then gives $\dim(p) \leq |\Sigma|^{|p|^+}$. $\square$

To define $\bar{f}$, we use the *identical* technique used earlier to map codes into domains in our study of $\leq_D$ and h_D. Given $p \in P(D)$, to each p_i, we associate a unit vector $|\psi_i\rangle \in \mathcal{H}^{d(p)}$ given by

$$|\psi_i\rangle := \left(\bigotimes_{j=1}^{|p_i|} f(p_i(j)) \right) \otimes \left(\bigotimes_{j=|p_i|+1}^{|p|^+} f(\perp) \right)$$

and then set

$$\bar{f}(p) := \sum_{i=1}^{\dim(p)} \frac{\mu p_i}{Z(p)} \cdot |\psi_i\rangle\langle\psi_i|.$$

Finally, let $g^* : P(D) \to \Lambda$ be the map which treats $g(p) \in \Lambda^{\dim(p)}$ as a state in $\Lambda^{d(p)}$. Intuitively, $g^*(p)$ is $g(p)$ with $(d(p) - \dim(p))$ many zeroes adjoined.

Theorem 8.9 *The diagram*

$$\begin{tikzcd}
\Omega \arrow[r, "\mathrm{spec}"] & \Lambda \\
P(D) \arrow[u, "\bar{f}"] \arrow[ur, "g^*"'] &
\end{tikzcd}$$

commutes and $\bar{f}(p \otimes q) = \bar{f}(p) \otimes \bar{f}(q)$ *whenever* $|p|$ *is a constant vector.*

Proof. First, each $|\psi_i\rangle$ is the tensor of $|p|^+$ different unit vectors in $\mathcal{H}^n$. Then $|\psi_i\rangle$ is a vector in $\mathcal{H}^d$, where $d := d(p)$. Its norm is

$$|\,|\psi_i\rangle\,| = \left(\prod_{j=1}^{|p_i|} |f(p_i(j))|\right) \cdot \left(\prod_{j=|p_i|+1}^{|p|^+} |f(\perp)|\right) = 1$$

using the fact that each vector in $f(\Sigma_\perp)$ is normalized. Then $|\psi_i\rangle\langle\psi_i|$ is a density operator on $\mathcal{H}^d$. But this means that $\bar{f}(p)$ is a convex sum of density operators, and hence also a density operator on $\mathcal{H}^d$. Now the important step: We show that the domain theoretic orthogonality of $\mathrm{Im}(p)$ implies that $\{|\psi_i\rangle\}$ is pairwise orthogonal.

Let $i \neq j$. Then $p_i \perp p_j$. Thus, these two strings differ at some index $k \leq \min\{|p_i|, |p_j|\}$, which means $p_i(k) \neq p_j(k)$. By the definition of f, these two symbols represent orthogonal vectors in $\mathcal{H}^n$, so $\langle f(p_i(k))|f(p_j(k))\rangle = 0$. We have

$$\langle\psi_i|\psi_j\rangle = \left(\prod_{m=1}^{\min\{|p_i|,|p_j|\}} \langle f(p_i(m))|f(p_j(m))\rangle\right) \cdot (\text{other terms}),$$

and since one value of m in the above product is k, we get $\langle\psi_i|\psi_j\rangle = 0$. Thus, $\{|\psi_i\rangle\}$ is a pairwise orthogonal subset of $\mathcal{H}^d$.

But this implies that each $\mu p_i/Z(p)$ is an eigenvalue of $\bar{f}(p)$. Since $\bar{f}(p)$ is a density operator, the sum of *all* its eigenvalues must be one. Thus, $\bar{f}(p)$ has exactly $\dim(p)$ nonzero eigenvalues, given by $g(p) = \mu p/Z(p)$. Its other $d(p) - \dim(p)$ eigenvalues must all be zero. By regarding $g(p) \in \Lambda^{\dim(p)}$ as a state in $\Lambda^{d(p)}$, we get the vector $g^*(p) \in \Lambda^{d(p)}$, which is the same as $\mathrm{spec}\,\bar{f}(p)$.

