Domain theory and measurement

Keye Martin

Naval Research Laboratory
Center for High Assurance Computer Systems
Washington DC 20375
keye.martin@nrl.navy.mil

Summary. Lecture notes on domain theory and measurement, driven by applications to physics, computer science and information theory, with a hint of provocation.

Contents

1 Introduction

History

I loved everything about being a graduate student except going to class, doing homework, taking exams, fulfilling requirements associated with earning a degree and being severely underpaid. What I especially loved was being able to do mathematics. One semester I signed up for a course on domain theory. I went to the first lecture and heard about dcpo's and the fixed point theorem: every Scott continuous map $f : D \to D$ on a dcpo D with a least element $\perp$ has a least fixed point given by

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

I thought it was neat, so I skipped the rest of my classes that day and immediately went home to try it out and see how it worked. I wrote down an example of a function on the interval domain $D = \mathbf{I}\mathbb{R}$, this one: for a continuous $f : \mathbb{R} \to \mathbb{R}$ on the real line, define

$$\mathrm{split}_f : C(f) \to C(f)$$

$$\mathrm{split}_f[a,b] = \begin{cases} \mathrm{left}[a,b] & \text{if } \mathrm{left}[a,b] \in C(f); \\ \mathrm{right}[a,b] & \text{otherwise,} \end{cases}$$

where $C(f)$ is the subset of $\mathbf{I}\mathbb{R}$ where f changes sign,

$$C(f) = \{[a,b] : f(a) \cdot f(b) \leq 0\},$$

and $\mathrm{left}[a,b] = [a,(a+b)/2]$ and $\mathrm{right}[a,b] = [(a+b)/2,b]$. If we begin from any interval $[a,b] \in C(f)$ on which f changes sign, then

$$\bigsqcup_{n\geq 0} \mathrm{split}_f^n[a,b]$$

is a fixed point of split_f, just like the fixed point theorem says it should be. Because $\mathrm{fix}(\mathrm{split}_f) = \{x \in \mathbf{I}\mathbb{R} : \mathrm{split}_f(x) = x\} = \{[r] : f(r) = 0\}$, iterating split_f is a scheme for calculating a solution of the equation $f(x) = 0$.

The problem was: split_f was *not* Scott continuous, so the fixed point theorem could not be used to explain its behavior on $\mathbf{I}\mathbb{R}$. And there was an especially easy way to see it: plenty of functions f have more than one zero on a given interval x – but if split_f is Scott continuous, its least fixed point on $\uparrow\!x$ is unique (being maximal), implying that f has only one zero on x. So then the question became: why did this function behave as though it were continuous? I set about to find an answer. In the process, I became so interested in domain theory that I never went back to class again.

What I learned was that there was a reason that this function behaved as though it were continuous: it wasn't, but its *measure* was. Its measure was an important thing: the length function $\mu[a, b] = b - a$ was different from other functions. I later learned that all recursive functions could be modeled in this way and that the *measurement* μ was intertwined with the structure of the domain itself. It provided a measure of information content and one could use this idea to measure the ‘rate’ at which a process on a domain manipulated information and to do more things than should be mentioned in an introduction.

I was never really able to finish telling the story in my doctoral thesis the way I thought it should have been told. Nevertheless, at two hundred pages, I decided to stop typing and go to sunny England, where the theory advanced, with the same structure found in computation (domains and measurements) also being found in quantum mechanics and general relativity. Later, it was realized that the same structure was also present in information theory: there was a domain of binary channels, for instance, with capacity as a measurement. In all these cases, there are neat applications and new perspectives offered on ideas we previously misled ourselves into believing we understood. The interaction between these areas is what makes the study of domains and measurements very exciting.

That brings us to now. This is a ‘tutorial’ on domain theory and measurement. It is about what we believe we know today. It’s also about what we believe we don’t know today. There are also new results and ideas here never published before.

Overview

In Section 2, the basic elements of domains and measurements are introduced where our goal is to explain partiality and content as concepts and to explain how one models them with domains and measurements in practice so that new problems can be solved. In essence, the goal is to teach “the method” of finding domains and measurements in nature. A dozen or so basic examples are given. In Section 3, we give an important example of what one does with domains and measurements: applies fixed point theorems. Applications include numerical methods and fractals. In Section 4, we give more advanced examples of partiality and content: the domain of real analytic mappings, the domain of finite probability distributions, the domain of quantum mixed states and from general relativity, the domain of spacetime intervals. Applications of these domains are to the computation of real analytic mappings, the maximum entropy state in statistical mechanics, classical and quantum communication and to the reconstruction of spacetime from a countable set, including its geometry.

In Section 5, we discuss the informatic derivative: when an operator on a domain iterates to a fixed point, its informatic derivative measures the rate at which the iterates converge. Applications are to numerical analysis, to the computation of the Holevo capacity in quantum information theory and to the complexity of list algorithms. The informatic derivative applies to both continuous and *discrete* data. In Section 6, we discuss additional models of 'process' that have proven themselves useful in the measurement formalism: the renee equation, trajectories, vectors. Applications include showing that each order on a domain gives rise to a natural notion of computability, such as the primitive and partial recursive functions; the analysis of algorithms as trajectories on domains, such as Grover's algorithm for quantum searching, whose complexity is the amount of time it takes a trajectory to reach its maximum in the information order; an analysis of how noise affects communication with qubits; the derivation of lower bounds on the complexity of algorithms such as sorting and searching and the fixed point theorem: entropy is the least fixed point of the copying operator that is above algorithmic complexity.

In Section 7, we give a brief overview of where things currently stand in the study of domains and measurements, and try to persuade domain theorists in search of a decent job to send us an email.

To the student

A *student* is any person young enough at heart to be open to new ideas. This paper is written for students. We have tried to strike a balance between *philosophy*, *mathematics* and *applicability*. Philosophy: what is the big picture? Mathematics: how do we learn about the big picture? Applicability: what can the big picture teach us about the world we live in? Philosophy is good because *thinking* is good. Mathematics is good because knowing *what* you are thinking about is good. Applicability is good because knowing *why* you are thinking what you are thinking about is good. It is pretty rare that a set of ideas starts off with all three of these in equal measure. Sometimes there is only philosophy, sometimes only math, sometimes only a question. But as a set of ideas evolves, one hopes to see the appearance of all three.

2 The basic elements

Most newcomers to domain theory stop reading when they see domains presented as a seemingly endless list of axioms satisfied by partial orders. If this is your first time reading about domain theory, perhaps you should consider a different approach. Try first reading Section 2.1 to understand the ideas intuitively. Then go to Section 2.2, but ignore the technical definitions and just look at the dozen or so examples given instead. After those examples, have a look at Section 4, where there are more involved examples. Then ask yourself a question: given the intuitions on partiality and information content combined with the numerous instances of the idea that you have seen, how would *you* formally capture those ideas?

If you find a formal mathematical definition of domain and measurement that captures all of the examples, compare it to the formalizations given in Sections 2.2 and 2.3. If your formalization differs, it might be time to stop reading these notes and to pursue your own direction. If it is the same, then you will understand the basic definitions of domain theory and measurement in a way few people do. And if you are unable to come up with a formalization that captures the basic examples, then you will better appreciate definitions like ‘continuous dcpo’ and ‘measurement’ – you will see them for what they are: a significant step toward a mathematical definition of ‘information’.

Major references: [1, 15].

2.1 Intuition

A *domain* $(D, \sqsubseteq)$ is a set of objects D together with a partial order $\sqsubseteq$ that has certain intrinsic notions of completeness and approximation defined by the order. The order $\sqsubseteq$ is thought of as an *information order*. Intuitively, $x \sqsubseteq y$ means “x contains information about y” or that “x carries information about y.” We might also say y is at least as informative as x – though this is really just mathematical uptightness that obscures the essence of the idea: when talking to one’s friends, people always just say that $x \sqsubseteq y$ means y is *more informative* than x. Elements that compare in the information order are *comparable* and the thing to remember about comparable elements is that *one of them carries information about the other*.

The *completeness* in a domain refers to the fact that certain results generated by processes have ‘limits’. For instance, if a process generates a sequence (x_n) of elements that *increase* with respect to the information order, $x_n \sqsubseteq x_{n+1}$ for all n, then it should ‘go somewhere’ i.e.

$$x_1 \sqsubseteq x_2 \sqsubseteq \ldots \Longrightarrow \bigsqcup_{n\in\mathbb{N}} x_n \in D$$

The element $\bigsqcup_{n\geq\mathbb{N}} x_n$ is not only above each x_n in the information order, it is the ‘best’ such object. Intuitively, if the process generating (x_n) is an algorithm repeatedly producing iterates x_n, then $\bigsqcup_{n\geq\mathbb{N}} x_n$ is the *final answer*.

The notion of *approximation* is a special case of the information order. If x approximates y, we write $x \ll y$. What it means intuitively is that x carries *essential* information about y. But what does “essential” mean? One view of essential is that any process that produces a sequence (x_n) of values with $\bigsqcup x_n = y$ must satisfy $x \sqsubseteq x_n$ for all but a finite number of the x_n. That is, we cannot compute y without first computing in *finite time* an object that x carries information about. Thus, x can also be thought of as a *finite approximation* of y. Put yet another way, $x \ll y$ means that all informatic paths to y must pass through x.

An *ideal* (or total) object x in a domain D is one that we can only get to using a process that constructs a sequence of finite approximations. For example, a *maximal element* $x \in D$ is an object that cannot be improved upon i.e.

$$(\forall y \in D)\ x \sqsubseteq y \Rightarrow x = y.$$

Each maximal element is an example of an ideal element. Any object that is not ideal (or total) is called *partial*. Let us give several intuitive examples of ideal and partial objects.

A compact interval $[a, b]$ of the real line provides a *partial* description of a real number; a one point interval $[x, x]$ is total. The uniform probability distribution $\perp = (1/n, \ldots, 1/n)$ provides incomplete information on the expected outcome of an experiment, while the finite probability distribution $(1, 0, \ldots, 0)$ predicts the outcome with certainty. The polynomial $1 + x$ is a finite approximation of the analytic mapping e^x. A pure state $|\psi\rangle\langle\psi|$ in quantum mechanics is *total*; a mixed state like $\perp = I/n$ is partial. An infinite set of natural numbers is total while a finite subset of it provides a finite approximation.

A *measurement* $\mu : D \to [0, \infty)$ is a function on a domain D that to each informative object $x \in D$ assigns a number μx that measures the *amount of partiality* in x. The amount of partiality, or *uncertainty*, in an object is also called its *information content*. For instance, we would expect uncertainty to decrease as we move up in the information order,

$$x \sqsubseteq y \Rightarrow \mu x \geq \mu y.$$

If a process calculates $x = \bigsqcup x_n$, we would expect

$$\mu\left(\bigsqcup_{n \in \mathbb{N}} x_n\right) = \lim_{n \to \infty} \mu x_n.$$

If x and y are comparable and $\mu x = \mu y$, then this means that one carries information about the other and that they have the same information content, so we would expect $x = y$. In particular, if $\mu x = 0$, so that x is an object with no uncertainty, then we would expect that x cannot be improved upon. That is, we would expect x to be maximal in the information order.

2.2 Domains

In this section, we give several basic examples of domains, including the formal definition of a continuous dcpo. At no point in this section will we define "domain," though we will quite frequently make statements like "such and such is an example of a domain." There is a good reason for our vagueness, but at this point in time, we intend to remain vague about it.

The intrinsic notion of completeness in a domain is at least partially captured by the fact that it forms a dcpo:

Definition 1. Let $(P, \sqsubseteq)$ be a partially ordered set or *poset*. 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.

One way to formalize the intrinsic notion of approximation possessed by a domain is *continuity*:

Definition 2. Let $(D, \sqsubseteq)$ be a dcpo. For elements $x, y \in D$, we write $x \ll y$ iff for every directed subset S with $y \sqsubseteq \bigsqcup S$, we have $x \sqsubseteq s$, for some $s \in S$. 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 $\uparrow x := \{y \in D : x \sqsubseteq y\}$

A set $B \subseteq D$ is a *basis* when $B \cap \twoheaddownarrow x$ is directed with supremum x for each $x \in D$. A dcpo is *continuous* when it has a basis and ω*-continuous* when it has a countable basis.

Remark: Any continuous dcpo is an example of a *domain.*

Example 1. The collection of compact intervals of the real line

$$\mathbf{I}\mathbb{R} = \{[a, b] : a, b \in \mathbb{R} \ \& \ a \leq b\}$$

ordered under reverse inclusion

$$[a, b] \sqsubseteq [c, d] \Leftrightarrow [c, d] \subseteq [a, b]$$

is an ω-continuous dcpo:

- For directed $S \subseteq \mathbf{I}\mathbb{R}$, $\bigsqcup S = \bigcap S$,
- $I \ll J \Leftrightarrow J \subseteq \mathrm{int}(I)$, and
- $\{[p, q] : p, q \in \mathbb{Q} \ \& \ p \leq q\}$ is a countable basis for $\mathbf{I}\mathbb{R}$.

The domain $\mathbf{I}\mathbb{R}$ is called the *interval domain.* If we replace $\mathbb{R}$ by $[0, 1]$, then we obtain the *interval domain* $\mathbf{I}[0, 1]$ over the unit interval.

A binary channel has two inputs ("0" and "1") and two outputs ("0" and "1"). An input is sent through the channel to a receiver. Because of noise in the channel, what arrives may not necessarily be what the sender intended. The effect of noise on input data is modelled by a noise matrix u. If data is sent through the channel according to the distribution x, then the output is distributed as $y = x \cdot u$. The noise matrix u is given by

$$u = \begin{pmatrix} a & \bar{a} \\ b & \bar{b} \end{pmatrix}$$

where $a = P(0|0)$ is the probability of receiving 0 when 0 is sent and $b = P(0|1)$ is the probability of receiving 0 when 1 is sent and $\bar{x} := 1 - x$ for $x \in [0, 1]$. Thus, the noise matrix of a binary channel can be represented by a point (a, b) in the unit square $[0, 1]^2$ and all points in the unit square represent the noise matrix of some binary channel.

Example 2. Binary channels. The set of nonnegative noise matrices

$$\mathbb{N} = \left\{ (a,b) = \begin{pmatrix} a\ \bar{a} \\ b\ \bar{b} \end{pmatrix} : a \geq b \ \&\ a, b \in [0,1] \right\}$$

is in bijective correspondence with $\mathbf{I}[0,1]$ via $(a,b) \mapsto [b,a]$. With the order it inherits from $\mathbf{I}[0,1]$, $\mathbb{N}$ is called the *domain of binary channels.*

Example 3. Let X be a locally compact Hausdorff space. Its *upper space*

$$\mathbf{U}X = \{\emptyset \neq K \subseteq X : K \text{ is compact}\}$$

ordered under reverse inclusion

$$A \sqsubseteq B \Leftrightarrow B \subseteq A$$

is a continuous dcpo:

- For directed $S \subseteq \mathbf{U}X$, $\bigsqcup S = \bigcap S$, and
- $A \ll B \Leftrightarrow B \subseteq \text{int}(A)$.

Example 4. Given a metric space (X,d), the *formal ball model* [6]

$$\mathbf{B}X = X \times [0,\infty)$$

is a poset when ordered via

$$(x,r) \sqsubseteq (y,s) \Leftrightarrow d(x,y) \leq r - s.$$

The approximation relation is characterized by

$$(x,r) \ll (y,s) \Leftrightarrow d(x,y) < r - s.$$

The poset $\mathbf{B}X$ is continuous. However, $\mathbf{B}X$ is a dcpo iff the metric d is complete. In addition, $\mathbf{B}X$ has a countable basis iff X is a separable metric space.

Definition 3. An element x of a poset is *compact* if $x \ll x$. A poset is *algebraic* if its compact elements form a basis; it is ω*-algebraic* if it has a countable basis of compact elements.

Example 5. The powerset of the naturals

$$\mathcal{P}\omega = \{x : x \subseteq \omega\}$$

ordered by inclusion $x \sqsubseteq y \Leftrightarrow x \subseteq y$ is an ω-algebraic dcpo:

- For directed set $S \subseteq \mathcal{P}\omega$, $\bigsqcup S = \bigcup S$,
- $x \ll y \Leftrightarrow x \sqsubseteq y \ \&\ x$ is finite, and
- $\{x \in \mathcal{P}\omega : x \text{ is finite}\}$ is a countable basis for $\mathcal{P}\omega$.

Example 6. Binary strings. The collection of functions

$$\Sigma^\infty = \{ \, s \mid s : \{1, \ldots, n\} \to \{0,1\}, 0 \leq n \leq \infty \, \}$$

ordered by extension

$$s \sqsubseteq t \Leftrightarrow |s| \leq |t| \;\&\; (\, \forall \, 1 \leq i \leq |s| \,) \; s(i) = t(i),$$

where $|s|$ is the cardinality of $\mathrm{dom}(s)$, is an ω-algebraic dcpo:

- For directed $S \subseteq \Sigma^\infty$, $\bigsqcup S = \bigcup S$,
- $s \ll t \Leftrightarrow s \sqsubseteq t \;\&\; |s| < \infty$,
- $\{s \in \Sigma^\infty : |s| < \infty\}$ is a countable basis for Σ^∞,
- The least element $\perp$ is the unique s with $|s| = 0$.

A *partial function* (or partial map) on a set X is a function $f : A \to X$ where $A \subseteq X$. We write $\mathrm{dom}(f) = A$ and denote partial maps as $f : X \rightharpoonup X$. They are equivalent to functions of the form $f : X \to X \cup \{\perp\}$. The next domain is of central importance in recursion theory:

Example 7. The set of partial mappings on the naturals

$$[\mathbb{N} \rightharpoonup \mathbb{N}] = \{ \, f \mid f : \mathbb{N} \rightharpoonup \mathbb{N} \text{ is a partial map}\}$$

ordered by extension

$$f \sqsubseteq g \Leftrightarrow \mathrm{dom}(f) \subseteq \mathrm{dom}(g) \;\&\; f = g \text{ on } \mathrm{dom}(f)$$

is an ω-algebraic dcpo:

- For directed set $S \subseteq [\mathbb{N} \rightharpoonup \mathbb{N}]$, $\bigsqcup S = \bigcup S$,
- $f \ll g \Leftrightarrow f \sqsubseteq g \;\&\; \mathrm{dom}(f)$ is finite, and
- $\{f \in [\mathbb{N} \rightharpoonup \mathbb{N}] : \mathrm{dom}(f) \text{ finite}\}$ is a countable basis for $[\mathbb{N} \rightharpoonup \mathbb{N}]$.