To prove $\bar{f}(p \otimes q) = \bar{f}(p) \otimes \bar{f}(q)$ when $|p|$ is constant, write

$$\bar{f}(p) = \sum_{i=1}^{\dim(p)} \frac{\mu p_i}{Z(p)} |\psi_i\rangle\langle\psi_i| \quad \text{and} \quad \bar{f}(q) = \sum_{j=1}^{\dim(q)} \frac{\mu q_j}{Z(q)} |\phi_j\rangle\langle\phi_j|$$

so that bilinearity and properties of the tensor on $P(D)$ immediately give

$$\bar{f}(p) \otimes \bar{f}(q) = \sum_{i=1}^{\dim(p)} \sum_{j=1}^{\dim(q)} \frac{\mu(p_i \otimes q_j)}{Z(p \otimes q)} \cdot |\psi_i\rangle\langle\psi_i| \otimes |\phi_j\rangle\langle\phi_j|$$

Now we write out

$$\bar{f}(p \otimes q) = \sum_{i=1}^{\dim(p)} \sum_{j=1}^{\dim(q)} \frac{\mu(p_i \otimes q_j)}{Z(p \otimes q)} \cdot |\alpha_{ij}\rangle\langle\alpha_{ij}|$$

The proof is finished if we can show that

$$|\alpha_{ij}\rangle\langle\alpha_{ij}| = |\psi_i\rangle\langle\psi_i| \otimes |\phi_j\rangle\langle\phi_j|.$$

But when $|p|$ is constant the calculation of $|\psi_i\rangle$ never involves $f(\perp)$,

$$\begin{aligned}
|\alpha_{ij}\rangle &= \left(\bigotimes_{k=1}^{|p_i|+|q_j|} f((p_i \otimes q_j)(k))\right) \otimes \left(\bigotimes_{k=|p_i|+|q_j|+1}^{|p|^+ + |q|^+} f(\perp)\right) \\
&= \left(\bigotimes_{k=1}^{|p_i|} f(p_i(k))\right) \otimes \left(\bigotimes_{k=1}^{|q_j|} f(q_j(k))\right) \otimes \left(\bigotimes_{k=|q_j|+1}^{|q|^+} f(\perp)\right) \\
&= |\psi_i\rangle \otimes |\phi_j\rangle
\end{aligned}$$

so we have exactly $|\alpha_{ij}\rangle\langle\alpha_{ij}| = |\psi_i\rangle\langle\psi_i| \otimes |\phi_j\rangle\langle\phi_j|$. $\square$

If the vector $|p|$ is constant, it means that p represents a density operator. In general, a process $p \in P(D)$ only partly represents a density operator, so we use the state $f(\perp)$ to 'complete' it to one. Now we can explain the iterates ρ^n. First, write ρ as

$$\rho = \sum_{i=1}^{n} \lambda_i |\psi_i\rangle\langle\psi_i|$$

where $\{|\psi_i\rangle\}$ is an orthonormal basis for $\mathcal{H}^n$, and we can assume all $\lambda_i > 0$. This information *implicitly defines* f as

$$f : \Sigma \to \mathcal{H}^n :: i \mapsto |\psi_i\rangle$$

where $\Sigma = \{1, \ldots, n\}$ and the measurement μ on Σ^∞ has $\mu i = \lambda_i$ extended homomorphically. For the process $q = (1, \ldots, n)$ we have $\bar{f}(q) = \rho$. And since $|q|$ is constant,

$$\bar{f}(q^n) = \bar{f}(q) \otimes \ldots \otimes \bar{f}(q) = \rho^n.$$

Thus, $q^n \in P(D)$ represents ρ^n. But what physical process does q represent? We think of it as the process of measuring the state of a system whose density operator is ρ.

Notice that ρ is an observable, being self-adjoint, so if we measure *it* when the system's density operator is ρ, the system changes to $|\psi_i\rangle\langle\psi_i|$ with probability λ_i. We have implicitly associated an algorithm to the measuring of observable ρ: If measuring ρ causes the system to jump to eigenstate $|\psi_i\rangle\langle\psi_i|$, this corresponds computationally to an algorithm q which has taken path q_i. Thus, the act of measuring itself corresponds to supplying the algorithm q with an input; only when this input is supplied does process q take the path q_i, and then it does so with probability λ_i.

Thus, it is possible to think about the work done by an algorithm q associated to the measuring of observable ρ; in our case, the work we expect q will do when the measurement of ρ is performed is

$$\begin{aligned}\langle q\rangle(g(q)) &= \sigma g(q) - \log Z(q)\\ &= \sigma g^*(q) - \log Z(q)\\ &= \sigma(\,\mathrm{spec}\,\bar{f}(q)\,) - \log Z(q)\\ &= \sigma(\mathrm{spec}(\rho)) - 0\\ &= \sigma\rho,\end{aligned}$$

the von Neumann entropy of density operator ρ. Similarly, we can represent the measurement of an arbitrary observable e: Define $f(i)$ to be the eigenstate of e associated to outcome i, and use both ρ and e to obtain the probabilities λ_i which define μ. Then $\bar{f}(q)$ is the density operator for the state of the system *after* the measurement of observable e, assuming ρ is its density operator before.

While the specifics of this section will not be needed in the next, the idea of associating an algorithm to various physical processes in such a way that the algorithm attempts to 'mimic' the process is a good one to keep in mind.

9 Entanglement and algorithmic complexity

Consider the molecule from Example 7.2 whose state is represented by a process p. Now consider two molecules $p \otimes p$ which are not interacting enough to worry about. Suppose that the first is in state p_i and that the second molecule is in state p_j. Then we would say that the state of the *joint system,* i.e., the two molecules considered together as a single system, is $p_i \otimes p_j$.

Conversely, we can ask, what does it mean to know the state of the joint system? That's simple, it means we know some element $p_i \otimes p_j \in \mathrm{Im}(p \otimes p)$, from which we can immediately deduce that the first molecule is in state p_i, while the second is in state p_j. It is this latter bit that does not always work in quantum mechanics: We can know the state of the joint system exactly without knowing the states of all its components exactly. Literally, the whole can be more than the sum of its parts.

To put some mathematics to this, suppose we have two systems, named 1 and 2, with respective state spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ of the same dimension n, and that the joint system is in the pure state described by the unit vector

$$|\psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2.$$

The state $|\psi\rangle$ is *entangled* if we can never write $|\psi\rangle = x \otimes y$ for $x \in \mathcal{H}_1$ and $y \in \mathcal{H}_2$.

How can we tell when a state is entangled, and if it is entangled, how can we measure 'how entangled' it is? In the bipartite setting that we are in, the *Schmidt decomposition* suggests a natural approach: For any normalized $|\psi\rangle$ in $\mathcal{H}_1 \otimes \mathcal{H}_2$, there are nonnegative real numbers $\lambda_i \geq 0$ such that

$$|\psi\rangle = \sum_{i=1}^{n} \lambda_i \, |\psi_i\rangle \otimes |\phi_i\rangle$$

where $\{|\psi_i\rangle\}$ and $\{|\phi_i\rangle\}$ are orthonormal bases of $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. By applying the Pythagorean identity on $\mathcal{H}_1 \otimes \mathcal{H}_2$,

$$|\,|\psi\rangle\,|^2 = \sum_{i=1}^{n} \lambda_i^2 \cdot |\,|\psi_i\rangle \otimes |\phi_i\rangle\,|^2 = \sum_{i=1}^{n} \lambda_i^2 \cdot |\,|\psi_i\rangle\,|^2 \cdot |\,|\phi_i\rangle\,|^2 = \sum_{i=1}^{n} \lambda_i^2 \cdot 1 \cdot 1 = \sum_{i=1}^{n} \lambda_i^2,$$

and since $|\psi\rangle$ is a unit vector,

$$\sum_{i=1}^{n} \lambda_i^2 = 1.$$

But more is true: This set of numbers is *unique.* The way to see this is to consider the information that $|\psi\rangle$ contains about the individual systems. For instance, the information $|\psi\rangle$ contains about 1 is the density operator on $\mathcal{H}_1$ given by

$$\rho_1 = \mathrm{tr}_2 \, |\psi\rangle\langle\psi|$$

where tr_2 is the operator that takes operators on $\mathcal{H}_1 \otimes \mathcal{H}_2$ to operators on $\mathcal{H}_1$. Intuitively, tr_2 removes the information $|\psi\rangle$ contains about 2, thereby yielding the information $|\psi\rangle$ contains about 1. Formally, it is defined on basic elements by

$$\mathrm{tr}_2 \, |x_1\rangle\langle x_2| \otimes |y_1\rangle\langle y_2| = \langle y_1|y_2\rangle \cdot |x_1\rangle\langle x_2|$$