Algebraic domains may seem ‘discrete’ in some sense, or at least more discrete than domains that are continuous but not algebraic, such as $\mathbf{I}\mathbb{R}$. However, the reader should not go around believing that the continuous and the discrete are irreversibly divided – they are not. In domain theory and measurement, it is often possible to take a unified view of the two. To partially illustrate this point, let us now consider a *continuous extension* of the finite powerset $\mathcal{P}\{1, \ldots, n\}$ to the set of finite probability distributions

$$\Delta^n := \left\{ x \in [0,1]^n : \sum x_i = 1 \right\}.$$

Each set $A \in \mathcal{P}\{1, \ldots, n\}$ has a characteristic map $\chi_A : \{1, \ldots, n\} \to \{0,1\}$ defined by

$$\chi_A(i) := \begin{cases} 1 & \text{if } i \in A; \\ 0 & \text{otherwise.} \end{cases}$$

for which we have

$$A \supseteq B \iff \chi_A \geq \chi_B$$

where $\geq$ is the pointwise order on functions of type $\{1, \ldots, n\} \to \{0, 1\}$ and $1 \geq 0$. But each $A \in \mathcal{P}\{1, \ldots, n\} \setminus \{\emptyset\}$ corresponds to a canonical $x \in \Delta^n$ given by

$$x_i := \begin{cases} x^+ & \text{if } i \in A; \\ 0 & \text{otherwise,} \end{cases}$$

where x^+ refers to the largest probability in x. Thus, we can think of *any* $x \in \Delta^n$ as having a characteristic function $\chi_x : \{1, \ldots, n\} \to [0, 1]$ given by

$$\chi_x(i) := \begin{cases} 1 & \text{if } x_i = x^+; \\ x_i & \text{otherwise.} \end{cases}$$

Example 8. The set of *classical states*

$$\Delta^n := \left\{ x \in [0, 1]^n : \sum x_i = 1 \right\}$$

is a continuous dcpo in its *implicative order* [23]

$$x \sqsubseteq y \equiv \chi_x \geq \chi_y.$$

The implicative order can also be characterized as

$$x \sqsubseteq y \equiv (\forall i)\ x_i < y_i \Rightarrow x_i = x^+$$

where again x^+ refers to the largest probability in x. Thus, only a maximum probability is allowed to increase as we move up in the information order on Δ^n. If the maximum probability refers to a solution of a problem, then moving up in this order ensures that we are getting closer to the answer.

Example 9. The set of *decreasing* classical states

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

with the *majorization relation* $\leq$ given by

$$x \leq y \equiv (\forall k < n)\ \sum_{i=1}^{k} x_i \leq \sum_{i=1}^{k} y_i$$

is a continuous dcpo $(\Lambda^n, \leq)$. If the implicative $\sqsubseteq$ order is restricted to Λ^n, then we have $(\Lambda^n, \sqsubseteq) \subseteq (\Lambda^n, \leq)$, and this inclusion is strict.

A *list* over S is a function $x : \{1, ..., n\} \to S$ for $n \geq 0$ and the set of all such x is denoted $[S]$. The *length* of a list x is $|\mathrm{dom}\, x|$. A list x can be written as $[x(1), ..., x(n)]$, where the *empty list* (the list of length 0) is written $[\,]$. We can also write lists as $a :: x$, where $a \in S$ is the *first element* of the list $a :: x$

and $x \in [S]$ is the *rest* of the list $a :: x$. For example, the list $[1,2,3]$ can be written $1 :: [2,3]$.

A set $K \subseteq \mathbb{N}$ is *convex* if $a, b \in K$ & $a \leq x \leq b \Rightarrow x \in K$. Given a finite convex set $K \subseteq \mathbb{N}$, the map $\text{scale}(K) : \{1, ..., |K|\} \to K$ given by

$$\text{scale}(K)(i) = \min K + i - 1$$

relabels the elements of K so that they begin with one.

Example 10. The domain of finite lists. The set of finite lists $[S]$ with $\sqsubseteq$ given by reverse convex containment

$$x \sqsubseteq y \equiv (\exists \text{ convex } K \subseteq \{1, \ldots, \text{length}(y)\})\ y \circ \text{scale}(K) = x.$$

is an algebraic dcpo in which all elements are compact. If $x \sqsubseteq y$, we say that y is a *sublist* of x.

For instance, if $L = [1,2,3,4,5,6]$, then $[1,2,3], [4,5,6], [3,4,5], [2,3,4], [3,4], [5]$ and $[\,]$ are all sublists of L, while $[1,4,5,6], [1,3]$ and $[2,4]$ are *not* sublists of L. The set $[S]$ is also called *the free monoid* over S.

Example 11. Products of domains. If D and E are dcpo's then

$$D \times E := \{(d, e) : d \in D \ \&\ e \in E\}$$

is a dcpo in the pointwise order

$$(x_1, y_1) \sqsubseteq (x_2, y_2) \equiv x_1 \sqsubseteq x_2 \ \&\ y_1 \sqsubseteq y_2.$$

If D and E are both continuous, then so is $D \times E$, where

$$(x_1, y_1) \ll (x_2, y_2) \equiv x_1 \ll x_2 \ \&\ y_1 \ll y_2.$$

Having discussed the information order, let us turn now to the question of *information content.*

2.3 Measurement

From Section 2.1, a measurement $\mu : D \to [0, \infty)$ should satisfy:

1. For all $x, y \in D$, $x \sqsubseteq y \Rightarrow \mu x \geq \mu y$, and
2. If (x_n) is an increasing sequence in D, then

$$\mu\left(\bigsqcup_{n \geq 1} x_n\right) = \lim_{n \to \infty} \mu x_n.$$

On all the domains that we will work with, a mapping will have these two properties exactly when it is *Scott continuous.*

Definition 4. For a subset $X \subseteq D$ of a dcpo D, define

$$\uparrow\!X := \bigcup_{x \in X} \uparrow\!x \quad \& \quad \downarrow\!X := \bigcup_{x \in X} \downarrow\!x$$

A subset $U \subseteq D$ of a dcpo D is *Scott open* when it is an *upper set* $U = \uparrow\!U$ that is *inaccessible by directed suprema*:

$$\bigsqcup S \in U \Rightarrow S \cap U \neq \emptyset$$

for all directed $S \subseteq D$.

The Scott open sets on a dcpo form a topology. A subset $C \subseteq D$ is *Scott closed* when it is a *lower set* $C = \downarrow\!C$ that contains the supremum of every directed set it contains. Of particular importance for us is that the Scott topology on a continuous dcpo has the collection $\{\twoheaduparrow\!x : x \in D\}$ as a basis. That is, it is a topology determined by approximation.

Example 12. A basic Scott open set in $\mathbf{I}[0,1]$ is

$$\twoheaduparrow[a,b] = \{x \in \mathbf{I}[0,1] : x \subseteq \mathrm{int}([a,b])\}.$$

In the domain of binary channels $\mathbb{N} = \{(a,b) : a \geq b \ \& \ a,b \in [0,1]\}$, drawn with a on the x-axis and b on the y-axis, such a set forms a right triangle whose hypotenuse lies along the diagonal, but whose other two sides are removed.

Definition 5. A function $f : D \to E$ between dcpo's is *Scott continuous* if the inverse image of a Scott open set in E is Scott open in D.

Scott continuity can be characterized order theoretically [1]:

Theorem 1. *A function $f : D \to E$ is Scott continuous iff f is monotone,*

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

and preserves directed suprema:

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

for all directed $S \subseteq D$.

Thus, on the *very* reasonable assumption that the domains which arise in practice always allow us to replace directed sets with increasing sequences, a measurement should at least be a Scott continuous function $\mu : D \to [0,\infty)^*$, where $[0,\infty)^*$ is the domain of nonnegative reals in its dual order:

$$x \sqsubseteq y \equiv y \leq x$$

and $\leq$ refers to the usual way of ordering real numbers. But there is more to the story of content than just continuity.

Imagine that we would like to compute an ideal element $x \in D$ in some domain D that cannot be computed exactly. How accurately would we like to calculate it? We wish to calculate it to within an accuracy of $a \ll x$. Now we proceed to calculate, using some process to determine a sequence of values $x_1, \ldots, x_n \ldots$ each x_i containing information about x, that is, $x_i \sqsubseteq x$. When do we stop? We stop when some *measure* of information content μ says that we are 'close enough' to the answer x. How does μ say this?

It tells us that *if* x_n contains information about x and *if* x and x_n are close enough in information content, *then* we have succeeded in calculating x to within the desired accuracy. That is, we have found an x_n such that $a \ll x_n$. In symbols,

$$(\exists \varepsilon > 0)(\forall n)(x_n \sqsubseteq x \;\&\; |\mu x - \mu x_n| < \varepsilon \Rightarrow a \ll x_n)$$

Now the thing to realize is that other computations may take *other* paths (x_n) to x and that we may also be interested in *other* levels of accuracy a. Since we want μ to guarantee accuracy for these processes too, we want μ to satisfy

$$(\forall a \ll x)(\exists \varepsilon > 0)(\forall y \in D)(y \sqsubseteq x \;\&\; |\mu x - \mu y| < \varepsilon \Rightarrow a \ll y)$$

If μ can provide this for the element $x \in D$, then μ must be measuring the *information content* of x. If the last statement holds, then it also holds when we can quantify over *all* Scott open sets U since sets of the form $\mathord{\uparrow\!\!\uparrow} a$ are a basis for the Scott topology at x. For a dcpo D, we arrive at the following:

Definition 6. A Scott continuous $\mu : D \to [0, \infty)^*$ is said to *measure the content* of $x \in D$ if for all Scott open sets $U \subseteq D$,

$$x \in U \Rightarrow (\exists \varepsilon > 0)\, x \in \mu_\varepsilon(x) \subseteq U$$

where

$$\mu_\varepsilon(x) := \{y \in D : y \sqsubseteq x \;\&\; |\mu x - \mu y| < \varepsilon\}$$

are called the *ε-approximations* of x.

We often refer to μ as simply 'measuring' $x \in D$ or as measuring $X \subseteq D$ when it measures each element of X. Minimally, a *measurement* should measure the content of its kernel:

Definition 7. A *measurement* $\mu : D \to [0, \infty)^*$ is a Scott continuous map that measures the content of $\ker(\mu) := \{x \in D : \mu x = 0\}$.

The order on a domain D defines a clear sense in which one object has 'more information' than another: a *qualitative* view of information content. The definition of measurement attempts to identify those monotone mappings μ which offer a *quantitative* measure of information content in the sense specified by the order. The essential point in the definition of measurement is that μ measure content in a manner that is consistent with the particular view offered by the order. There are plenty of monotone mappings that are not measurements – and while some of them may measure information content in *some other sense*, each sense must first be specified by a different information order. The definition of measurement is then a minimal test that a function μ must pass if we are to regard it as providing a measure of information content.

Lemma 1. *Let $\mu : D \to [0,\infty)^*$ be a measurement.*

(i) *If $x \in \ker(\mu)$, then $x \in \max(D) = \{x \in D : \uparrow\!x = \{x\}\}$.*
(ii) *If μ measures the content of $y \in D$, then*

$$(\forall x \in D)\ x \sqsubseteq y \ \&\ \mu x = \mu y \Rightarrow x = y.$$

These results say (i) elements with no uncertainty are maximal in the information order and (ii) comparable elements with the same information content are equal. The converse of (i) is not true and there are many important cases (see Section 3.4 for instance) where the applicability of measurement is greatly heightened by the fact that $\ker\mu$ need not consist of *all* maximal elements.

Example 13. Canonical measurements.

(i) $(\mathbf{I}\mathbb{R}, \mu)$ the interval domain with the length measurement $\mu[a,b] = b - a$.
(ii) $(\mathcal{P}\omega, |\cdot|)$ the powerset of the naturals with $|\cdot| : \mathcal{P}\omega \to [0,\infty)^*$ given by

$$|x| = 1 - \sum_{n \in x} \frac{1}{2^{n+1}}.$$

(iii) $([\mathbb{N} \rightharpoonup \mathbb{N}], \mu)$ the partial functions on the naturals with

$$\mu f = |\mathrm{dom}(f)|$$

where $|\cdot|$ is the previous measurement on $\mathcal{P}\omega$.
(iv) $(\Sigma^\infty, 1/2^{|\cdot|})$ the binary strings where $|\cdot| : \Sigma^\infty \to [0,\infty]$ is the length of a string.
(v) $(\mathbf{U}X, \mathrm{diam})$ the upper space of a locally compact metric space (X,d) with

$$\mathrm{diam}\, K = \sup\{d(x,y) : x, y \in K\}.$$

(vi) $(\mathbf{B}X, \pi)$ the formal ball model of a complete metric space (X,d) with

$$\pi(x,r) = r$$

(vii) (Δ^n, μ) the classical states in their implicative order with $\mu x = 1 - x^+$. Shannon entropy

$$H(x) = -\sum_{i=1}^{n} x_i \log_2(x_i)$$

is also a measurement on Δ^n.

(viii) $(\mathbb{N}, c)$ the nonnegative binary channels with capacity from information theory (Shannon)

$$c(a,b) = \log_2 \left(2^{\frac{\bar{a}H(b) - \bar{b}H(a)}{a-b}} + 2^{\frac{bH(a) - aH(b)}{a-b}} \right)$$

where $c(a,a) := 0$ and $H(x) = -x\log_2(x) - (1-x)\log_2(1-x)$ is the binary entropy.

(ix) $([S], \text{length})$ lists with length as a measurement

(x) Products: if (D,μ) and (E,λ), are domains with measurements, then $D \times E$ is a domain with $\max\{\mu,\lambda\}$ and $\mu+\lambda$ as measurements[1].

In each case, we have $\ker\mu = \max(D)$.

We will see other examples in Section 4, including the domains of analytic mappings, quantum states and spacetime intervals. The reader who is impatient to find out what one does with a measurement can skip ahead to any of the other sections as long as they promise to eventually return. The reader interested in understanding the ideas should continue reading.

The view of information content taken in the study of measurement is that of a structural relationship between two classes of objects which, generally speaking, arises when one class may be viewed as a simplification of the other. The process by which a member of one class is simplified and thereby 'reduced' to an element of the other is what we mean by 'the measurement process' in domain theory [16]. One of the classes may well be a subset of real numbers, but the 'structural relationship' underlying content should not be forgotten. Here is the definition of measurement in this more general case:

Definition 8. A Scott continuous map $\mu : D \to E$ between dcpo's is said to *measure the content of* $x \in D$ if

$$x \in U \Rightarrow (\exists \varepsilon \in \sigma_E)\, x \in \mu_\varepsilon(x) \subseteq U,$$

whenever $U \in \sigma_D$ is Scott open and

$$\mu_\varepsilon(x) := \mu^{-1}(\varepsilon) \cap \downarrow\! x$$

are the elements ε close to x in content. The map μ *measures* X if it measures the content of each $x \in X$.

Definition 9. A *measurement* is a Scott continuous map $\mu : D \to E$ between dcpo's that measures $\ker\mu := \{x \in D : \mu x \in \max(E)\}$.

In the case $E = [0,\infty)^*$, the new definition of "measures the content of x" is equivalent to the one given earlier, so we reserve the right to denote the set $\mu_{[0,\varepsilon)}(x)$ by $\mu_\varepsilon(x)$ in contrast to how we first defined $\mu_\varepsilon(x)$, though

[1] In *principle*, it is possible to measure the dcpo of Scott continuous maps $[D \to E]$. In practice, though, the question is how to do so *simply*. See [21, 43] for more.

we will always be clear about how we are using this notation. In addition, Lemma 1 remains valid in this more general case; the 'reflective' nature of measurement is covered in more detail in [15]. In addition, with the more abstract formulation of measurement, it becomes clear that measurements compose. That, for example, is why it was easy to measure the domain $[\mathbb{N} \rightharpoonup \mathbb{N}]$ of partial functions in the last example.

2.4 Distance, content and topology

Why should there be any relation between *topology* and *information content*? To answer this, we have to remember that we are not just talking about any topology, but rather, the Scott topology, which as we have seen is the topology of approximation. Second, we have to recall the subtle relation between information content and the desire to obtain accurate approximations of ideal elements discussed in the last section.

As it turns out, one way to think of a measurement is essentially as being the informatic analogue of 'metric' for domain theory. There are several senses in which this is true. Let us consider one by returning to the elements ε close to $x \in D$, abbreviated to

$$\mu_\varepsilon(x) := \mu_{[0,\varepsilon)}(x) = \{y \in D : y \sqsubseteq x \ \&\ \mu y < \varepsilon\},$$

for $\varepsilon > 0$.

Theorem 2. *Let D be a continuous dcpo. If $\mu : D \to [0,\infty)^*$ measures $X \subseteq D$, then*

$$\{\uparrow\!\mu_\varepsilon(x) \cap X : x \in X, \varepsilon > 0\}$$

is a basis for the relative Scott topology on X.

Thus, in the presence of a measurement, we can understand the Scott topology as being derived from ε-approximations of points, similar to the way the topology of a metric space is specified.

To further develop the analogy between metric and measurement hinted at in the last result, suppose that a continuous dcpo D has the property that for any $x, y \in D$ there is $z \in D$ with $z \sqsubseteq x, y$. Notice that in Example 13, domains (i)–(ix) all have this property. If we encounter a continuous dcpo that does not have this property, we can always adjoin a bottom element $\bot$, and scale the measurement so that $\mu\bot = 1$. See chapter five of [15] for more. Then we can define $d : D^2 \to [0,\infty)^*$ given by

$$d(x,y) = \inf\{\mu z : z \ll x, y\} = \inf\{\mu z : z \sqsubseteq x, y\}$$

Because μ is monotone, d is Scott continuous. Because μ is Scott continuous, we have $d(x,y) = \mu x$ when $x \sqsubseteq y$. The distance function d associated to μ is sometimes denoted $d(\mu)$.

Definition 10. *For a monotone map $\mu : D \to [0,\infty)^*$ on a continuous dcpo D with $d = d(\mu)$ defined,*

$$B_\varepsilon(x) := \{y \in D : d(x,y) < \varepsilon\}$$

for all $x \in D$, $\varepsilon > 0$.

Happily, distance and content are related as follows.

Theorem 3. *If $\mu : D \to [0,\infty)^*$ is Scott continuous on a continuous dcpo D with $d = d(\mu)$ defined, then*

$$B_\varepsilon(x) = \uparrow\!\mu_\varepsilon(x),$$

for each $x \in D$ and $\varepsilon > 0$. Consequently,

$$\{B_\varepsilon(x) \cap X : x \in X, \varepsilon > 0\}$$

is a basis for the relative Scott topology on X whenever μ measures X.

Example 14. For $(\mathbf{I}\mathbb{R}, \mu)$,

$$d([a],[b]) = |a-b|,$$

for all $a, b \in \mathbb{R}$. Because d is the Euclidean metric on $\mathbb{R}$, we can conclude that $\max(\mathbf{I}\mathbb{R})$ in its relative Scott topology is homeomorphic to $\mathbb{R}$.

The last example is also true for $\mathbf{I}[0,1]$. But now something interesting happens, because Theorem 3 says that *any measurement* on $\mathbf{I}[0,1]$ induces the Euclidean topology on its kernel. Recalling that the *capacity* $c : \mathbb{N} \to [0,1]^*$ of a binary channel

$$c(a,b) = \log_2\left(2^{\frac{\bar{a}H(b)-\bar{b}H(a)}{a-b}} + 2^{\frac{bH(a)-aH(b)}{a-b}}\right)$$

is a measurement on the domain of binary channels $\mathbb{N} \simeq \mathbf{I}[0,1]$ from Example 13(viii), its associated distance function on $\ker(c) = \max(\mathbb{N})$ is

$$\rho([a],[b]) = c(a,b) = c(b,a)$$

Then, just like Euclidean distance, capacity $c : [0,1]^2 \to [0,1]$ also has the following three properties:

(i) $c(a,b) = c(b,a)$,
(ii) $c(a,b) = 0$ iff $a = b$,
(iii) The sets $\{y \in [0,1] : c(x,y) < \varepsilon\}$ for $\varepsilon > 0$ form a basis for the Euclidean topology on $[0,1]$.

Capacity does not satisfy the triangle inequality, so it is not a priori obvious that the sets in (iii) form a basis for any topology, let alone the Euclidean topology. Let us state this another way: the topology of certain spaces can be derived from a notion of distance that is defined in terms of the amount of information that can be transmitted between two points [29].

Now we consider another sense in which measurements are the domain theoretic counterpart to metrics.

Definition 11. A measurement $\mu : D \to [0,\infty)^*$ on a continuous dcpo D satisfies *the triangle inequality* if for all consistent pairs $x, y \in D$, there is an element $z \sqsubseteq x, y$ such that $\mu z \leq \mu x + \mu y$.

When a measurement satisfies the triangle inequality, its corresponding notion of distance is a metric on the set of elements with measure zero.

Theorem 4. *Let (D,μ) be a domain with a measurement satisfying the triangle inequality. Then $d(\mu) : \ker(\mu) \times \ker(\mu) \to [0,\infty)$ is a metric which yields the relative Scott topology on $\ker(\mu)$.*

For instance, many of the measurements in Example 13 satisfy the triangle inequality, including (i)–(v) and (xi). More generally, the class of *Lebesgue measurements*, discussed in Section 3.4, allows one to conclude that $\ker(\mu)$ is metrizable. In fact, in most cases, we can construct a metric from μ, though the construction is more involved. One such case is when there is an element $z \sqsubseteq x, y$ with $\mu z \leq 2 \cdot \max\{\mu x, \mu y\}$, see chapter five of [15] for more on this.

The relation between measurement and topology does not end with the observation that they are like metrics. It turns out that measuring a domain is *equivalent* to being able to generate a certain topology.

Definition 12. The μ *topology* on a continuous dcpo D has

$$\{\Uparrow a \cap \downarrow x : a, x \in D\}$$

as a basis.

Unexpectedly, the μ topology is *always* zero dimensional and Hausdorff.

Theorem 5. *Let D be a continuous dcpo. A Scott continuous $\mu : D \to [0,\infty)^*$ measures D iff $\{\mu_\varepsilon(x) : x \in D \ \&\ \varepsilon > 0\}$ is a basis for the μ topology.*

In the above result, $\mu_\varepsilon(x) = \mu_{[0,\varepsilon)}$ is defined as it was earlier in this section. We pause for a moment now to look at a few of the things one does with domains and measurements.

3 Fixed points

A *least element* in a dcpo D is an element $\perp$ such that $\perp \sqsubseteq x$ for all $x \in D$. The first theorem I ever heard about in domain theory is:

Theorem 6. *Let D be a dcpo with a least element $\perp$. If $f : D \to D$ is Scott continuous, it has a least fixed point given by*

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

A useful corollary is that f has a least fixed point on $\uparrow\!x$ if $x \sqsubseteq f(x)$.

Exercise: Prove that split_f from the introduction is *not* Scott continuous by showing that it is *not* monotone. (*Hint*: Cheat, by reading this section).

Major references: [15]

3.1 Fixed points of nonmonotonic mappings

Ordinarily, this discussion would be deferred to the section "forms of process evolution" but we include it here so that the reader gets some quick examples of what one does with measurement.

Definition 13. A *splitting* on a dcpo D is a function $s : D \to D$ with $x \sqsubseteq s(x)$ for all $x \in D$.

Theorem 7. *Let D be a dcpo with a measurement μ that measures D. If $I \subseteq D$ is closed under directed suprema and $s : I \to I$ is a splitting whose measure*

$$\mu \circ s : I \to [0, \infty)^*$$

is Scott continuous, then

$$(\forall x \in I) \ \bigsqcup_{n \geq 0} s^n(x) \ \textit{is a fixed point of } s.$$

Moreover, the set of fixed points $\mathrm{fix}(s) = \{x \in I : s(x) = x\}$ *is a dcpo.*

In applications, a slightly weaker formulation can be useful: if for every increasing sequence (x_n) in I we have

$$\mu s \left(\bigsqcup x_n \right) = \lim_{n \to \infty} \mu s(x_n),$$

then

$$\bigsqcup_{n \geq 0} s^n(x) \in \mathrm{fix}(s),$$

for every $x \in I$. In addition, $\mathrm{fix}(s) = I \cap \ker(\mu)$ iff $\mu s(x) < \mu x$ for all $x \in I$ with $\mu x > 0$. The point being: we do not need to check that $\mu \circ s$ is monotone in order to establish the existence of fixed points.

Example 15. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous map on the real line. Denote by $C(f)$ the subset of $\mathbf{I}\mathbb{R}$ where f changes sign, that is,

$$C(f) = \{[a,b] : f(a) \cdot f(b) \leq 0\}.$$

The continuity of f ensures that this set is closed under directed suprema, and the mapping

$$\mathrm{split}_f : C(f) \to C(f)$$

given by

$$\mathrm{split}_f[a,b] = \begin{cases} \mathrm{left}[a,b] & \text{if } \mathrm{left}[a,b] \in C(f); \\ \mathrm{right}[a,b] & \text{otherwise,} \end{cases}$$

is a splitting where $\mathrm{left}[a,b] = [a,(a+b)/2]$ and $\mathrm{right}[a,b] = [(a+b)/2,b]$. The measure of this mapping

$$\mu\,\mathrm{split}_f[a,b] = \frac{\mu[a,b]}{2}$$

is Scott continuous, so Theorem 7 implies that

$$\bigsqcup_{n \geq 0} \mathrm{split}_f^n[a,b] \in \mathrm{fix}(\mathrm{split}_f).$$

However, $\mathrm{fix}(\mathrm{split}_f) = \{[r] : f(r) = 0\}$, which means that iterating split_f is a scheme for calculating a solution of the equation $f(x) = 0$. This numerical technique is called *the bisection method.*

Proposition 1. *For a continuous selfmap $f : \mathbb{R} \to \mathbb{R}$ which has at least one zero, the following are equivalent:*

(i) *The map* split_f *is monotone.*
(ii) *The map f has a unique zero r and*

$$C(f) = \{[a,r] : a \leq r\} \cup \{[r,b] : r \leq b\}.$$

That is, if split_f is monotone, then in order to calculate the solution r of $f(x) = 0$ using the bisection method, we must first *know* the solution r.

Example 16. A function $f : [a,b] \to \mathbb{R}$ is *unimodal* if it has a maximum value assumed at a unique point $x^* \in [a,b]$ such that

(i) f is strictly increasing on $[a,x^*]$, and
(ii) f is strictly decreasing on $[x^*,b]$.

Unimodal functions have the important property that

$$x_1 < x_2 \Rightarrow \begin{cases} x_1 \leq x^* \leq b \text{ if } f(x_1) < f(x_2), \\ a \leq x^* \leq x_2 \text{ otherwise.} \end{cases}$$

This observation leads to an algorithm for computing x^*. For a unimodal map $f : [a,b] \to \mathbb{R}$ with maximizer $x^\star \in [a,b]$ and a constant $1/2 < r < 1$, define a dcpo by

$$I_{x^*} = \{\bar{x} \in \mathbf{I}\mathbb{R} : [a,b] \sqsubseteq \bar{x} \sqsubseteq [x^*]\},$$

and a splitting by

$$\max_f : I_{x^*} \to I_{x^*}$$

$$\max_f[a,b] = \begin{cases} [l(a,b), b] & \text{if } f(l(a,b)) < f(r(a,b)); \\ [a, r(a,b)] & \text{otherwise,} \end{cases}$$

where $l(a,b) = (b-a)(1-r)+a$ and $r(a,b) = (b-a)r+a$. The measure of $\max_f$ is Scott continuous since $\mu \max_f(\bar{x}) = r \cdot \mu(\bar{x})$, for all $\bar{x} \in I_{x^*}$. By Theorem 7,

$$\bigsqcup_{n\geq 0} \max_f^n(\bar{x}) \in \text{fix}(\max_f),$$

for any $\bar{x} \in I_{x^*}$. However, any fixed point of $\max_f$ has measure zero, and the only element of I_{x^*} with measure zero is $[x^*]$. Thus, $\bigsqcup \max_f^n[a,b] = [x^*]$, which means that iterating $\max_f$ yields a method for calculating x^*. This technique is called the *r-section search.*

Finally, observe that $\max_f$ is not monotone. Let $\text{-}1 < \alpha < 1$ and $f(x) = 1-x^2$. The function f is unimodal on any compact interval. Since $\max_f[\text{-}1,1] = [\text{-}1, 2r-1]$, we see that

$$\begin{aligned}\max_f[\text{-}1,1] \sqsubseteq \max_f[\alpha,1] &\Rightarrow 1 \leq 2r-1 \text{ or } r(\alpha,1) \leq 2r-1 \\ &\Rightarrow 1 \leq r \text{ or } \alpha+1 \leq r(\alpha+1) \\ &\Rightarrow r \geq 1,\end{aligned}$$

which contradicts $r < 1$. Thus, for no value of r is the algorithm monotone.

The previous examples make it clear that there are natural and important examples of processes on domains that are fundamentally nonmonotonic but which nevertheless have fixed points whose existence can be easily established by measurement based results. Moreover, the previous fixed point theorem is a strict generalization of the usual fixed point theorem in domain theory:

Example 17. If $f : D \to D$ is a Scott continuous map on a dcpo D with a measurement μ that measures D, then we consider its restriction to the set of points where it improves

$$I(f) = \{x \in D : x \sqsubseteq f(x)\}.$$

This yields a splitting $f : I(f) \to I(f)$ on a dcpo with continuous measure. By Theorem 7,

$$(\forall x \in I(f)) \bigsqcup_{n\geq 0} f^n(x) \text{ is a fixed point of } f.$$

For instance, if D is ω-continuous with basis $\{b_n : n \in \mathbb{N}\}$, then

$$\mu x = |\{n : b_n \ll x\}|$$

defines such a measurement. Notice, however, that with this construction we normally have $\ker \mu = \emptyset$.

3.2 Numerical methods

Numerical methods provide an interesting application of domains and measurements. The two examples in the last section, the bisection and the golden section search, really only scratch the surface of what is possible in this regard. So in this section, we take a closer look.

A topological question: what the **** are we computing?

By Theorem 2, a measurement μ allows one to derive the Scott topology on $\ker(\mu)$. This fundamental fact ensures that what *appears* to be computation *actually is* computation.

Example 18. Recall the bisection method $\mathrm{split}_f : C(f) \to C(f)$ from Example 15. By Theorem 7,

$$\bigsqcup_{n\geq 0} \mathrm{split}_f^n(x) \in \mathrm{fix}(\mathrm{split}_f),$$

for all $x \in C(f)$. But $\mathrm{fix}(\mathrm{split}_f) = \{[r] : f(r) = 0\}$, which means that iterating split_f is a scheme for calculating a zero of f. Right?

Well, almost. Let's take a closer look at things. In the zero finding problem, the desired result is a *number* that approximates the zero r, not an interval. In practice, we calculate a small enough interval x, and then choose a point within it as an approximation of r. The true reason that split_f is an algorithm for computing r is that if we begin with any $x \in C(f)$, and then choose any sequence $x_n \in \mathrm{split}_f^n(x)$, we *always* have

$$|x_n - r| \leq \mu\, \mathrm{split}_f^n(x) \leq \frac{\mu x}{2^n},$$

and hence $x_n \to r$ in the *usual topology* on the real line.

Then what we need to know is that computation on a domain actually corresponds to computation in reality. For the splittings of Prop. 7, the following result confirms exactly this.

Proposition 2. *Let D be a continuous dcpo with a map μ that measures D, $I \subseteq D$ a set closed under suprema of increasing sequences and $s : I \to I$ a splitting with $\mu s \leq c \cdot \mu$ for a constant $0 \leq c < 1$. Then for all $x \in I$, if $x_n \in \uparrow s^n(x) \cap \ker \mu$, we have*

$$x_n \to \bigsqcup_{n \geq 0} s^n(x) \in \operatorname{fix}(s) \subseteq \ker \mu,$$

in the relative Scott topology on $\ker \mu$.

For instance, in the case of the interval domain $\mathbf{I}\mathbb{R}$, we have

$$\ker \mu = \max(\mathbf{I}\mathbb{R}) = \{[x] : x \in \mathbb{R}\} \simeq \mathbb{R}$$

where the homemorphism is between the relative Scott topology on $\ker \mu$ and the usual topology on the real line. Thus, to say that $\bigsqcup_{n \geq 0} \mathrm{split}_f^n(x)$ computes a zero of f means exactly the same thing as it does in numerical analysis.

Numerical methods and the information order

Some numerical methods manipulate information in a manner that is fundamentally different than a *bracketing method*, such as the bisection, or a *one point method*, like Newton's method. Each way of manipulating information corresponds to a different information order.

Example 19. Let $D = [0, 1]$ be the unit interval in its usual order. Then

$$\mathcal{P}_C(D) = \{[a, b] : a, b \in [0, 1] \ \& \ a \leq b\}$$

is called *the convex powerdomain* over D and its order is given by

$$[a, b] \sqsubseteq [x, y] \Leftrightarrow a \leq x \ \& \ b \leq y.$$

We can measure this object by

$$\mu[a, b] = (1 - a) + (1 - b).$$

Note that $\ker \mu = \max(\mathcal{P}_C(D)) = \{[1]\}$.

The measurement above has a natural explanation [19].

One point methods

A one point method amounts to iterating a continuous $f : [a, b] \to [a, b]$ until we reach a fixed point, so it should come as no surprise that we can model them domain theoretically with a copy of $[a, b] \simeq [0, 1]$ in its usual order. However, there is another way. We can exploit the fact that

$$[0, 1] \simeq \{[x] : x \in [0, 1]\} \subseteq \mathcal{P}_C[0, 1].$$

This subset we name the *total reals* and for this reason we refer to the other elements of $\mathcal{P}_C[0, 1]$ as *partial reals.*

Example 20. Let f be concave increasing on $[a,b]$ with $f(a) < 0$ and $f(b) > 0$. Then consider the partial function $I_f : \mathcal{P}_C[a,r] \rightharpoonup \mathcal{P}_C[a,r]$ given by

$$[x] \mapsto [x - f(x)/f'(x)],$$

which is defined only on the subset of total reals. By Theorem 7,

$$\bigsqcup_{n\geq 0} I_f^n[x] \in \mathrm{fix}(I_f) = \{[r]\},$$

and so *Newton's method* converges for any initial guess $x \in \mathrm{dom}(I_f)$.

One of the standard reasons for avoiding Newton's method is that it requires the calculation of a derivative. A common method for overcoming this difficulty is to approximate the derivative by calculating a difference quotient using *two values* which simultaneously also serve to approximate the zero r. The most famous of the *interpolation methods,* as they are called, is probably the secant method.

An analysis of the secant method

One point methods are nothing more than iterating a function on some part of the real line, so domain theory is not *necessary* for describing them. However, with multi-point or *interpolation methods,* i.e., those which use more than one point to determine the next approximation in an iterative scheme, we arrive at our first example where pursuit of the uniformity ideal mandates a domain theoretic approach.

Example 21. The secant method. If we have a real valued function f, the following scheme is very useful for zero finding: choose *two* initial guesses x_0 and x_1 and then proceed inductively according to

$$x_{n+1} = x_n - f(x_n) \cdot \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}$$

for $n \geq 1$. The hope is that this sequence converges to a zero of f.

At each iteration of this algorithm, instead of one value, as with Newton's method, there are two values to be used in calculating the next approximation. We visualize it as a sequence of intervals:

$$[x_0, x_1] \to [x_1, x_2] \to [x_2, x_3] \to \cdots$$

The arrow indicates that we are moving up in the information order. These intervals are almost never nested. Happily, though, they often form an increasing sequence in the domain $\mathcal{P}_C[a,b]$ of partial reals.

If we have a function f, its derivative $df[x] = f'(x)$ can be extended from the total reals to the set of *all* partial reals $\mathcal{P}_C[a,b]$ by

$$df[x,y] = \frac{f(y)-f(x)}{y-x} \text{ if } y > x.$$

And just like that, we can model the secant method.

Theorem 8. *Let f be concave and increasing on $[a,b]$ with a zero $r \in (a,b)$. Then iterating the splitting $\sec_f : \mathcal{P}_C[a,r] \to \mathcal{P}_C[a,r]$ given by*

$$\sec_f[x,y] = \left[y, y - \frac{f(y)}{df[x,y]}\right]$$

is an algorithm for calcuating r. That is,

$$\bigsqcup_{n \geq 0} \sec_f^n(x) = [r],$$

for any $x \in \mathcal{P}_C[a,r]$.