for $|x_i\rangle \in \mathcal{H}_1$ and $|y_i\rangle \in \mathcal{H}_2$, and then extended additively to all density operators. Thus, we can find the density operator ρ_1 for system 1 by first noting that

$$|\psi\rangle\langle\psi| = \sum_{1 \leq i,j \leq n} \lambda_i \lambda_j \cdot |\psi_i\rangle\langle\psi_j| \otimes |\phi_i\rangle\langle\phi_j|$$

and then calculating

$$\rho_1 = \mathrm{tr}_2 \, |\psi\rangle\langle\psi| = \sum_{1 \leq i,j \leq n} \lambda_i \lambda_j \cdot \langle\phi_i|\phi_j\rangle \cdot |\psi_i\rangle\langle\psi_j|$$

By the pairwise orthogonality of $\{|\phi_i\rangle\}$, this simplifies to

$$\rho_1 = \sum_{i=1}^{n} \lambda_i^2 |\psi_i\rangle\langle\psi_i|.$$

Because $\{|\psi_i\rangle\}$ is an orthonormal basis for $\mathcal{H}_1$, we have $\mathrm{spec}(\rho_1) = \{\lambda_i^2\}$, and since the spectrum of an operator is a unique set, the *Schmidt coefficients* in the decomposition of a normalized $|\psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$ define a *function* $\mathrm{sc}_1 : \mathcal{H}_1 \otimes \mathcal{H}_2 \to \Lambda^n$ given by

$$\mathrm{sc}_1 \, |\psi\rangle = \mathrm{sort}^-(\, \mathrm{spec}(\mathrm{tr}_2 \, |\psi\rangle\langle\psi|)\,) = \mathrm{sort}^-(\lambda_1^2, \ldots, \lambda_n^2).$$

Interestingly, the function sc_2, which extracts the classical information $|\psi\rangle$ contains about subsystem 2, is *equal* to sc_1. Thus, we simply name them both by

$$\mathrm{sc}\,|\psi\rangle := \mathrm{sc}_1\,|\psi\rangle = \mathrm{sc}_2\,|\psi\rangle$$

However, were we studying *multipartite entanglement* we would expect *minimally* to consider maps sc_i, one for each subsystem i.

Thus, $|\psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$ is *not entangled* iff $(\exists i)\,\lambda_i = 1$ iff $\sigma(\mathrm{sc}|\psi\rangle) = 0$, a wonderful application of the measurement $\sigma : (\Lambda^n, \leq) \to [0,\infty)^*$. And generally speaking, we can see that it is reasonable to regard $\sigma(\mathrm{sc}|\psi\rangle)$ as measuring the extent to which $|\psi\rangle$ is entangled. Knowing the connection between σ and $\leq$ given in Theorem 5.1, we would expect $\leq$ to be relevant in the study of entanglement as well. And it is:

Theorem 9.1 (Nielsen [10]) *State $|\psi\rangle$ can be converted into state $|\phi\rangle$ by means of local operations and classical communication iff* $\mathrm{sc}|\psi\rangle \leq \mathrm{sc}|\phi\rangle$.

This result characterizes exactly when it is possible to transform entanglement using *local operations and classical communication.* Here is what this means: There are two spatially separated systems, one controlled by a party named 1, the other controlled by party 2. To say they control their local systems means they are free to do anything they want to them. The two parties are also allowed to communicate with one another using classical communication. This process may be used to transform the state of the joint system from one pure state to another if and only if the spectrum of each subsystem's density operator moves up in the domain $(\Lambda^n, \leq)$. You will notice that the effect of this process is that it reduces entanglement:

$$\mathrm{sc}|\psi\rangle \leq \mathrm{sc}|\phi\rangle \;\Rightarrow\; \sigma(\mathrm{sc}|\psi\rangle) \geq \sigma(\mathrm{sc}|\phi\rangle)$$

Thus, corresponding to a measure of entanglement σ, there is a process by which entanglement is reduced $\leq$. But intuitively, in order to reduce entanglement, we must do algorithmic work.