For a total real $[x]$, we have $\sec_f[x] = [x, x - f(x)/f'(x)]$, which says that the secant method arises as the extension of a reversible formulation of Newton's method from the set of total reals to the set of *all* partial reals.

An interesting consequence here is that if we are able to compute the value of f' at just one $x \in [a,r)$, then the problem of generating two initial guesses for the secant method is eliminated: given such an $[x]$, we are then assured that we have enough information to calculate the partial real $\sec_f[x]$, and from there, Theorem 8 ensures that the iterates $\sec_f^n[x]$ converge to $[r]$.

So we have seen enough to find it plausible that the one point methods, the bracketing methods and the interpolation methods all have natural domain theoretic models and that the question of their correctness amounts *in all cases* to proving that some operator has a fixed point. Notice that numerical analysis only uses the fixed point approach for one point methods. This provides a nice uniform approach. But to really be able to believe in it, we need domain theory to teach us something new and significant about zero finding – perhaps something that someone other than a domain theorist would care about.

A new method for zero finding

The zero finding problem really is one of the great problems in the history of mathematics: given a real valued function f on an interval $[a,b]$, find a *zero* of f, that is, a number x such that $f(x) = 0$. Evariste Galois proved that one must resort to algorithms in solving this problem by showing that polynomials of degree five and higher have no solution by radicals, i.e., their zeroes are not in general expressible by a formula.

If one assumes nothing about f except continuity, then there are many senses in which *the bisection method* is the optimal algorithm for zero finding ([3][12]). However, for a class of Lipschitz mappings [4], the bisection method is no longer optimal. But Lipschitz mappings have derivatives almost everywhere [38]. In addition, the optimal algorithm makes use of the Lipschitz constant [4], which is a bound on its derivative. Another case in which bisection is not optimal is the class of convex mappings [8]. But there again, one finds that convex mappings are differentiable everywhere except on a countable set [38]. These two examples raise the following question: If we have a nontrivial class $\mathcal{C}$ of functions and a zero finding algorithm which is better than the bisection for the members of $\mathcal{C}$, must the functions in $\mathcal{C}$ possess some amount of differentiability? In short, is differentiability in some form necessary in order to beat the bisection method?

We are going to prove that the answer to this question is no. For the class of *Hölder continuous* mappings, which contains all of the well-known examples of nowhere differentiable functions, including those arising in the analysis of Brownian motion and fractals [7], we use domain theory and measurement to design and analyze a new algorithm for zero finding which is better than the bisection method at every iteration.

We will design the method for Hölder continuous functions which have a simple zero on a compact interval $[a, b]$.

Definition 14. A map $f : [a, b] \to \mathbb{R}$ is *Hölder continuous* if there are positive constants $c > 0$ and $\alpha > 0$ such that

$$|f(x) - f(y)| \leq c \cdot |x - y|^{\alpha}$$

for all $x, y \in [a, b]$.

Example 22. Weierstrass's function. The function introduced in 1872 by Weierstrass,

$$f(x) = \sum_{n=0}^{\infty} a^n \cos\left(b^n \pi x\right),$$

is nowhere differentiable for $0 < a < 1$ and b an odd integer with $ab > 1+3\pi/2$. It is Hölder continuous [46] with $\alpha = \log(1/a)/\log b$.

Definition 15. A function $f : [a, b] \to \mathbb{R}$ has a *simple zero* $r \in [a, b]$ if

$$\operatorname{sgn} f(x) = \operatorname{sgn}(x - r)$$

for $x \in [a, b]$, where $\operatorname{sgn}(x) = x/|x|$ for $x \neq 0$, and $\operatorname{sgn}(0) = 0$. Write

$$\square f := \{x \in \mathbf{I}\mathbb{R} : [a, b] \sqsubseteq x \sqsubseteq [r]\}$$

for the set of intervals where f changes sign.

Then a function has a simple zero r if it is positive to the right of r and negative to the left of r. We will make use of the following operators on $\mathbf{I}\mathbb{R}$:

Definition 16.

- $l : \mathbf{I}\mathbb{R} \to \mathbb{R} :: [a,b] \mapsto a$
- $m : \mathbf{I}\mathbb{R} \to \mathbb{R} :: [a,b] \mapsto (a+b)/2$
- $r : \mathbf{I}\mathbb{R} \to \mathbb{R} :: [a,b] \mapsto b$

These are abbreviated $l_x := l(x)$, $r_x := r(x)$ and $m_x := m(x)$.

For instance, if $f : [a,b] \to \mathbb{R}$ has a simple zero r on $[a,b]$, then the bisection $\text{split}_f : \Box f \to \Box f$ can be written compactly as

$$\text{split}_f(x) = \begin{cases} [l_x, m_x] & \text{if } f(m_x) > 0; \\ [m_x, r_x] & \text{otherwise.} \end{cases}$$

This formulation of the bisection will help us understand its relation to the new method:

Theorem 9. *Let $f : [a,b] \to \mathbb{R}$ be a Hölder continuous map with a simple zero r. Then iterating the splitting $s_f : \Box f \to \Box f$ given by*

$$s_f(x) = \begin{cases} \left[l_x, m_x - (f(m_x)/c)^{1/\alpha}\right] & \textit{if } f(m_x) > 0; \\ \\ \left[m_x + (|f(m_x)|/c)^{1/\alpha}, r_x\right] & \textit{otherwise;} \end{cases}$$

is an algorithm for computing r. That is,

$$\bigsqcup_{n \geq 0} s_f^n(x) = [r],$$

for all $x \in \Box f$. Thus, for all $x \in \Box f$, if $x_n \in s_f^n(x)$ for each n, then $x_n \to r$.

The method also easily extends to the case where $|fx - fy| \leq c \cdot g(|x-y|)$, for a left invertible $g : [0,\infty) \to [0,\infty)$ satisfying $g(0) = 0$.

A comparison with the bisection

If s_1 and s_2 are two algorithms, then a natural intuition stemming from domain theory is to say that s_2 is a better algorithm than s_1 if

$$s_1 \sqsubseteq s_2 \equiv (\forall x)\, s_1(x) \sqsubseteq s_2(x).$$

However, in the analysis of numerical techniques, one should not expect to be able to make absolute statements such as “Algorithm 1 is better than Algorithm 2 always and there is nothing more to be said.” For instance, sometimes the bisection method is better than Newton’s method, if the derivatives of a function are difficult (or impossible) to calculate, while an advantage of Newton’s method is its quadratic convergence when close enough to the root. Aside from the fact that our method requires one to determine the constants α and c – which is not necessarily a simple matter – we can in a lot of cases say that s_f is simply better than the bisection:

Proposition 3. *Let $f : [a, b] \to \mathbb{R}$ be a Hölder continuous map with a simple zero r. Then* $\mathrm{split}_f \sqsubseteq s_f$ *and for any $x \in \Box f$,*

$$\mu s_f(x) = \mu\,\mathrm{split}_f(x) - \left(\frac{|f(m_x)|}{c}\right)^{1/\alpha} \leq \mu\,\mathrm{split}_f(x),$$

with equality only in the unlikely event that $m_x = r$.

For instance, if r is a computable irrational and we begin with an input x having rational endpoints, then s_f is a strict improvement over the bisection.

Corollary 1. *Let $f : [a, b] \to \mathbb{R}$ be a Hölder continuous map with a simple irrational zero r. Then for any $x \in \Box f$ with rational endpoints $l_x, r_x \in \mathbb{Q}$,*

$$\mu s_f^n(x) < \mu\,\mathrm{split}_f^n(x),$$

for all iterations $n \geq 1$.

And in general we can see the same is true anytime the input interval does not contain r as its midpoint: once s_f gains an advantage over the bisection, it keeps this advantage forever. While the qualitative statement $\mathrm{split}_f \sqsubseteq s_f$ is certainly a strong one for numerical methods, when taken on its own, it leaves something to be desired: how are we to know the inputs where they are equal? Even if we know that $\mathrm{split}_f \sqsubseteq s_f$ and $\mathrm{split}_f \neq s_f$, they may only differ on a single input, which doesn't say very much.

But when we incorporate the quantitative as well, then the clarity of what we are saying improves greatly: split_f and s_f are equal iff their measures are iff one can magically choose an input whose midpoint is the zero (which amounts to guessing the answer). This provides a simple and clear example of the "extra something" that measurement adds to the standard order theoretic setting and illustrates how precise an analysis is possible when the qualitative and quantitative are united.

To summarize, domain theory and measurement provides a language for expressing zero finding algorithms which renders the verification process systematic and *uniform:* it enables us to turn the question of correctness into one about fixed points for *all* zero finding methods, whereas this is only normally achieved in numerical analysis for one point schemes like Newton's method. And because it also produced something new, it is okay to believe in it now if you want to.

3.3 Unique fixed points

So measurement can be used to generalize the Scott fixed point theorem so as to include important nonmonotonic processes. But it can also improve upon it for monotone maps as well, by giving a technique that guarantees *unique* fixed points.

Definition 17. Let D be a continuous dcpo with a measurement μ. A monotone map $f : D \to D$ is a *contraction* if there is a constant $c < 1$ with

$$\mu f(x) \leq c \cdot \mu x$$

for all $x \in D$.

Theorem 10. *Let D be a continuous dcpo with a measurement μ such that*

$$(\,\forall\, x, y \in \ker\mu\,)(\,\exists\, z \in D\,)\; z \sqsubseteq x, y.$$

If $f : D \to D$ is a contraction and there is a point $x \in D$ with $x \sqsubseteq f(x)$, then

$$x^\star = \bigsqcup_{n \geq 0} f^n(x) \in \max(D)$$

is the unique fixed point of f on D. Furthermore, $x^\star$ is an attractor in two different senses:

(i) *For all $x \in \ker\mu$, $f^n(x) \to x^\star$ in the Scott topology on $\ker\mu$, and*
(ii) *For all $x \sqsubseteq x^\star$, $\bigsqcup_{n\geq 0} f^n(x) = x^\star$, and this supremum is a limit in the Scott topology on D.*

When a domain has a least element, the last result is easier to state.

Corollary 2. *Let D be a domain with least element $\perp$ and measurement μ. If $f : D \to D$ is a contraction, then*

$$x^\star = \bigsqcup_{n \geq 0} f^n(\perp) \in \max D$$

is the unique fixed point of f on D. In addition, the other conclusions of Theorem 10 hold as well.

Example 23. Let $f : X \to X$ be a contraction on a complete metric space X with Lipschitz constant $c < 1$. The mapping $f : X \to X$ extends to a monotone map on the formal ball model $\bar{f} : \mathbf{B}X \to \mathbf{B}X$ given by

$$\bar{f}(x, r) = (fx, c \cdot r),$$

which satisfies

$$\pi \bar{f}(x, r) = c \cdot \pi(x, r),$$

where $\pi : \mathbf{B}X \to [0, \infty)^*$ is the standard measurement on $\mathbf{B}X$, $\pi(x, r) = r$. Now choose r so that $(x, r) \sqsubseteq \bar{f}(x, r)$. By Theorem 10, $\bar{f}$ has a unique attractor which implies that f does also because $X \simeq \ker\pi$.

We can also use the upper space $(\mathbf{U}X, \mathrm{diam})$ to prove the Banach contraction theorem for compact metric spaces by applying the technique of the last example. In [17], a domain theoretic result is given which generalizes the Banach contraction theorem. Next up: probably the most overused example of a Scott continuous map in domain theory. Here is something new about it:

Example 24. Consider the well-known functional

$$\phi : [\mathbb{N} \rightharpoonup \mathbb{N}] \to [\mathbb{N} \rightharpoonup \mathbb{N}]$$

$$\phi(f)(k) = \begin{cases} 1 & \text{if } k = 0, \\ kf(k-1) & \text{if } k \geq 1 \text{ \& } k-1 \in \operatorname{dom} f. \end{cases}$$

which is easily seen to be monotone. Applying $\mu : [\mathbb{N} \rightharpoonup \mathbb{N}] \to [0, \infty)^*$, we compute

$$\begin{aligned} \mu\phi(f) &= |\operatorname{dom}(\phi(f))| \\ &= 1 - \sum_{k \in \operatorname{dom}(\phi(f))} \frac{1}{2^{k+1}} \\ &= 1 - \left(\frac{1}{2^{0+1}} + \sum_{k-1 \in \operatorname{dom}(f)} \frac{1}{2^{k+1}} \right) \\ &= 1 - \left(\frac{1}{2} + \sum_{k \in \operatorname{dom}(f)} \frac{1}{2^{k+2}} \right) \\ &= \frac{1}{2} \left(1 - \sum_{k \in \operatorname{dom}(f)} \frac{1}{2^{k+1}} \right) \\ &= \frac{\mu f}{2} \end{aligned}$$

which means ϕ is a contraction on the domain $[\mathbb{N} \rightharpoonup \mathbb{N}]$. By the contraction principle,

$$\bigsqcup_{n \in \mathbb{N}} \phi^n(\perp) = \mathrm{fac}$$

is the unique fixed point of ϕ on $[\mathbb{N} \rightharpoonup \mathbb{N}]$, where $\perp$ is the function defined nowhere.

3.4 Fractals

We now consider certain nontrivial examples of contractions and some of their fixed points: *fractals*. By induction, a continuous map $\mu : D \to [0, \infty)^*$ is a measurement iff for all *finite* $F \subseteq \ker \mu$ and all open sets $U \subseteq D$,

$$F \subseteq U \Rightarrow (\exists\, \varepsilon > 0)(\forall x \in F)\, \mu_\varepsilon(x) \subseteq U.$$

If we require this to hold, not only for finite sets F, but for *all* compact sets K, we have exactly a *Lebesgue measurement.*

Definition 18. A *Lebesgue measurement* $\mu : D \to [0, \infty)^*$ is a continuous map such that for all compact sets $K \subseteq \ker \mu$ and all open sets $U \subseteq D$,

$$K \subseteq U \Rightarrow (\exists\, \varepsilon > 0)(\forall x \in K)\, \mu_\varepsilon(x) \subseteq U.$$

Not all measurements are Lebesgue (Example 5.3.2 of [15]). The existence of a Lebesgue measurement on a domain implies an important relationship between the Scott topology and the *Vietoris topology*:

Definition 19. The *Vietoris hyperspace* of a Hausdorff space X is the set of all nonempty compact subsets $\mathcal{P}_{com}(X)$ with the *Vietoris topology:* it has a basis given by all sets of the form

$$\sigma(U_1, \cdots, U_n) := \{K \in \mathcal{P}_{com}(X) : K \subseteq \bigcup_{i=1}^{n} U_i \text{ and } K \cap U_i \neq \emptyset, 1 \leq i \leq n\},$$

where U_i is a nonempty open subset of X, for each $1 \leq i \leq n$.

Given a finite number of contractions on a domain (D, μ) with a Lebesgue measurement μ, their union is modelled by a contraction on the *convex powerdomain* which then has a unique fixed point and yields the following result from [24]:

Theorem 11. *Let D be a continuous dcpo such that*

$$(\,\forall\, x, y \in D\,)(\,\exists\, z \in D\,)\; z \sqsubseteq x, y.$$

If $f : D \to D$ and $g : D \to D$ are contractions for which

$$(\,\exists\, x \in D\,)\; x \sqsubseteq f(x) \;\&\; x \sqsubseteq g(x),$$

then there is a unique $K \in \mathcal{P}_{com}(\ker \mu)$ such that $f(K) \cup g(K) = K$. In addition, it is an attractor:

$$(\forall C \in \mathcal{P}_{com}(\ker \mu))\, (f \cup g)^n(C) \to K,$$

in the Vietoris topology on $\mathcal{P}_{com}(\ker \mu)$.

In order to apply these results, we need a simple and clear way to recognize Lebesgue measurements. Let $f : [0, \infty)^2 \to [0, \infty)$ be a function such that $f(x_n, y_n) \to 0$ whenever $x_n, y_n \to 0$.

Theorem 12. *If $\mu : D \to [0, \infty)^*$ is a measurement such that for all pairs $x, y \in D$ with an upper bound,*

$$(\,\exists\, z \sqsubseteq x, y\,)\; \mu z \leq f(\mu x, \mu y),$$

then μ is a Lebesgue measurement.

The value of this result is that it identifies a condition satisfied by many of the Lebesgue measurements encountered in practice. For instance, just consider the number of examples covered by $f(s, t) = 2 \cdot \max\{s, t\}$.

Example 25. Lebesgue measurements.

(i) The domain of streams $(\Sigma^\infty, 1/2^{|\cdot|})$.
(ii) The powerset of the naturals $(\mathcal{P}\omega, |\cdot|)$.
(iii) The domain of partial maps $([\mathbb{N} \rightharpoonup \mathbb{N}], |\mathrm{dom}|)$.
(iv) The interval domain $(\mathbf{I}\mathbb{R}, \mu)$.
(v) The upper space $(\mathbf{U}X, \mathrm{diam})$ of a locally compact metric space (X, d).
(vi) The formal ball model $(\mathbf{B}X, \pi)$ of a complete metric space (X, d).

In fact, $f(s,t) = s + t$ applies to (i)–(v), the *triangle inequality.*

We are now going to apply Theorem 11 to obtain the classical result of [11] for hyperbolic iterated function systems on complete metric spaces.

Definition 20. An *iterated function system* (IFS) on a space X is a nonempty finite collection of continuous selfmaps on X. We write an IFS as $(X; f_1, \ldots, f_n)$.

Definition 21. An IFS $(X; f_1, \ldots, f_n)$ is *hyperbolic* if X is a complete metric space and f_i is a contraction for all $1 \leq i \leq n$.

Definition 22. Let (X, d) be a metric space. The *Hausdorff metric* on $\mathcal{P}_{com}(X)$ is

$$d_H(A, B) = \max\{\sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A), \}$$

for $A, B \in \mathcal{P}_{com}(X)$.

Hyperbolic iterated function systems are used to model fractals: Given a fractal image, one searches for a hyperbolic IFS which models it. But what does it mean to model an *image*? The answer is given by Hutchinson's fundamental result [11].

Theorem 13 (Hutchinson). *If $(X; f_1, \ldots, f_n)$ is a hyperbolic IFS on a complete metric space X, then there is a unique nonempty compact subset $K \subseteq X$ such that*

$$K = \bigcup_{i=1}^{n} f_i(K).$$

Moreover, for any nonempty compact set $C \subseteq X$, $(\bigcup_{i=1}^{n} f_i)^k(C) \to K$ in the Hausdorff metric d_H as $k \to \infty$.

At this stage, we can see that what will be most difficult in proving such a result is the convergence in the Hausdorff metric. Luckily, this topology is *independent* of the metric d on X.

Theorem 14. *Let (X, d) be a metric space. Then the topology induced by the Hausdorff metric d_H on $\mathcal{P}_{com}(X)$ is the Vietoris topology on $\mathcal{P}_{com}(X)$.*

In [6], the formal ball model $\mathbf{B}X$ is used to give a domain theoretic proof of the existence and uniqueness of the set K in Theorem 13 for any complete metric space (X, d). What is missing from that discussion is the important issue that K is also an attractor with respect to the Hausdorff metric d_H.

Example 26. If we have two contractions $f, g : X \to X$ on a complete metric space X, they have Scott continuous extensions

$$\bar{f}, \bar{g} : \mathbf{B}X \to \mathbf{B}X$$

which are contractions on $\mathbf{B}X$ with respect to $\pi(x, r) = r$. But π is a Lebesgue measurement on a domain which has the property that for all $(x, r), (y, s) \in \mathbf{B}X$, there is an element $z = (x, r + s + d(x, y)) \in \mathbf{B}X$ with $z \sqsubseteq (x, r), (y, s)$. In addition, for any $x \in X$, choosing r so that

$$r \geq \frac{d(x, fx)}{1 - c_f} \text{ and } r \geq \frac{d(x, gx)}{1 - c_g},$$

where $c_f, c_g < 1$ are the Lipschitz constants for f and g, respectively, gives a point $(x, r) \sqsubseteq \bar{f}(x, r), \bar{g}(x, r)$. By Theorem 11,

$$(\, \exists! K \in \mathcal{P}_{com}(\ker \pi) \,)\ \bar{f}(K) \cup \bar{g}(K) = K.$$

However, because $\ker \pi \simeq X$ and the mappings $\bar{f}, \bar{g}$ extend f and g, it is clear that

$$(\, \exists! K \in \mathcal{P}_{com}(X) \,)\ f(K) \cup g(K) = K$$

Finally, by Theorems 11 and 14, K is an attractor for $f \cup g$ on $\mathcal{P}_{com}(X)$.

If a space may be realized as the kernel of a Lebesgue measurement on a continuous *dcpo* D, then Theorem 11 implies that Hutchinson's result holds for any finite family of contractions which extend to D. *Necessarily,* two questions arise:

- Which spaces arise as the kernel of a Lebesgue measurement?
- When does a domain admit a Lebesgue measurement?

The answer to the first question is that a space is completely metrizable iff it is the kernel of a Lebesgue measurement on a continuous dcpo, and metrizable iff it is the kernel of a Lebesgue measurement on a continuous poset. The answer to the second question, for an ω continuous dcpo D, is that the set of maximal elements $\max(D)$ is regular iff it is metrizable iff it is the kernel of a Lebesgue measurement on D.

All of this is explained in more detail in [24]. Such results also have interesting implications for general relativity [27].

4 Instances of partiality

We now consider four examples of domains whose descriptions are nontrivial. Our first example is from analysis: the domain of real analytic mappings. The basic idea is to be able to write things like

$$1 \sqsubseteq 1 + x \sqsubseteq 1 + x + \frac{x^2}{2!} \sqsubseteq \ldots \sqsubseteq \bigsqcup_{n\geq 0}\left(1 + \ldots + \frac{x^n}{n!}\right) = e^x$$

Here the polynomials $(1+\ldots+x^n/n!)$ in a Taylor expansion are *partial*, while the analytic map e^x is *total.* On this domain, we see that analytic mappings arise as fixed points of monotone operators which provide schemes for how to compute them as a limit of 'finite approximations' (polynomials).

Our second example concerns finite probability distributions or *classical states*: $e_1 = (1, 0, \ldots, 0)$, $e_2 = (0, 1, 0, \ldots, 0), \ldots, e_n = (0, 0, \ldots, 1)$ are *total*, while all others are *partial*; in particular, the least informative distribution is $\perp = (1/n, \ldots, 1/n)$, which we expect to be a least element in the 'domain' of classical states. On this domain, the maximum entropy state of statistical mechanics arises as the least fixed point of a Scott continuous operator that gives a scheme for calculating it.

Our third example is the quantum analogue of the second: the domain of quantum states. In it, pure states $|\psi\rangle\langle\psi|$ are *total*, while all others (the mixed states) are *partial*; in particular, its least element is the completely mixed state $\perp = I/n$. On this domain, unital quantum channels will be seen to have the same domain theoretic properties as binary symmetric channels from classical information theory: they are Scott continuous and have a Scott closed set of fixed points. Later, after we have studied the informatic derivative, we will see that this domain enables us to calculate the Holevo capacity of a unital qubit channel. The domain of quantum states can also be used to recover classical and quantum logic in a unified manner.

Our fourth example is from general relativity: the domain of spacetime intervals. In it, single events $[x, x]$ are *total*, while nontrivial intervals $[p, q]$ are *partial*, in a manner completely analogous to the interval domain $\mathbf{I}\mathbb{R}$. In fact, these two domains have the exact same formal structure, as we will see. The domain of spacetime intervals is used to explain how spacetime, including its geometry, can be reconstructed in a purely order theoretic manner beginning from only a countable dense set. This result may be of interest to those concerned with the causal set approach to quantum gravity.

References: The results in this section are from the following sources: Section 3.1 is from some of the author's unpublished notes (1998), Section 3.2 is from [5, 31], 3.3 is from [5, 30] and 3.4 is from [26, 32, 27].

4.1 Analytic mappings

Real analytic mappings will be represented as infinite lists of rational numbers.

Definition 23. A *list* over $\mathbb{Q}$ is a function $x : \{0..., n\} \to \mathbb{Q}$ for $n \in \mathbb{N} \cup \{\infty\}$. The *length* of a list x is $|\mathrm{dom}(x)|$. $\mathbb{Q}^\infty$ is the set of both finite and infinite lists over $\mathbb{Q}$.

A finite list x is usually written as a vector $[x_0, ..., x_n]$, where $x_i = x(i)$. The empty list has been excluded above because there is no empty polynomial.

Definition 24. The prefix order $\sqsubseteq$ on $\mathbb{Q}^\infty$ is given by

$$x \sqsubseteq y \equiv \mathrm{dom}(x) \subseteq \mathrm{dom}(y) \text{ and } (\forall i \in \mathrm{dom}(x))\, x_i = y_i.$$

In this way, $\mathbb{Q}^\infty$ is an ω-algebraic Scott domain whose compact elements are exactly $\mathbb{Q}_{fin}$.

Definition 25. The *norm* of a power series

$$f(x) = \sum_{n=0}^{\infty} a_n\, x^n$$

is

$$\|f\|(x) = \sum_{n=0}^{\infty} |a_n\, x^n|$$

provided that this sum exists.

To say that a power series has a norm on $[a, b]$ means exactly that it converges absolutely on $[a, b]$.

Lemma 2. *If a function f is defined by an absolutely convergent power series on $[a, b]$, then f and $\|f\|$ are both continuous on $[a, b]$.*

Recall that $C[a, b]$ denotes the space of continuous real value function defined on $[a, b]$.

Definition 26. The *degree* of a list p is $|p| :=$ (length p) $- 1$, with the understanding that the degree of an infinite list is ∞. For a list of rationals $a \in \mathbb{Q}^\infty$, we set

$$(\sigma a)x = \sum_{n=0}^{|a|} a_n\, x^n,$$

whenever such a sum exists for all $x \in [a, b]$. This gives a partial mapping $\sigma : \mathbb{Q}^\infty \to C[a, b]$. A list p is *analytic* on $[a, b]$ if $\|(\sigma p)r\| < \infty$ where $r = \max\{|a|, |b|\}$.

Observe that σ is defined for *any* analytic list.

Corollary 3. *For every analytic $p \in \mathbb{Q}^\infty$, $\sigma p \in C[a, b]$ & $\|\sigma p\| \in C[a, b]$.*

For $f, g \in C[a, b]$, the *uniform metric* is

$$d(f, g) = \sup\{|f(x) - g(x)| : x \in [a, b]\}$$

With these preliminaries out the way, we can now order analytic mappings:

Definition 27. The set

$$\mathbb{P}^\infty[a,b] := \{(p,r) : p \textit{ analytic}, r \in [0,\infty)^*\}$$

is ordered by

$$(p,r) \sqsubseteq (q,s) \equiv p \sqsubseteq q \text{ and } d(\|\sigma p\|, \|\sigma q\|) \leq r - s.$$

Theorem 15. *$\mathbb{P}^\infty[a,b]$ is an ω-continuous dcpo with a countable basis given by*

$$\{(p,r) : p \textit{ finite}, r \in \mathbb{Q} \;\&\; r \geq 0\}.$$

Its approximation relation is

$$(p,r) \ll (q,s) \Leftrightarrow p \textit{ finite} \;\&\; d(||\sigma p||, ||\sigma q||) < r - s$$

and its natural measurement $\mu : \mathbb{P}^\infty[a,b] \to [0,\infty)^*$ given by

$$\mu(p,r) = r + \frac{1}{2^{|p|}}$$

measures all of $\mathbb{P}^\infty[a,b]$, has $\ker\mu = \max(\mathbb{P}^\infty[a,b])$ and satisifies the triangle inequality: for all pairs $x, y \in \mathbb{P}^\infty[a,b]$ with an upper bound, there is $z \sqsubseteq x, y$ with $\mu z \leq \mu x + \mu y$.

We adopt the convention of writing

$$\mathcal{P}_{fin}[a,b] = \{(p,r) \in \mathbb{P}^\infty[a,b] : p \text{ finite}, r \geq 0\}$$

Proposition 4. *If $f : \mathcal{P}_{fin}[a,b] \to E$ is a monotone map into a dcpo such that*

$$f(p,r) = \bigsqcup f(p, r + 1/n)$$

for p finite, then f may be extended uniquely to a Scott continuous map on all of $\mathbb{P}^\infty[a,b]$.

A mapping f of the type discussed in the previous result is said to be *invariant on polynomials.*

Example 27. The unary operation addition by 1

$$(p,r) \mapsto ([a_0 + 1, \ldots, a_n], r)$$

is monotone and invariant on polynomials, so it extends uniquely to $\mathbb{P}^\infty[a,b]$.

At times we may blur the distinction between polynomials and lists of rational numbers, that is, we will treat them as one and the same for the purpose of illustrating various points about mappings on $\mathbb{P}^\infty[a,b]$ and the *functions* they act on. Now for a nontrivial example.

Example 28. Let $[a, b]$ be an interval containing 0 and define

$$I : \mathcal{P}_{fin}[a, b] \to \mathbb{P}^{\infty}[a, b]$$

$$I(p, r) = (\int_0^x p(t)\ dt, m \cdot r)$$

where $m = \max\{|a|, |b|\}$ and $\int_0^x p(t)\ dt$ is the list operation taking $p = [a_0, ..., a_n]$ to $[0, a_0, ..., a_n/(n + 1)]$. This mapping is monotone and invariant on polynomials so it has a unique Scott continuous extension to all of $\mathbb{P}^{\infty}[a, b]$, which we denote by $\int_0^x$.

From a *symbolic* definition of integral for polynomials, the one we normally program when implementing the polynomial data type, we systematically obtain a definition of integral for analytic mappings.

Example 29. The exponential map. Consider the operator

$$\exp : \mathbb{P}^{\infty}[-c, c] \to \mathbb{P}^{\infty}[-c, c]$$

$$\exp(p, r) = 1 + \int_0^x (p, r)$$

for $0 < c < 1$, the Scott continuous map $\int_0^x$ composed with the unary Scott continuous operator that adds 1. Since $(1, r) \sqsubseteq \exp(1, r)$ for $r \geq 1$, Scott continuity gives a fixed point

$$\text{fix}(\exp) = \bigsqcup_{n \geq 0} \exp^n(1, 1)$$

that is easily seen to be the exponential map e^x. This fixed point is *unique.*

First, because exp is a contraction with respect to the natural measurement μ with $\mu(\exp) \leq \max\{c, 1/2\} \cdot \mu$, any other fixed point $\exp(p, r) = (p, r)$ yields $\mu(p, r) = 0$. But any fixed point $(p, 0)$ must also have $1 \sqsubseteq p$. Let $r := d(1, \|\sigma p\|) + 1 \geq 1$. Then since $(1, r) \sqsubseteq (p, 0)$, we have a lower bound for both $(p, 0)$ and fix(exp), which gives $(p, 0) = \text{fix}(\exp)$ since exp is a contraction.

Notice that $\exp^n(1, 1)$ is the n^{th}-degree Taylor approximation of the maximal element e^x. In addition, the smaller the interval $[-c, c]$, the smaller that c is, the quicker that exp converges to e^x. Thus, using the domain of analytic mappings we see that *fewer terms of the Taylor series* are required to approximate e^x on the interval $[-c/2, c/2]$ than on the interval $[-c, c]$.

In a similar way one can realize the sine and cosine functions as unique fixed points of Scott continuous mappings.

Example 30. The trigonometric functions. The operator for the sine is

$$\phi(p, r) = x - \int_0^x \int_0^y (p, r).$$

The operator for the cosine is

$$\phi(p,r) = 1 - \int_0^x \int_0^y (p,r).$$

The iteration for the first begins with the polynomial x, and the second begins with the polynomial 1.

4.2 Classical states

Definition 28. Let $n \geq 2$. The *classical states* are

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

A classical state $x \in \Delta^n$ is *pure* when $x_i = 1$ for some $i \in \{1, \ldots, n\}$; we denote such a state by e_i.

Pure states $\{e_i\}_i$ are the actual states a system can be in, while general mixed states x and y are epistemic entities. Imagine that one of n different outcomes is possible. If our knowledge of the outcome is $x \in \Delta^n$, and then by some means we determine that outcome i is not possible, our knowledge improves to

$$p_i(x) = \frac{1}{1-x_i}(x_1, \ldots, \hat{x_i}, \ldots, x_{n+1}) \in \Delta^n,$$

where $p_i(x)$ is obtained by first removing x_i from x and then renormalizing. The partial mappings which result, $p_i : \Delta^{n+1} \rightharpoonup \Delta^n$ with $\mathrm{dom}(p_i) = \Delta^{n+1} \setminus \{e_i\}$, are called the *Bayesian projections* and lead one to the following relation on classical states.

Definition 29. For $x, y \in \Delta^{n+1}$,

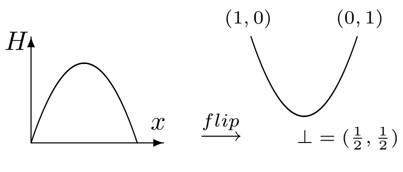
$$x \sqsubseteq y \equiv (\forall i)(x, y \in \mathrm{dom}(p_i) \Rightarrow p_i(x) \sqsubseteq p_i(y)).$$

For $x, y \in \Delta^2$,

$$x \sqsubseteq y \equiv (y_1 \leq x_1 \leq 1/2) \text{ or } (1/2 \leq x_1 \leq y_1).$$

The relation $\sqsubseteq$ on Δ^n is called the *Bayesian order.*

As we can see, the definition of Δ^{n+1} from Δ^n is natural. The order on Δ^2, is derived from the graph of entropy $H(x) = -x\log_2(x) - (1-x)\log_2(1-x)$ as follows:



It is canonical in the following sense:

Theorem 16. *There is a unique partial order on Δ^2 that satisfies the mixing law*

$$x \sqsubseteq y \text{ and } p \in [0,1] \;\Rightarrow\; x \sqsubseteq (1-p)x + py \sqsubseteq y$$

and has $\perp := (1/2, 1/2)$ as a least element. It is the Bayesian order on classical two states.

The Bayesian order was discovered in [5] where the following is proven:

Theorem 17. $(\Delta^n, \sqsubseteq)$ *is a dcpo with least element $\perp := (1/n, \ldots, 1/n)$ and* $\max(\Delta^n) = \{e_i : 1 \leq i \leq n\}$. *It has Shannon entropy*

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

as a measurement of type $\Delta^n \to [0, \infty)^*$.

A more subtle example of a measurement on Δ^n in its Bayesian order is the retraction $r : \Delta^n \to \Lambda^n$ which rearranges the probabilities in a classical state into descending order.

The Bayesian order has a more direct description: the *symmetric formulation.* Let $S(n)$ denote the group of permutations on $\{1, \ldots, n\}$ and

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

denote the collection of *monotone* decreasing classical states. It can then be shown [5] that for $x, y \in \Delta^n$, we have $x \sqsubseteq y$ iff there is a permutation $\sigma \in S(n)$ such that $x \cdot \sigma, y \cdot \sigma \in \Lambda^n$ and

$$(x \cdot \sigma)_i (y \cdot \sigma)_{i+1} \leq (x \cdot \sigma)_{i+1} (y \cdot \sigma)_i$$

for all i with $1 \leq i < n$. Thus, $(\Delta^n, \sqsubseteq)$ can be thought of as $n!$ many copies of the domain $(\Lambda^n, \sqsubseteq)$ identified along their common boundaries, where $(\Lambda^n, \sqsubseteq)$ is

$$x \sqsubseteq y \equiv (\forall i < n)\, x_i y_{i+1} \leq x_{i+1} y_i.$$

It should be remarked though that the problems of ordering Λ^n and Δ^n are very different, with the latter being far more challenging, especially if one also wants to consider quantum mixed states. Let us now consider an important application of the Bayesian order to give a method for calculating the maximum entropy state of statistical mechanics.

The maximum entropy principle

The possible outcomes of an event are $a_1, \ldots, a_n$. It is repeated many times and an average value of E is observed. What is the probability p_i of a_i? The *maximum entropy principle* provides an approach to solve this problem: because Shannon entropy has a maximum value on the set

$$\left\{ p \in \Delta^n : \sum_{i=1}^{n} p_i \cdot a_i = E \right\}$$

that is assumed at exactly one point, one possibility is to use this state as the probability distribution that models our observed data. Beautiful – but how do we calculate this distribution?

Define

$$f(x) = \frac{\sum_{i=1}^{n} a_i e^{x a_i}}{\sum_{i=1}^{n} e^{x a_i}} - E, \quad I_f(x) = x - \frac{f(x)}{(a_n - a_1)^2}$$

for any $x \in \mathbb{R}$. Define $\lambda : \Delta^n \to \mathbb{R} \cup \{\pm\infty\}$ by

$$\lambda(x) = \begin{cases} \frac{\log\left(\frac{\mathrm{sort}(x)_1}{\mathrm{sort}(x)_2}\right)}{a_n - a_{n-1}} & \text{if } I_f(0) > 0; \\ \\ \frac{\log\left(\frac{\mathrm{sort}(x)_1}{\mathrm{sort}(x)_2}\right)}{a_1 - a_2} & \text{otherwise.} \end{cases}$$

with the understanding for pure states that $\lambda x = \infty$ in the first case and $\lambda x = -\infty$ in the other. The map sort puts states into *decreasing* order.