Imagine an entangled pure state $|\psi\rangle$ of a system with several subsystems whose entanglement is successively reduced to zero; write this as

$$\lim_{n\to\infty} p^n|\psi\rangle = |\phi\rangle.$$

The state $|\phi\rangle$ is pure and *not entangled,* so during this process the density operator ρ of any given subsystem must be approaching a pure state, which means its entropy $\sigma\rho$ eventually decreases to zero. With no loss of generality, we can assume that $\sigma\rho$ decreases with n. Now imagine an algorithm q which 'mimics' ρ, perhaps like the one we saw in the last section. The classical state $g^*(q)$ representing our knowledge about the output of q is mathematically equal to $\mathrm{spec}(\rho)$. But since the entropy of ρ decreases with n, so too must the entropy of $g^*(q)$, which means our knowledge of the output of q is increasing. The only way our knowledge about the output of q can increase is if we do algorithmic work. The result of this work is that we move to a new state of knowledge from which the expected complexity should be *less* than it was originally. Thus, writing $\mathrm{ent}|\psi\rangle$ for the amount of entanglement, we expect

$$\frac{d}{dt}\mathrm{ent}|\psi\rangle \leq 0 \;\Rightarrow\; \frac{d}{dt}\langle q\rangle \leq 0.$$

That is, in order to reduce entanglement, we must do algorithmic work. In the bipartite setting, the converse is also possible:

Proposition 9.2 *$|\psi\rangle \to |\phi\rangle$ using local operations and classical communication iff*

$$\langle p\rangle(\mathrm{sc}|\psi\rangle) \geq \langle p\rangle(\mathrm{sc}|\phi\rangle)$$

for every binary decision tree $p : \{1, \ldots, n\} \to 2^\infty$.

Proof. A process $p \in P^n(2^\infty)$ on the domain of binary streams gives rise to a binary tree $\downarrow\mathrm{Im}(p)$ whose leaves are $\mathrm{Im}(p)$. Conversely, a binary tree with n leaves yields a process $p \in P^n(2^\infty)$. Now the result follows from Theorem 4.7. $\Box$

By Theorem 4.7, the last result can be stated for an arbitrary symbolic domain $(D, \otimes, \mu)$ in possession of a single binary process. We have simply chosen to emphasize the case of binary trees. An interesting question is to ask if there are other forms of entanglement transformation that are equivalent to simultaneously reducing the complexity of a class of algorithms. But a more interesting question is to ask if there are forms which are *not*.

Denouement

To review some of what we have seen:

- Instead of using measurement to understand the calculation of a fixed point, we have used fixed points to understand a measurement (σ).
- Instead of using a partial order to understand computation, we have used computation to understand a partial order ($\leq$).
- We have shown how to use entropy in the study of algorithmic complexity; we have shown how to derive entropy from basic concepts in computer science.
- Algorithms? Processes. Monotone functions? Antichains.

We have shown that information arises as the least fixed point of copying that is above complexity; this is one possible formulation of the idea that information is complexity relative to the class of all conceivable processes. Another is given by the universal limit, which we like to write as

$$\lim_{D\in\Sigma} (h_D, \leq_D) = (\sigma, \leq).$$

The viewpoint we developed about these ideas in the classical setting applies to the quantum setting as well, and we have proven this. This should be taken as evidence that *one possible* pragmatic definition of information is as an *extension of complexity.* An illustration of the effectiveness of these simple ideas is the ease with which a rigorous connection between entanglement transformation and algorithmic complexity was established. We have also unintentionally answered questions like "How can physics benefit domain theory?" and "What is the quantum mechanical significance of the Scott topology on Λ^n?"

Ideas

We have introduced a number of new ideas in this paper. A strong desire to keep things as short and to the point as possible has prevented examining them in any real detail.

- The order theoretic structure of $(\Lambda^n, \leq, \sigma)$ is richer than we have indicated here, and the author believes this additional structure has fundamental consequences for thermodynamics.
- There is a lot one can say about symbolic domains: Induction, integration, etc. The idea of algebra coexisting with measurement is exciting.
- Domain theoretic orthogonality needs lots of attention in view of its importance for complexity in computer science and physics; intuitively, a more general definition is possible which oddly enough leads to the exact same quantitative theory. Why?
- Orthogonality will be useful for ordering $P(D)$. The map $p \mapsto 1 - Z(p)$ is its natural measurement.
- There should be a class of measurements which unifies the triangle inequality with the subadditivity needed for orthogonal sets.
- Far more is possible with the inequalities relating h to σ; a single inequality should give $h = \sigma$ on $\mathbf{I}[0,1]$ and $h \leq 1 + \sigma$ on 2^∞.
- The universal limit should also hold over the class of ω algebraic Scott domains (D, μ).
- Define the complexity of a domain (D, μ) to be the number

  $$c(D) := h_D(\perp).$$

  Is there a domain constructor $*$ such that

  $$c(D * E) = c(D) \cdot c(E)$$

  where $\cdot$ is multiplication? If so, then notice that one should now be able to discuss polynomial time computability by repeatedly multiplying copies of the lazy naturals.