Theorem 18. *Let $a_1 < E < a_n$. The map*

$$\phi : \Delta^n \to \Delta^n$$

given by

$$\phi(x) = (e^{I_f(\lambda x) a_1}, \ldots, e^{I_f(\lambda x) a_n}) \cdot \frac{1}{Z(x)}$$

$$Z(x) = \sum_{i=1}^{n} e^{I_f(\lambda x) a_i}$$

is Scott continuous in the Bayesian order. Its least fixed point is the maximum entropy state.

The maximum entropy principle has been successfully applied to perform image reconstruction from noisy data, probabilistic link extraction from intelligence data, natural language processing, stock price volatility, thermodynamics.

4.3 Quantum states

Let $\mathcal{H}^n$ denote an n-dimensional complex Hilbert space with specified inner product $\langle\cdot|\cdot\rangle$.

Definition 30. A *quantum state* is a density operator $\rho : \mathcal{H}^n \to \mathcal{H}^n$, i.e., a self-adjoint, positive, linear operator with $\mathrm{tr}(\rho) = 1$. The quantum states on $\mathcal{H}^n$ are denoted Ω^n.

Definition 31. A quantum state ρ on $\mathcal{H}^n$ is *pure* if

$$\mathrm{spec}(\rho) \subseteq \{0, 1\}.$$

The set of pure states is denoted Σ^n. They are in bijective correspondence with the one dimensional subspaces of $\mathcal{H}^n$.

Classical states are distributions on the set of pure states $\max(\Delta^n)$. By Gleason's theorem, an analogous result holds for quantum states: Density operators encode distributions on the set of pure states Σ^n.

Definition 32. A *quantum observable* is a self-adjoint linear operator $e : \mathcal{H}^n \to \mathcal{H}^n$.

An observable of a physical system is anything about it that we can measure. For example, *energy* is an observable. Observables in quantum mechanics are represented mathematically by self-adjoint operators.

If we have the operator e representing the energy observable of a system (for instance), then its set of eigenvalues $\mathrm{spec}(e)$, called the *spectrum* of e, consists of the actual energy values a system may assume. If our knowledge about the state of the system is represented by density operator ρ, then quantum mechanics predicts the probability that a measurement of observable e yields the value $\lambda \in \mathrm{spec}(e)$. It is

$$\mathrm{pr}(\rho \to e_\lambda) := \mathrm{tr}(p_e^\lambda \cdot \rho),$$

where p_e^λ is the projection corresponding to eigenvalue λ and e_λ is its associated eigenspace in the *spectral representation* of e.

Definition 33. Let e be an observable on $\mathcal{H}^n$ with $\mathrm{spec}(e) = \{1, \ldots, n\}$. For a quantum state ρ on Ω^n,

$$\mathrm{spec}(\rho|e) := (\mathrm{pr}(\rho \to e_1), \ldots, \mathrm{pr}(\rho \to e_n)) \in \Delta^n.$$

We assume that all observables e have $\mathrm{spec}(e) = \{1, \ldots, n\}$. For our purposes it is enough to assume $|\mathrm{spec}(e)| = n$; the set $\{1, \ldots, n\}$ is chosen for the sake of aesthetics. Intuitively, then, e is an experiment on a system which yields one of n different outcomes; if our a priori knowledge about the state of the system is ρ, then our knowledge about what the result of experiment e

will be is $\mathrm{spec}(\rho|e)$. Thus, $\mathrm{spec}(\rho|e)$ determines our ability to *predict* the result of the experiment e.

Let us point out that $\mathrm{spec}(\rho) = \mathrm{Im}(\mathrm{spec}(\rho|e))$ and $\mathrm{spec}(\sigma) = \mathrm{Im}(\mathrm{spec}(\sigma|e))$ are equivalent to $[\rho, e] = 0$ and $[\sigma, e] = 0$, where $[a, b] = ab - ba$ is the commutator of operators.

Definition 34. Let $n \geq 2$. For quantum states $\rho, \sigma \in \Omega^n$, we have $\rho \sqsubseteq \sigma$ iff there is an observable $e : \mathcal{H}^n \to \mathcal{H}^n$ such that $[\rho, e] = [\sigma, e] = 0$ and $\mathrm{spec}(\rho|e) \sqsubseteq \mathrm{spec}(\sigma|e)$ in Δ^n.

This is called the *spectral order* on quantum states.

Theorem 19. $(\Omega^n, \sqsubseteq)$ *is a dcpo with maximal elements* $\max(\Omega^n) = \Sigma^n$ *and least element* $\perp = I/n$*, where* I *is the identity matrix. It has von Neumann entropy*

$$\sigma\rho = -\mathrm{tr}(\rho \log \rho)$$

as a measurement of type $\Omega^n \to [0, \infty)^*$.

Another natural measurement on Ω^n is the map $q : \Omega^n \to \Lambda^n$ which assigns to a quantum state its spectrum rearranged into descending order. It can be thought of as an important link between classical and quantum information theory.

There is one case where the spectral order can be described in an elementary manner.

Example 31. The 2×2 density operators Ω^2 can be represented as points on the unit ball in $\mathbb{R}^3$:

$$\Omega^2 \simeq \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \leq 1\}.$$

For example, the origin $(0, 0, 0)$ corresponds to the completely mixed state $I/2$, while the points on the surface of the sphere describe the pure states. The order on Ω^2 then amounts to the following: $x \sqsubseteq y$ iff the line from the origin $\perp$ to y passes through x.

Let us now consider an application of Ω^2 to the study of communication.

Classical and quantum communication

The classical channels $f : \Delta^2 \to \Delta^2$ which increase entropy $(H(f(x)) \geq H(x))$ are exactly those f with $f(\perp) = \perp$. They are the *strict* mappings of domain theory, which are also known as *binary symmetric channels* in information theory. Similarly, the entropy increasing qubit channels are exactly those channels[2] $\varepsilon : \Omega^2 \to \Omega^2$ for which $\varepsilon(\perp) = \perp$. These are called *unital* in quantum information theory.

[2] Quantum channels are completely positive and convex linear, see [35] for more.

Definition 35. A qubit channel $\varepsilon : \Omega^2 \to \Omega^2$ is *unital* if $\varepsilon(\bot) = \bot$.

Theorem 20.

- *A classical channel $f : \Delta^2 \to \Delta^2$ is binary symmetric iff it is Scott continuous and its set of fixed points is Scott closed.*
- *A quantum channel $f : \Omega^2 \to \Omega^2$ is unital if and only if it is Scott continuous and its set of fixed points is Scott closed.*

In fact, this last result hints at how to establish the uniqueness of Ω^2, in a manner completely similar to the corresponding result for Δ^2:

Theorem 21. *There is a unique partial order on Ω^2 with the following three properties:*

(i) *It has least element $\bot = I/2$,*
(ii) *It satisfies the mixing law: if $r \sqsubseteq s$, then $r \sqsubseteq tr + (1-t)s \sqsubseteq s$, for all $t \in [0,1]$,*
(iii) *Every unital channel $f : \Omega^2 \to \Omega^2$ is Scott continuous and has a Scott closed set of fixed points.*

It is the spectral order, and gives Ω^2 the structure of a Scott domain.

Finally, let us turn to one last application of the spectral order.

Classical and quantum logic

The logics of Birkhoff and von Neumann consist of the propositions one can make about a physical system. Each proposition takes the form "The value of observable e is contained in $E \subseteq \text{spec}(e)$." For classical systems, the logic is $\mathcal{P}\{1,\ldots,n\}$, while for quantum systems it is $\mathbb{L}^n$, the lattice of (closed) subspaces of $\mathcal{H}^n$. In each case, implication of propositions is captured by inclusion, and a fundamental distinction between classical and quantum – that there are pairs of quantum observables whose exact values cannot be simultaneously measured at a single moment in time – finds lattice theoretic expression: $\mathcal{P}\{1,\ldots,n\}$ is distributive; $\mathbb{L}^n$ is not.

Remarkably, the classical and quantum logics can be *derived* from the Bayesian and spectral orders using the *same* order theoretic technique.

Definition 36. An element x of a dcpo D is *irreducible* when

$$\bigwedge(\uparrow\!x \cap \max(D)) = x$$

The set of irreducible elements in D is written $\text{Ir}(D)$.

The order dual of a poset $(D, \sqsubseteq_D)$ is written D^*; its order is $x \sqsubseteq y \Leftrightarrow y \sqsubseteq_D x$.

Theorem 22. *For $n \geq 2$, the classical lattices arise as*

$$\mathrm{Ir}(\Delta^n)^* \simeq \mathcal{P}\{1,\ldots,n\} \setminus \{\emptyset\},$$

and the quantum lattices arise as

$$\mathrm{Ir}(\Omega^n)^* \simeq \mathbb{L}^n \setminus \{0\}.$$

4.4 Spacetime intervals

General relativity is Einstein's theory of gravity in which gravity is understood not in terms of mysterious "universal" forces but rather as part of the geometry of spacetime. It is profoundly beautiful and beautifully profound from both the physical and mathematical viewpoints and it teaches us clear lessons about the universe in which we live that are easily explainable. For example, it offers a wonderful explanation of gravity: if an apple falls from a tree, the path it takes is not determined by the Newtonian ideal of an "invisible force" but instead by the curvature of the space in which the apple resides: gravity is the curvature of spacetime. In addition, the presence of matter in spacetime causes it to "bend" and Einstein even gives us an equation that relates the curvature of spacetime to the matter present within it. *However.*

Since everything attracts everything else, a gravitating mass of sufficient size will eventually collapse. In 1965, Penrose [36] showed that any such collapse eventually leads to a singularity where the mathematical description of spacetime as a continuum breaks down. This leads to the need to reformulate gravity. It is hoped that the elusive quantum theory of gravity will resolve this problem.

Since the first singularity theorems [36, 10], causality has played a key role in understanding spacetime structure. The analysis of causal structure relies heavily on techniques of differential topology [37]. For the past decade Sorkin and others [42] have pursued a program for quantization of gravity based on causal structure. In this approach the causal relation is regarded as the fundamental ingredient and the topology and geometry are secondary.

In this section, we will see that the causal structure of spacetime is captured by a *domain* and learn the surprising connection between measurement, the Newtonian concept of time, and the geometry of spacetime.

Definition 37. A continuous poset $(P, \leq)$ is *bicontinuous* if

- For all $x, y \in P$, $x \ll y$ iff for all filtered $S \subseteq P$ with an infimum,

  $$\bigwedge S \leq x \Rightarrow (\exists s \in S)\, s \leq y,$$

  and
- For each $x \in P$, the set $\Uparrow\!x$ is filtered with infimum x.

We tend to prefer the notation $\leq$ for the order on a poset that is known to be bicontinuous. For x, y in a poset $(P, \leq)$,

$$x < y \equiv x \leq y \ \& \ x \neq y.$$

In general, $<$ and $\ll$ are completely different ideas.

Example 32. $(\mathbb{R}, \leq)$, $(\mathbb{Q}, \leq)$ are bicontinuous.

Definition 38. The *interval topology* on a continuous poset P exists when sets of the form

$$(a, b) = \{x \in P : a \ll x \ll b\} \quad \& \quad \twoheaduparrow x = \{y \in P : x \ll y\}$$

form a basis for a topology on P.

Notice that on a *bicontinuous poset*, the interval topology exists and has

$$(a, b) := \{x \in P : a \ll x \ll b\}$$

as a basis.

A *manifold* $\mathcal{M}$ is a locally Euclidean Hausdorff space that is connected and has a countable basis. Such spaces are paracompact. A *Lorentz metric* on a manifold is a symmetric, nondegenerate tensor field of type $(0, 2)$ whose signature is $(- + ++)$.

Definition 39. A *spacetime* is a real four-dimensional[3] smooth manifold $\mathcal{M}$ with a Lorentz metric g_{ab}.

Let $(\mathcal{M}, g_{ab})$ be a time-orientable spacetime. Let $\Pi^+_\leq$ denote the future directed causal curves, and $\Pi^+_\ll$ denote the future directed time-like curves.

Definition 40. For $p \in \mathcal{M}$,

$$I^+(p) := \{q \in \mathcal{M} : (\exists \pi \in \Pi^+_\ll)\, \pi(0) = p, \pi(1) = q\}$$

and

$$J^+(p) := \{q \in \mathcal{M} : (\exists \pi \in \Pi^+_\leq)\, \pi(0) = p, \pi(1) = q\}$$

Similarly, we define $I^-(p)$ and $J^-(p)$.

We write the relation J^+ as

$$p \leq q \equiv q \in J^+(p).$$

The "Alexandroff topology" on a spacetime has $\{I^+(p) \cap I^-(q) : p, q \in \mathcal{M}\}$ as a basis; a spacetime $\mathcal{M}$ is strongly causal iff its Alexandroff topology is Hausdorff iff its Alexandroff topology is the manifold topology. Penrose has called *globally hyperbolic* spacetimes "the physically reasonable spacetimes [44]."

[3] The results in the present paper work for any dimension $n \geq 2$ [26].

Definition 41. A spacetime $\mathcal{M}$ is *globally hyperbolic* if it is strongly causal and if $\uparrow a \cap \downarrow b$ is compact in the manifold topology, for all $a, b \in \mathcal{M}$.

Theorem 23. *If $\mathcal{M}$ is globally hyperbolic, then $(\mathcal{M}, \leq)$ is a bicontinuous poset with $\ll \, = I^+$ whose interval topology is the manifold topology.*

This result motivates the following definition:

Definition 42. A poset $(X, \leq)$ is *globally hyperbolic* if it is bicontinuous and each interval $[a, b] = \{x : a \leq x \leq b\}$ is compact in the interval topology.

Globally hyperbolic posets have rich enough structure that we can deduce many properties of spacetime from them *without* appealing to differentiable structure or geometry, such as the compactness of the space of causal curves [27]. We can also deduce new aspects of spacetime. Globally hyperbolic posets are very much like the real line. In fact, a well-known domain theoretic construction pertaining to the real line extends in perfect form to the globally hyperbolic posets:

Theorem 24. *The closed intervals of a globally hyperbolic poset X*

$$\mathbf{I}X := \{[a, b] : a \leq b \ \& \ a, b \in X\}$$

ordered by reverse inclusion

$$[a, b] \sqsubseteq [c, d] \equiv [c, d] \subseteq [a, b]$$

form a continuous domain with

$$[a, b] \ll [c, d] \equiv a \ll c \ \& \ d \ll b.$$

The poset X has a countable basis iff $\mathbf{I}X$ is ω-continuous. Finally,

$$\max(\mathbf{I}X) \simeq X$$

where the set of maximal elements has the relative Scott topology from $\mathbf{I}X$.

In fact, more is true: in [26] it is shown that the category of globally hyperbolic posets is naturally isomorphic to the category of *interval domains*. This observation – that spacetime has a canonical domain theoretic model – teaches us something new: from only a countable set of events and the causality relation, one can reconstruct spacetime in a purely order theoretic manner. Explaining this requires domain theory.

Reconstruction of the spacetime manifold

An *abstract basis* is a set $(C, \ll)$ with a *transitive* relation that is *interpolative* from the $-$ *direction*:

$$F \ll x \Rightarrow (\exists y \in C)\, F \ll y \ll x,$$

for all finite subsets $F \subseteq C$ and all $x \in F$. Suppose, though, that it is also interpolative from the $+$ *direction*:

$$x \ll F \Rightarrow (\exists y \in C)\, x \ll y \ll F.$$

Then we can define a new abstract basis of *intervals*

$$\mathrm{int}(C) = \{(a,b) : a \ll b\} = \ll \subseteq C^2$$

whose relation is

$$(a,b) \ll (c,d) \equiv a \ll c \;\&\; d \ll b.$$

Let $\mathbf{I}C$ denote the ideal completion of the abstract basis $\mathrm{int}(C)$.

Theorem 25. *Let C be a countable dense subset of a globally hyperbolic spacetime $\mathcal{M}$ and $\ll = I^+$ be timelike causality. Then*

$$\max(\mathbf{I}C) \simeq \mathcal{M}$$

where the set of maximal elements have the Scott topology.

Theorem 25 is very different from results like “Let $\mathcal{M}$ be a certain spacetime with relation $\leq$. Then the interval topology is the manifold topology.” Here we identify, in abstract terms, a process by which a countable set with a causality relation determines a space. The process is entirely order theoretic in nature, spacetime is not required to understand or execute it (i.e., if we put $C = \mathbb{Q}$ and $\ll = <$, then $\max(\mathbf{I}C) \simeq \mathbb{R}$). In this sense, our understanding of the relation between causality and the topology of spacetime is now explainable independently of geometry. Ideally, one would now like to know what constraints on C in general imply that $\max(\mathbf{I}C)$ is a manifold.

Time and measurement

A global time function $t : \mathcal{M} \to \mathbb{R}$ on a globally hyperbolic spacetime $\mathcal{M}$ is a continuous function such that $x < y \Rightarrow t(x) < t(y)$ and $t^{-1}(r) = \Sigma$ is a Cauchy surface for $\mathcal{M}$, for each $r \in \mathbb{R}$.

Theorem 26. *For any global time function $t : \mathcal{M} \to \mathbb{R}$ on a globally hyperbolic spacetime, the function $\Delta t : \mathcal{M} \to [0,\infty)^*$ given by $\Delta t[a,b] = t(b) - t(a)$ measures all of $\mathbf{I}(\mathcal{M})$. It is a measurement with $\ker(\Delta t) = \max(\mathbf{I}(\mathcal{M}))$.*

Let $d : \mathbf{I}(\mathcal{M}) \to [0,\infty)^*$ denote the Lorentz distance on a globally hyperbolic spacetime

$$d[a,b] = \sup_{\pi_{ab}} \mathrm{len}(\pi_{ab})$$

where the sup is taken over all causal curves that join a to b.

A function between continuous posets is *interval continuous* when each poset has an interval topology and the inverse image of an interval open set is interval open. By the bicontinuity of $\mathcal{M}$, the interval topology on $\mathbf{I}(\mathcal{M})$ exists, so we can consider interval continuity for functions $\mathbf{I}(\mathcal{M}) \to [0,\infty)^*$.