As one would expect, the complexity h seems most effective on domains that you can draw using Hasse diagrams. Its calculation should probably not be taken lightly.

Acknowledgement

it seems like such a simple thing,
as does a little girl smiling on a swing that goes a lot;
it seems like such a simple thing,
but i have watched long enough to know that it is not.
– from 'stargazing' (written for an angel in Katowice).

I thank my darling Maja for *putting on* the dress which inspired this paper, and for *putting up* with a man determined to perfect the art of not sleeping.

References

[1] S. Abramsky and A. Jung. *Domain theory.* In S. Abramsky, D. M. Gabbay, T. S. E. Maibaum, editors, Handbook of Logic in Computer Science, vol. III. Oxford University Press, 1994.

[2] P. M. Alberti and A. Uhlmann. *Stochasticity and partial order: doubly stochastic maps and unitary mixing.* Dordrecht, Boston, 1982.

[3] L. G. Kraft. *A device for quantizing, grouping and coding amplitude modulated pulses.* M.S. Thesis, Electrical Engineering Department, MIT, 1949.

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

[5] R. F. Muirhead. *Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters.* Proc. Edinburgh Math. Soc., 21:144-157, 1903.

[6] *A foundation for computation.* Ph.D. Thesis, Department of Mathematics, Tulane University, 2000.

[7] *A triangle inequality for measurement.* To appear.

[8] *Unique fixed points in domain theory.* Electronic Notes in Theoretical Computer Science, vol. 45, 2001.

[9] M. A. Nielsen. *Probability distributions consistent with a mixed state.* Physical Review A **62**, 052308, 2000.

[10] M. A. Nielsen. *Conditions for a class of entanglement transformations.* Physical Review Letters, vol.83 (2), p.436–439, 1999.

[11] C. E. Shannon. *A mathematical theory of communication.* Bell Systems Technical Journal 27, 379–423 and 623–656, 1948.

[12] E. Schrödinger. Proceedings of the Cambridge Philosophical Society **32**, 446 (1936).

[13] A. Uhlmann. *On the Shannon entropy and related functionals on convex sets.* Reports on Mathematical Physics, Volume 1, Issue 2, p. 147–159, 1970.

Appendix: A domain from 1903

Here is the proof that majorization is a domain; those unfamiliar with domains may find the appendix 'Domain theory' at the end of this paper helpful.

Lemma 9.3 *Λ^n is a dcpo with least element $\perp := (1/n, \ldots, 1/n)$.*

Proof. Increasing sequences have suprema. To prove that all directed sets do, we use the usual trick from the study of measurement, and demonstrate the existence of a strictly monotone map $\sigma : \Lambda^n \to [0,\infty)^*$ that preserves suprema of increasing sequences. We can generate them at will as follows: The map

$$\Lambda^n \to \prod_{i=1}^{n}[0,1] :: x \mapsto (s^1 x, \ldots, s^n x)$$

is an order embedding between posets, so if f_i is a measurement on $[0,1]$, then

$$\sigma x = \sum_{i=1}^{n} f_i(s^i x)$$

is such a map on Λ^n. Now the proof is finished. □

Lemma 9.4 *Let $\pi : [0,1] \to \Lambda^n$ be the straight line path from $\perp$ to x,*

$$\pi(t) = (1-t)\perp + tx.$$

Then it is Scott continuous and $\pi(t) \ll x$ for each $t < 1$. Thus, Λ^n is continuous.

Proof. The following equality is helpful in this proof:

$$s^i \pi(t) = \frac{i}{n}(1-t) + t \cdot s^i x.$$

For monotonicity of π, if $s < t$, then

$$s^i \pi(s) \leq s^i \pi(t) \Leftrightarrow \frac{i}{n} \leq s^i x$$

which is clear using $\perp \sqsubseteq x$. Scott continuity follows from Euclidean continuity.

To prove $\pi(t) \ll x$ for $t < 1$, consider an increasing sequence $(x_k) \in \Lambda^n$ with $x \sqsubseteq \bigsqcup x_k$. Because $t < 1$, $s^i \pi(t) < s^i x$ for all $1 \leq i < n$; for $i = n$ they are both equal to one. Since $x \sqsubseteq \bigsqcup x_k$, we have

$$s^i \pi(t) < s^i x \leq \lim_{k \to \infty} s^i x_k,$$

for all $1 \leq i < n$. Since there are only finitely many i, there must exist some k with $s^i \pi(t) \leq s^i x_k$ for all $1 \leq i < n$. For this same k, $s^n \pi(t) = s^n x_k = 1$, which gives $\pi(t) \leq x_k$. This proves $\pi(t) \ll x$.

Using the Scott continuity of π, $\mathord{\downarrow\!\!\downarrow} x$ contains an increasing sequence with supremum x, and this means that $\mathord{\downarrow\!\!\downarrow} x$ itself is directed with supremum x, which is the continuity of Λ^n. □

Theorem 9.5 *Λ^n is a continuous dcpo.*

Appendix: The tensor product

Given two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, of any dimension whatsoever, their Hilbert space *tensor product* is the completion of their algebraic tensor product. Specifically, we form the algebraic tensor product, then introduce an inner product, and then complete it. First, the algebra, where we assume for simplicity that we have complex vector spaces.

Definition 9.6 The *algebraic tensor product* of two vector spaces V and W is a pair $(T, \otimes)$ where T is a vector space and $\otimes : V \times W \to T$ is a bilinear map such that

$$\begin{tikzcd} T \arrow[r, "\exists!\ \bar{f}"] & \bullet \\ V \times W \arrow[u, "\otimes"] \arrow[ur, "\forall f"'] & \end{tikzcd}$$

That is, for every bilinear f into any other vector space $\bullet$, there is a unique linear $\bar{f}$ that makes the diagram commute. We write $T = V \otimes W$.

This construction is purely algebraic, usually encountered in the study of modules over rings. The elements of $V \otimes W$ are finite linear combinations of elements $v \otimes w$ where $v \in V$ and $w \in W$. Given two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$, then, we form the complex vector space $\mathcal{H}_1 \otimes \mathcal{H}_2$, and then introduce an inner product on it

$$\langle \alpha(x_1 \otimes y_1) | \beta(x_2 \otimes y_2) \rangle = \alpha^* \beta \langle x_1 | x_2 \rangle \langle y_1 | y_2 \rangle$$

using the inner products from $\mathcal{H}_1$ and $\mathcal{H}_2$. This definition is extended linearly using the convention of quantum mechanics, conjugate linearity in the first argument. Next, if V is a vector space with an inner product, then its metric completion $\bar{V}$ (with respect to the metric derived from the inner product) can be achieved by an inner product on $\bar{V}$ which extends the one on V. Thus, the metric completion of the inner product space $(\mathcal{H}_1 \otimes \mathcal{H}_2, \langle \cdot | \cdot \rangle)$ is a Hilbert space called *the tensor product* $\mathcal{H}_1 \otimes \mathcal{H}_2$.

If we have two Hilbert spaces of finite dimension m and n, respectively, their tensor product will have dimension mn. However, complex/real Hilbert spaces are isomorphic iff they have the same dimension, which means that there is a unitary surjective linear map between them which preserves the inner products. Thus, we can take the tensor product of such spaces to be $\mathcal{H}^{mn}$, where $\mathcal{H}$ is a one dimensional complex Hilbert space.

Appendix: Domain theory

Let $(P, \sqsubseteq)$ be a partially ordered set or *poset* [1]. A nonempty subset $S \subseteq P$ is *directed* if $(\forall x, y \in S)(\exists z \in S)\, x, y \sqsubseteq z$. The *supremum* $\bigsqcup S$ of $S \subseteq P$ is the least of its upper bounds when it exists. A *dcpo* is a poset in which every directed set has a supremum.

For elements x, y of a dcpo D, we write $x \ll y$ iff for every directed subset S with $y \sqsubseteq \bigsqcup S$, we have $x \sqsubseteq s$, for some $s \in S$.