Theorem 27. *The Lorentz distance $d : \mathbf{I}(\mathcal{M}) \to [0,\infty)^*$ has the following properties:*

(i) *It is monotone: $x \leq y \Rightarrow d(x) \geq d(y)$,*
(ii) *It preserves the way below relation: $x \ll y \Rightarrow d(x) > d(y)$,*
(iii) *It is interval continuous and hence, by* (i)*, Scott continuous.*

It does not measure $\mathbf{I}(\mathcal{M})$ at any point of $\ker(d)$.

That the Lorentz distance is not a measurement has all to do with relativity: it is a direct consequence of the fact that a clock travelling at the speed of light records no time as having elapsed i.e. the set of null intervals is equal to

$$\ker(d) \setminus \max(\mathbf{I}(\mathcal{M})) \neq \emptyset$$

but measurements μ always satisfy $\ker(\mu) \subseteq \max(D)$ (Lemma 1).

In fact, no interval continuous function $\mu : \mathbf{I}(\mathcal{M}) \to [0,\infty)^*$ can be a measurement: by interval continuity, $\mu x = 0$ for any x with $\mathord{\uparrow\!\!\uparrow} x = \emptyset$. Then just like the Lorentz distance, an interval continuous μ will also assign 0 to "null intervals." In this way, we see that interval continuity captures an essential aspect of the Lorentz distance: interval continuous functions do not distinguish between single events and null intervals. In addition, since Δt is a measurement, it cannot be interval continuous. This provides a surprising *topological* distinction between the Newtonian and relativistic concepts of time: d is interval continuous, Δt is not. Put another way, Δt can be used to reconstruct the *topology* of spacetime (Theorem 2), while d is used to reconstruct its *geometry.*

Reconstruction of spacetime geometry

Specifically, if in addition to $\mathrm{int}(C)$ we also begin with a countable collection of numbers l_{ab} chosen for each $(a,b) \in \mathrm{int}(C)$ in such a way that the map

$$\mathrm{int}(C) \to [0,\infty)^* :: (a,b) \mapsto l_{ab}$$

is monotone, then in the process of reconstructing spacetime, we can also construct the Scott continuous function $d : \mathbf{I}C \to [0,\infty)^*$ given by

$$d(x) = \inf\{l_{ab} : (a,b) \ll x\}.$$

In the event that the countable number of l_{ab} chosen are the Lorentz distances $l_{ab} = d[a,b]$, then the function d constructed above yields the Lorentz distance for any spacetime interval, the reason being that both are Scott continuous and are equal on a basis of the domain.

Thus, from a countable dense set of events and a countable set of distances, we can reconstruct the spacetime manifold together with its geometry in a purely order theoretic manner.

5 The informatic derivative

Major references: [15, 20, 30]

5.1 In a single measurement

Recall the seemingly innocent definition of the μ *topology* from Section 2.4:

Definition 43. The μ *topology* on a continuous dcpo D has as a basis all sets of the form $\mathrel{\uparrow\!\!\!\uparrow} x \cap \downarrow y$ where $x, y \in D$. It is denoted μ_D.

This also turns out to be the topology one needs to define rates of change on a domain. This comes as something of a surprise since the μ topology is *always zero-dimensional and Hausdorff.*

Definition 44. Let D be a continuous dcpo with a map $\mu : D \to [0,\infty)^*$ that measures $X \subseteq D$. If $f : D \to D$ is a function and $p \in X$ is not a compact element of D, then

$$df_\mu(p) := \lim_{x \to p} \frac{\mu f(x) - \mu f(p)}{\mu x - \mu p}$$

is called *the informatic derivative* of f at p with respect to μ, provided that it exists. The limit above is taken with respect to the μ topology.

If the limit above exists, then it is unique, since the μ topology is Hausdorff, and we are taking a limit at a point that is not *isolated*: $\{p\}$ is μ open iff p is compact. Notice too the importance of strict monotonicity of μ in Lemma 1: without it, we could not define the derivative. The definition of informatic derivative has a simple extension to functions $f : D \to E$ between domains with measurements (D, μ) and (E, λ) [15].

Our first example comes from calculus and provided the first relationship between domain theory and the differential calculus [15].

Theorem 28. *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous map on the real line with $p \in \mathbb{R}$. If $f'(p)$ exists, then*

$$d\bar{f}_\mu[p] = |f'(p)|$$

where $\bar{f}(x) = f(x)$ is the canonical extension of f to $\mathbf{I}\mathbb{R}$ and $\mu[a,b] = b - a$.

In particular, any iterative process with a classical derivative has an informatic derivative, and from the complexity viewpoint, they are equal. In fact, it can be shown that $d\bar{f}_\mu$ exists and is continuous iff f has a continuous first derivative i.e. the informatic derivative is equivalent to the classical derivative for C^1 functions. However, in general, informatic differentiability of $\bar{f}$ is strictly more general than classical differentiability [25].

5.2 The derivative at a fixed point

It often happens that partial maps on spaces have fixed points which are *unknown.* For example, the polynomial $p : \mathbb{R} \to \mathbb{R}$ given by $p(x) = x^3 + x - 1$ has a zero on $[0, 1]$ because $p(0) \cdot p(1) < 0$. Consequently, $f(x) = x - p(x)$ has a fixed point on $[0, 1]$, even though we are not sure of what it is.

Because a partial map $f : X \rightharpoonup X$ on a space X may have an *unknown* fixed point p, methods for calculating it are important. A minimal requirement is usually that p be an *attractor*: that there exist an open set $U \subseteq X$ such that for all $x \in U$, $f^n(x) \to p$. This provides a simple scheme for approximating p: simply calculate the iterates $f^n(x)$ beginning with any $x \in U$.

In Sections 3 and 4 we saw many examples of numerical methods and monotone maps (when restricted to $I(f) = \{x : x \sqsubseteq f(x)\}$) which give rise to partial splittings that converge to fixed points.

Lemma 3. *Let $s : D \rightharpoonup D$ be a partial splitting which maps into* $\mathrm{dom}(s)$*. If $s(p) = p$ and $ds_\mu(p)$ exists, then $ds_\mu(p) \leq 1$.*

So we consider partial maps f with fixed points p such that $df_\mu(p) \leq 1$. The identity map $1 : D \to D$ has $d(1)_\mu(p) = 1$ at any element which is not compact, meaning that a map whose derivative is unity need not have an attractive point. However, if $df_\mu(p) < 1$, then we can say something: for monotone maps with fixed points in the kernel, we have an attractor in the Scott topology.

Theorem 29. *Let $f : (D, \mu) \to (D, \mu)$ be a monotone mapping with $f(\ker\mu) \subseteq \ker\mu$. If $df_\mu(p) < 1$ at a fixed point $f(p) = p \in \ker\mu$, then there is an approximation $a \ll p$ such that*

(i) *For all $x \in D$, if $a \sqsubseteq x \sqsubseteq p$, then*

$$\bigsqcup_{n\geq 0} f^n(x) = p,$$

and this is a limit in the μ topology on D.

(ii) *The unique fixed point of f on $\uparrow a$ is p.*

(iii) *For all $x \in \ker\mu \cap \uparrow a$, $f^n(x) \to p$ in the Scott topology on* $\ker\mu$*.*

In [15], it is shown that (i) is equivalent to f being μ *continuous* at p, so we can take (i) as a definition of μ continuity at a fixed point. The bisection method split_f is not necessarily μ continuous at a fixed point if the corresponding zero of f is not isolated.

Corollary 4. *Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous map on the real line with a fixed point $f(p) = p$. If $d\bar{f}_\mu[p] < 1$, then there is an $\varepsilon > 0$ such that*

$$(\,\forall\, x \in (p-\varepsilon, p+\varepsilon)\,)\; f^n(x) \to p.$$

In particular, this holds if f is differentiable at p and $|f'(p)| < 1$.

The last corollary applies to continuous maps on the real line that have informatic derivatives but do not have classical derivatives [25]. As an application of Theorem 29, we will prove the correctness of Newton's Method *without* using Taylor's Theorem.

Example 33. Let $f : [a,b] \to \mathbb{R}$ be a continuous function with a zero $r \in (a,b)$. If f' is nonzero and continuous on $[a,b]$ and $f''(r)$ exists, we consider the continuous map $I_f : [a,b] \to \mathbb{R}$, given by

$$I_f(x) = x - \frac{f(x)}{f'(x)}.$$

It is easy to see that $I_f(r) = r$. By extending I_f to the real line in any way whatsoever, we appeal to Theorem 28 and obtain

$$\frac{d\bar{I}_f}{d\mu}[r] = 0.$$

By Corollary 4, we see that there is an $\varepsilon > 0$ such that $I_f(x) \to r$ for all $x \in (r-\varepsilon, r+\varepsilon)$.

But what is achieved by avoiding Taylor's theorem? To prove the correctness of Newton's method using Taylor's theorem, we must assume that f'' exists on an *open interval* containing the zero r. The proof we gave in Example 33 assumes only that $f''(r)$ exists. This gives one definite advantage to using Theorem 29 in place of Taylor's theorem: we can prove that Newton's method works on a larger class of functions.

Of course, once we know that an iterative process works *correctly,* the next question inevitably concerns the *rate* at which it works. In classical numerical analysis, the efficiency of an iterative algorithm is determined by calculating its *order of convergence.*

Definition 45. Let (x_n) be a sequence of reals with $x_n \to p$. If

$$0 < \lim_{n\to\infty} \frac{|x_{n+1} - p|}{|x_n - p|^\alpha} = r < \infty,$$

for some $\alpha \geq 1$, then α is called the *order of convergence* of the sequence. If $\alpha = 1$ then r is called the *rate of convergence* of (x_n).

In this definition, the sequence (x_n) is generated by a numerical algorithm designed to calculate p. The larger that α is, the quicker the convergence of (x_n) to p, the better the algorithm.

If $\alpha = 1$, the algorithm is said to converge *linearly.* For $\alpha = 2$, the convergence is *quadratic.* Two linearly convergent algorithms may be compared based on their rates of convergence.

Notice that orders of convergence are calculated using the uncertainty $|x_n - p|$. To extend the idea to the setting of domains with measurements, we consider sequences (x_n) which converge to their suprema p in the μ topology on D, and replace $|x_n - p|$ with $|\mu x_n - \mu p|$.

Definition 46. Let D be a dcpo and let μ measure $X \subseteq D$. If (x_n) is a sequence in D which converges to its supremum $p \in X$ in the μ topology and

$$0 < \lim_{n\to\infty} \frac{\mu x_{n+1} - \mu p}{(\mu x_n - \mu p)^\alpha} = r < \infty,$$

for some $\alpha \geq 1$, then α is called the *order of convergence* of the sequence. If $\alpha = 1$ then r is called the *rate of convergence* of (x_n).

An *increasing* sequence (x_n) converges to its supremum p in the μ topology. We begin with linear processes: the informatic derivative enables the systematic computation of rates of convergence.

Lemma 4. *Let $s : (D,\mu) \rightharpoonup (D,\mu)$ be a partial map which maps into* $\mathrm{dom}(s)$ *and has a fixed point $p = \bigsqcup s^n x$ in the μ topology. If $ds_\mu(p)$ exists, then*

$$\lim_{n\to\infty} \frac{\mu s^{n+1}(x) - \mu p}{\mu s^n(x) - \mu p} = \frac{ds}{d\mu}(p),$$

provided $\mu s^n(x) - \mu p > 0$ for all $n \geq 0$.

Thus, to find the rate at which a linear algorithm s converges to a fixed point p, we find its derivative at p. But why is this a measure of efficiency?

Proposition 5. *Let $s : (D,\mu) \rightharpoonup (D,\mu)$ be a partial map which maps into* $\mathrm{dom}(s)$. *If s is μ continuous at a fixed point p and $0 < ds_\mu(p) < 1$, then for all $0 < \varepsilon < 1 - ds_\mu(p)$, there is an $a \ll p$ such that for all $x \in \mathrm{dom}(s)$,*

$$a \sqsubseteq x \sqsubseteq p \text{ and } n \geq \frac{\log(\varepsilon/(\mu x - \mu p))}{\log\,(ds_\mu(p) + \varepsilon)} \;\Rightarrow\; s^n x \sqsubseteq p \text{ and } |\mu s^n x - \mu p| < \varepsilon,$$

provided $x \neq p$ and $n \geq 1$.

Prop. 5 gives an upper bound on the number of iterations a linear process must do before it achieves ε accuracy. In order that this estimate hold, the input x must be sufficiently close. However, even in the presence of this mathematical annoyance, we can still use it to understand why rate of convergence is a measure of efficiency.

Suppose we have two linear processes s, t which have a common fixed point p and that $0 < ds_\mu(p) < dt_\mu(p) < 1$. Let $\varepsilon > 0$. Imagine we have different inputs for s and t which both have measure λ and that $\lambda - \mu p > \varepsilon$. (If $\lambda - \mu p \leq \varepsilon$, each process is already ε close.) Then

$$\frac{\log(\varepsilon/(\lambda - \mu p))}{\log\left(ds_\mu(p) + \varepsilon\right)} < \frac{\log(\varepsilon/(\lambda - \mu p))}{\log\left(dt_\mu(p) + \varepsilon\right)},$$

that is, the number of iterations which ensure t is ε close to p also guarantee that s is ε close to p. However, it may be that s can achieve ε accuracy with fewer iterations than t. Roughly speaking, s is a better algorithm than t for calculating p.

The estimate on the number of iterations in Prop. 5 is useful because of its generality. However, we often encounter linear processes which satisfy $\mu s(x) - \mu s(p) \leq (ds_\mu(p))(\mu x - \mu p)$ for $x \sqsubseteq p$. In this case, we use the estimate

$$n \geq \frac{\log\left(\varepsilon/(\mu x - \mu p)\right)}{\log ds_\mu(p)}.$$

One question which springs to mind is: How can we know the values of μp and $ds_\mu(p)$ when p itself is unknown? Though we cannot always calculate these quantities independent of p, the estimates given for the number of iterations are still useful for comparing processes, as we saw above. On the other hand, in the case of Newton's Method, we actually know *a priori* that $\mu p = ds_\mu(p) = 0$. We can also calculate these quantities independent of p for the bisection method, the golden section search and for contraction mappings on complete metric spaces.

Example 34. For a continuous map $f : \mathbb{R} \to \mathbb{R}$, the bisection method is captured by the partial splitting

$$\mathrm{split}_f : \mathbf{I}\mathbb{R} \rightharpoonup \mathbf{I}\mathbb{R}$$

and the data

- $\mathrm{dom}(\mathrm{split}_f) = C(f) = \{[a, b] \in \mathbf{I}\mathbb{R} : f(a) \cdot f(b) \leq 0\}$
- $\mathrm{fix}(\mathrm{split}_f) = \{[r] : f(r) = 0\}$
- $d(\mathrm{split}_f)_\mu[r] = 1/2$ for all $[r] \in \mathrm{fix}(\mathrm{split}_f)$

If r is an isolated zero of f, then split_f is μ continuous at the associated fixed point $[r]$. By the remarks following Prop. 5, if r is an isolated zero of f and $x \in C(f)$ is a sufficiently small input around r, then

$$\mathrm{split}_f^n x \text{ for } n \geq \frac{\log(\varepsilon/\mu x)}{\log\left(1/2\right)}$$

is an ε-approximation of r.

The estimate for the number of iterations given in the last example can fail without μ continuity. If we take $f(x) = x \cdot \sin(1/x)$ for $x \neq 0$ and $f(0) = 0$, then there are arbitrarily small intervals $\bar{x} \in C(f)$ with $\bar{x} \sqsubseteq [0]$, but for which $\text{split}_f \bar{x} \not\sqsubseteq [0]$. Beginning with *any* one of these intervals as input, and then doing $n \geq \log(\varepsilon/\mu x)/\log{(1/2)}$ iterations of split_f, leaves an interval of length $< \varepsilon$. The problem is that we are now on track to calculate a *different* zero $[r]$, rather than the one we *intended* to calculate, $[0]$.

The point is this: an estimate for the number of iterations is of little use if we do not know what we are calculating. This is why zeroes are normally assumed isolated in numerical analysis, as in Newton's method, where we assume $f'(r) \neq 0$. Thus, we expect iterative numerical methods to be μ continuous at fixed points when realized as partial maps on domains.

Example 35. The Golden Section Search. In Example 16, given a function $f : \mathbb{R} \to \mathbb{R}$ and a constant $1/2 < r < 1$, we defined the splitting

$$\max_f : \mathbf{I}\mathbb{R} \to \mathbf{I}\mathbb{R}$$

$$max_f[a, b] = \begin{cases} [l(a, b), b] & \text{if } f(l(a, b)) < f(r(a, b)), \\ [a, r(a, b)] & \text{otherwise.} \end{cases}$$

where $l(a, b) = (b - a)(1 - r) + a$ and $r(a, b) = (b - a)r + a$.

If f is unimodal on $[a, b]$ and its unique maximizer is $x^* \in \text{int}[a, b]$, then $\max_f$ is μ continuous at $[x^*]$ because

$$[a, b] \ll \bar{x} \sqsubseteq [x^*] \Rightarrow \max_f \bar{x} \sqsubseteq [x^*],$$

which was shown in Example 16, and because it has a derivative at $[x^*]$, given by

$$\frac{d(\max_f)}{d\mu}[x^*] = r.$$

Thus, if f is unimodal on $[a, b]$ and $x^* \in \text{int}[a, b]$, then

$$\max_f^n[a, b] \text{ for } n \geq \frac{\log(\varepsilon/(b - a))}{\log(r)}$$

is an ε-approximation of x^*.

Example 36. Contraction maps. If $f : X \to X$ is a contraction on a complete metric space (X, d) with constant $0 < c < 1$, its extension to the formal ball model

$$\bar{f} : \mathbf{B}X \to \mathbf{B}X, \quad \bar{f}(x, r) = (fx, c \cdot r)$$

has derivative $d\bar{f}_\pi(p) = c$, for all $p \in \mathbf{B}X$. The map $\bar{f}$ is Scott continuous and hence μ continuous at all points. If we take any $x \in X$ and $r \geq d(x, fx)/(1-c)$, then

$$\bar{f}^{\,n}(x, r) \text{ for } n \geq \frac{\log(r/\varepsilon)}{\log c}$$

is an ε-approximation of the unique attractor of f.

The presence of informatic linearity in the last three examples enables us to use the estimate mentioned after Prop. 5. The next example is more interesting.