Definition 9.7 Let $(D, \sqsubseteq)$ be a dcpo. We set

- $\twoheaddownarrow x := \{y \in D : y \ll x\}$ and $\twoheaduparrow x := \{y \in D : x \ll y\}$
- $\downarrow x := \{y \in D : y \sqsubseteq x\}$ and $\uparrow x := \{y \in D : x \sqsubseteq y\}$

and say D is *continuous* if $\twoheaddownarrow x$ is directed with supremum x for each $x \in D$.

The *Scott topology* on a continuous dcpo D has as a basis all sets of the form $\twoheaduparrow x$ for $x \in D$. A function $f : D \to E$ between domains is *Scott continuous* if it reflects Scott open sets. This is equivalent to saying that f is *monotone*,

$$(\forall x, y \in D)\, x \sqsubseteq y \Rightarrow f(x) \sqsubseteq f(y),$$

and that it *preserves directed suprema*:

$$f(\bigsqcup S) = \bigsqcup f(S),$$

for all directed $S \subseteq D$. Like complete metric spaces, domains also have a result which guarantees the existence of *canonical* fixed points.

Theorem 9.8 *If $f : D \to D$ is a Scott continuous map on a dcpo and there is an $x \in D$ with $x \sqsubseteq f(x)$, then*

$$\mathrm{fix}(f) := \bigsqcup_{n \geq 0} f^n(x)$$

is the least fixed point of f on the set $\uparrow x$.

Notice the case of the above result when a dcpo has a least element $x = \perp$.

Definition 9.9 A *basis* for a domain D is a subset $B \subseteq D$ such that $B \cap \twoheaddownarrow x$ is directed with supremum x, for all $x \in D$.

Definition 9.10 A domain is *ω-continuous* if it has a countable basis.

Definition 9.11 An element x in a domain D is *compact* if $x \ll x$. The set of compact elements is denoted by $K(D)$. A domain D is *algebraic* if $K(D)$ forms a basis for D. An algebraic domain is *ω-algebraic* if its set of compact elements is countable.

Definition 9.12 A continuous dcpo is *compact* if it is Scott compact and the intersection of any two Scott compact upper sets is Scott compact.

The terminology "compact" is based on the fact that the above condition is equivalent to compactness in the Lawson topology.

Definition 9.13 A domain is *semantic* if it is ω-algebraic and compact.

By carefully analyzing the behavior of a while loop which generates successively better approximations to some desired ideal object, we become aware of the fact that our ability to distinguish between algorithms which progress and those which do not depends in a subtle way on the relationship between the ε-*approximations* around a point x,

$$\mu_\varepsilon(x) := \{y \in D : y \sqsubseteq x \ \& \ \varepsilon \ll \mu y\},$$

where $\mu : D \to E$ is a continuous map between domains, and our qualitative understanding expressed by $\sqsubseteq$. The study of measurement connects this pragmatic desire with an element of the imagination: "information content." Notice that the former is real, meaning that it is specific and concrete in purpose; the latter is anyone's guess.

Definition 9.14 A continuous map $\mu : D \to E$ between domains is said to *measure* the set $X \subseteq D$ if for all $x \in X$ and all Scott open sets $x \in U \subseteq D$, we have $x \in \mu_\varepsilon(x) \subseteq U$, for some $\varepsilon \ll \mu x$. We sometimes write $\mu \to \sigma_X$.

The terminology "induces the Scott topology near X" is used in [6], but we prefer to reserve it mostly for topological discussions. One of the fundamental properties of measurements is *strict monotonicity.*

Proposition 9.15 *Let $\mu : D \to E$ be a map that measures $X \subseteq D$. Then for all $x \in D$ and $y \in X$, we have $x \sqsubseteq y$ and $\mu x = \mu y \Rightarrow x = y$.*

The case $E = [0, \infty)^*$, the nonnegative reals in their *opposite* order, warrants special attention.

Definition 9.16 A *measurement* is a continuous map $\mu : D \to [0, \infty)^*$ that measures $\ker \mu = \{x \in D : \mu x = 0\}$.

By Prop. 9.15, note that $\ker \mu \subseteq \max D = \{x \in D : \uparrow\!x = \{x\}\}$. That is, an element with *no uncertainty* is maximal in the information order.