Example 37. The Regula Falsi Method. For a function $f : [a,b] \to \mathbb{R}$ such that

(i) $f(a) < 0$ and $f(b) > 0$,
(ii) $f'(x) > 0$ for all $x \in [a,b]$, and
(iii) $f''(x) \geq 0$ for all $x \in [a,b]$,

we define the partial mapping

$$r_f : \mathbf{I}\mathbb{R} \rightharpoonup \mathbf{I}\mathbb{R}$$

$$r_f[x,b] = \left[b - f(b) \left(\frac{b-x}{f(b)-f(x)} \right), b \right]$$

whose domain is

$$\mathrm{dom}(r_f) = \{[x,b] : a \leq x \leq r\}$$

where $r \in (a,b)$ is the unique zero of f on $[a,b]$.

The map r_f is a Scott continuous splitting which maps the dcpo $\mathrm{dom}(r_f)$ into itself. For if $a \leq x \leq y \leq r$, we have the string of inequalities

$$a \leq x \leq b - f(b) \left(\frac{b-x}{f(b)-f(x)} \right) \leq b - f(b) \left(\frac{b-y}{f(b)-f(y)} \right) \leq r,$$

where the second follows from $f(x) \leq 0$, and the last two follow from

$$\frac{f(b)-f(x)}{b-x} \leq \frac{f(b)-f(y)}{b-y} \leq \frac{f(b)-f(r)}{b-r},$$

which is a consequence of the fact that f' is nondecreasing. This proves that r_f is a monotone splitting which takes $\mathrm{dom}(r_f)$ into itself. Finally, r_f is Scott continuous because its measure is Scott continuous.

By Proposition 7, if $\bar{x} \in \mathrm{dom}(r_f)$, then

$$\bigsqcup_{n \geq 0} r_f^n(\bar{x}) \in \mathrm{fix}(r_f),$$

but it is easy to see that $\mathrm{fix}(r_f) = \{[r,b]\}$. Thus, iterating r_f is an algorithm for approximating r, called the *Regula Falsi method.* But how efficient is it?

To answer this question, we calculate the informatic derivative of r_f at the fixed point $[r,b]$ as follows:

$$\begin{aligned}\frac{dr_f}{d\mu}[r,b] &= \lim_{\bar{x}\to[r,b]} \frac{\mu r_f(\bar{x}) - \mu r_f[r,b]}{\mu\bar{x} - \mu[r,b]}\\ &= \lim_{x\to r^-} \frac{f(b)(r-x) + f(x)(b-r)}{(f(b)-f(x))(r-x)}\\ &= \lim_{x\to r^-} \left[\frac{f(b)}{f(b)-f(x)} + \frac{f(x)-f(r)}{r-x}\cdot\frac{b-r}{f(b)-f(x)}\right]\\ &= \frac{f(b)}{f(b)-f(r)} + (-1)f'(r)\cdot\frac{b-r}{f(b)-f(r)}\\ &= 1 - \frac{f'(r)(b-r)}{f(b)}.\end{aligned}$$

By monotonicity of r_f, this derivative is nonnegative, and hence a number in the interval $[0,1)$. In fact, we can see that

$$d(r_f)_\mu[r,b] \to 0 \quad \text{as} \quad b \to r$$

so the efficiency of this algorithm is determined by the closeness of b to r. Notice that it does *not* depend on a.

Once we have the derivatives of two different algorithms which solve the same problem, we can compare them to understand their respective strengths and weaknesses.

Example 38. The Bisection versus Regula Falsi. If $f : \mathbb{R} \to \mathbb{R}$ is a continuous map and $[a,b]$ is an interval such that $f(a) < 0$ and $f(b) > 0$, $f' > 0$ on $[a,b]$ and $f'' \geq 0$ on $[a,b]$, then

$$\bigsqcup_{n\geq 0} \mathrm{split}_f^n[a,b] = [r] \quad \text{and} \quad \bigsqcup_{n\geq 0} r_f^n[a,b] = [r,b]$$

are both schemes for calculating the unique zero r of f on $[a,b]$. But which one is better? We consider two examples.

If $f(x) = x^2 - x - 1$ and $[a,b] = [1,2]$, then $r = (1+\sqrt{5})/2$. Thus,

$$d(\mathrm{split}_f)[r] = \frac{1}{2} \quad \text{and} \quad d(r_f)[r,b] = \frac{7-3\sqrt{5}}{2} \approx 0.145898,$$

which means that *eventually* $\mu r_f(x) - \mu[r,b] \approx 0.14(\mu x - \mu[r,b])$, as compared to $\mu\,\mathrm{split}_f(x) - \mu[r] = 0.5(\mu x - \mu[r])$ for the bisection. In other words, *eventually* the Regula Falsi method reduces the uncertainty in an interval by about 86%, while for the bisection uncertainty is always reduced by 50%. This suggests that r_f is preferable in this case. Six iterations of each gives

$$\mathrm{split}_f^6[1,2] = [1.59375, 1.625] \quad \text{and} \quad r_f^6[1,2] \approx [1.618025, 2].$$

The approximation of r offered by the bisection is the midpoint of $\mathrm{split}_f^6[1,2]$, 1.609375, while the approximation given by the Regula Falsi method is the left endpoint of $r_f^6[1,2]$, around 1.618025. Thus, the Regula-Falsi method is accurate to four decimal places, while the bisection is only accurate to one. This supports the intuition offered by the informatic derivatives calculated above: r_f converges faster than split_f in this case.

If $f(x) = x^6 - x - 1$ and $[a,b] = [1,2]$, then $r \approx 1.13472$. The informatic derivatives in this case are

$$d(\text{split}_f)[r] = \frac{1}{2} \text{ and } d(r_f)[r,b] \approx 0.85407,$$

which suggests that now it is split_f which converges faster. If we do sixteen iterations of each, we find that

$$\text{split}_f^{16}[1,2] \approx [1.134719, 1.134735] \text{ and } r_f^{16}[1,2] \approx [1.121308, 2].$$

Thus, the bisection gives the approximation $r \approx 1.13472$, while the Regula Falsi method is only accurate to one decimal place. In fact, it is only after 68 iterations that the Regula Falsi method can duplicate what the bisection achieves in 16:

$$r_f^{68}[1,2] \approx [1.13472, 2].$$

The intuition imparted by informatic derivative is also correct in this instance.

Example 39. The secant method. Recall from Theorem 8, the *secant method* $\sec_f : \mathcal{P}_C[a,r] \to \mathcal{P}_C[a,r]$ given by

$$\sec_f[x,y] = \left[y, y - \frac{f(y)}{df[x,y]}\right]$$

yields an algorithm for calculating r with $f(r) = 0$ given by

$$\bigsqcup_{n \geq 0} \sec_f^n(x) = [r],$$

for any $x \in \mathcal{P}_C[a,r]$. For the secant method $\sec_f$, we have $d(\sec_f)[r] = 0$.

Let $\bar{x} = [x,y] \sqsubseteq [r]$ with $\mu\bar{x} > 0$. Then by the mean value theorem and the triangle inequality,

$$0 \leq \frac{\mu \sec_f(\bar{x})}{\mu\bar{x}} \leq \frac{2(r-y)}{r-x+r-y} + \frac{|f(y)|}{f'(c)(r-x+r-y)},$$

where $c \in \bar{x}$. But since $r - x + r - y \geq 2(r-y)$ and $r - x + r - y \geq r - y$, the expression on the right is bounded by

$$\frac{2(r-y)}{2(r-y)} + \frac{|f(y)|}{f'(c)(r-y)} = 1 - \frac{|f(y) - f(r)|}{f'(c)(y-r)}.$$

As $\bar{x} \to [r]$ in the μ topology, we have $x, y \to r$ and $c \to r$. Hence,

$$0 \leq \lim_{\bar{x} \to [r]} \frac{\mu \sec_f(\bar{x})}{\mu\bar{x}} \leq \lim_{c,y \to r} \left(1 - \frac{|f(y) - f(r)|}{f'(c)(y-r)}\right) = 1 - 1 = 0,$$

proving the claim.

Thus, the convergence of the secant method is *superlinear,* in agreement with numerical analysis. This is an interesting example. The function $\sec_f$ does *not* correspond to iterating a classical real valued function, and the informatic derivative is *not* a classical derivative: the formula in Example 21 takes *two* real numbers as input, but returns only *one* as output.

Thus, to prove that a numerical method works correctly, we show it iterates to a fixed point. To go along with this uniform approach to the problem of correctness, we now have a uniform method for calculating rates of convergence of linear processes: simply take the informatic derivative of a map on a domain at a fixed point. This extends what is done is numerical analysis, enabling a unified treatment not previously possible. For instance, the secant method, the golden section search and the bisection method are iterative processes which have no classical descriptions as differentiable functions on the real line. Nevertheless, we have seen that they may be naturally described as mappings on domains which possess informatic derivatives.

5.3 Rates of change in the communication process

A classical binary channel $f : \Delta^2 \to \Delta^2$ takes an input distribution to an output distribution. In a similar way, a qubit channel is a function of the form $\varepsilon : \Omega^2 \to \Omega^2$ that is convex linear and completely positive [35]. For our purposes, there is no need to get lost in too many details of the Hilbert space formulation: qubit channels can be represented as linear selfmaps on the unit ball in Euclidean three space as follows.

There is a 1-1 correspondence between density operators on a two dimensional state space and points on the unit ball $\mathbb{B}^3 = \{x \in \mathbb{R}^3 : |x| \leq 1\}$: each density operator $\rho : \mathcal{H}^2 \to \mathcal{H}^2$ can be written uniquely as

$$\rho = \frac{1}{2}\begin{pmatrix} 1 + r_z & r_x - ir_y \\ r_x + ir_y & 1 - r_z \end{pmatrix}$$

where $r = (r_x, r_y, r_z) \in \mathbb{R}^3$ satisfies $|r| = \sqrt{r_x^2 + r_y^2 + r_z^2} \leq 1$. The vector $r \in \mathbb{B}^3$ is called the *Bloch vector* associated to ρ. Bloch vectors have a number of aesthetically pleasing properties.

If ρ and σ are density operators with respective Bloch vectors r and s, then (i) the eigenvalues of ρ are $(1 \pm |r|)/2$, (ii) the von Neumann entropy of ρ is $S\rho = H((1 + |r|)/2) = H((1 - |r|)/2)$, where $H : [0, 1] \to [0, 1]$ is the base two Shannon entropy, (iii) if ρ and σ are pure states and $r + s = 0$, then ρ and σ are orthogonal, and thus form a basis for the state space; conversely, the Bloch vectors associated to a pair of orthogonal pure states form antipodal points on the sphere, (iv) the Bloch vector for a convex sum of mixed states is the convex sum of the Bloch vectors, (v) the Bloch vector for the completely mixed state $I/2$ is $0 = (0, 0, 0)$.

Because of the correspondence between Ω^2 and $\mathbb{B}^3$, let us now regard these two as equal.

A standard way of measuring the capacity of a quantum channel in quantum information is the Holevo capacity; it is sometimes called the product state capacity since input states are not allowed to be entangled across two or more uses of the channel.

Definition 47. For a quantum channel f, the *Holevo capacity* is given by

$$\mathrm{C}(f) = \sup_{\{x_i,\rho_i\}} \left[S\left(f\left(\sum_i x_i \rho_i \right) \right) - \sum_i x_i \cdot S(f(\rho_i)) \right]$$

where the supremum is taken over all ensembles $\{x_i, \rho_i\}$ of possible input states ρ_i to the channel.

The possible input states ρ_i to the channel are in general mixed and the x_i are probabilities with $\sum_i x_i = 1$. If f is the Bloch representation of a qubit channel, the Holevo capacity of f is given by

$$\mathrm{C}(f) = \sup_{\{x_i,r_i\}} \left[H\left(\frac{1 + |f\left(\sum_i x_i r_i\right)|}{2} \right) - \sum_i x_i \cdot H\left(\frac{1 + |f(r_i)|}{2} \right) \right]$$

where r_i are Bloch vectors for density operators in an ensemble, and we recall that eigenvalues of a density operator with Bloch vector r are $(1 \pm |r|)/2$.

Recall that the classical channels $f : \Delta^2 \to \Delta^2$ which increase entropy $(H(f(x)) \geq H(x))$ are exactly those f with $f(\bot) = \bot$. They are the *strict* mappings of domain theory, which are also known as *binary symmetric channels* in information theory. Similarly, the entropy increasing qubit channels are exactly those f for which $f(\bot) = \bot$. These are called *unital* in quantum information theory.

Theorem 30. *Let $\mu(x) = 1 - |x|$ denote the standard measurement on Ω^2. For any unital channel f and any $p \in \Omega^2$ different from $\bot$,*

$$df_\mu(p) = \frac{|f(p)|}{|p|}$$

Thus, the Holevo capacity of f is determined by the largest value of its informatic derivative. Explicitly,

$$\mathrm{C}(f) = 1 - H\left(\frac{1}{2} + \frac{1}{2} \sup_{x \in \ker(\mu)} df_\mu(x) \right)$$

Then $\mathrm{C}(f) = 1$ for *any* rotation f since $df_\mu = 1$. Notice that $df_\mu \equiv 1$ iff f is a rotation. For each $p \in [0,1]$, the unique channel $f \sqsubseteq 1$ with $df_\mu = p$ is the depolarization channel $f = d_p = p \cdot I$, so that $\mathrm{C}(d_p) = 1 - H((1 + p)/2)$. In fact, there is an isomorphism from binary symmetric channels onto the depolarization channels. The only unital qubit channel with capacity zero is 0 itself.

Example 40. The two Pauli channel in Bloch form is

$$\varepsilon(r) = p\,r + \left(\frac{1-p}{2}\right) s_x(r) + \left(\frac{1-p}{2}\right) s_y(r)$$

where s_x and s_y are the Bloch representations of the unitary channels derived from the Pauli spin operators σ_x and σ_y. This simplifies to

$$\varepsilon(r_x, r_y, r_z) = (pr_x, pr_y, -(1-p)r_z)$$

The matrix associated to ε is diagonal, so the diagonal element (eigenvalue) that has largest magnitude also yields the largest value of its informatic derivative. The capacity of the two Pauli channel is then

$$1 - H\left(\frac{1+\max\{p, 1-p\}}{2}\right)$$

where $p \in [0,1]$.

The set of unital channels $\mathcal{U}$ is compact hence closed and thus forms a dcpo as a subset of the domain $[\Omega^2 \to \Omega^2]$.

Corollary 5. *The Holevo capacity* $\mathrm{C} : \mathcal{U} \to [0,1]$ *is Scott continuous.*

Thus, the ability of a unital qubit channel to transmit information is determined by the largest value of its informatic derivative.

5.4 The derivative at a compact element: a discrete derivative.

If one looks closely at the definition of the informatic derivative above, it has a computationally restrictive aspect: the requirement that p not be isolated in the μ topology. This is equivalent to saying that p must not be a *compact* element of D. From the mathematical viewpoint, one does not object to this: mathematics offers us no way of obtaining unique 'limits' at isolated points of topological spaces. Nevertheless, computationally, it is easy to write down simple examples of mappings on domains which *should have* derivatives, but are excluded simply because they work only with compact elements.

For instance, on the domain of lists $[S]$, the map rest: $[S] \to [S]$ which removes the first element from a nonempty list and sends the empty list to itself, satisfies

$$\mu\,\mathrm{rest}(x) = \mu(x) - 1$$

for $x \neq [\,]$, where μ is the length measurement. Thus, we ought to be able to say that $d(\mathrm{rest})_\mu(x) = 1$ for $x \neq [\,]$.

We now consider an extension of the definition of informatic derivative which applies at compact elements as long as they are not minimal. One of the benefits of this extension is that we are finally able to understand the sense in which the asymptotic notions of complexity used in numerical analysis (rates of convergence) are the same as those used in the analysis of 'discrete' algorithms (for example, list processing). Another is the identification of an idea which allows us to systematically calculate both of these complexity notions in a uniform manner: informatic rates of change apply in both the continuous and discrete realms.

The informatic derivative at a compact element

Defining the informatic derivative of a selfmap on a domain D really only depends on our ability to define it for functions of the form $f : D \to \mathbb{R}$. If we set

$$df_\mu(p) = \lim_{x \to p} \frac{f(x) - f(p)}{\mu x - \mu p}$$

then for $f : D \to D$, we can set $df_\mu(p) = d(\mu f)_\mu(p)$, obtaining the usual definition of the informatic derivative. Of course, the problem is that this is only works when p is not compact i.e. when

$$p \notin K(D) := \{x \in D : x \ll x\}$$

These are precisely the points that are not isolated in the μ topology. The reason we must work with points which are not isolated is that there must be enough *nontrivial* μ open sets around p so that we can take a limit in the formal sense of topology – without enough nontrivial open sets, a limit may not be unique.

However, any point $p \notin \min(D) := \{x \in D : {\downarrow}x = \{x\}\}$ can be approximated from below using the nontrivial μ open subsets of D which are contained in ${\downarrow}p$ and which themselves contain p and at least one other element:

$$\text{approx}_\mu(p) = \{V \in \mu_D : p \in V \subseteq {\downarrow}p \text{ and } V \neq \{p\}\}.$$

Thus, the existence of approximations is not the problem – the problem is that we need a concept more applicable than 'limit'.

Definition 48. Let $f : D \to \mathbb{R}$ be a function and $p \in D$. We set

$$d^+ f_\mu(p) := \sup\{c : (\exists V \in \text{approx}_\mu(p))(\forall x \in V)\, f(x) - f(p) \geq c \cdot (\mu x - \mu p)\}$$

and

$$d^- f_\mu(p) := \inf\{c : (\exists V \in \text{approx}_\mu(p))(\forall x \in V)\, f(x) - f(p) \leq c \cdot (\mu x - \mu p)\},$$

provided p is not a *minimal element* of D, i.e., $p \notin \min(D)$.

The existence of the informatic derivative of a real-valued function in the usual case is expressible entirely in terms of $d^+ f_\mu$ and $d^- f_\mu$ as follows:

Theorem 31. *Let $f : D \to \mathbb{R}$ be a function with $p \in D \setminus K(D)$. Then $df_\mu(p)$ exists iff $d^+ f_\mu(p)$ exists, $d^- f_\mu(p)$ exists and $d^- f_\mu(p) \leq d^+ f_\mu(p)$. In either case, we have $df_\mu(p) = d^+ f_\mu(p) = d^- f_\mu(p)$.*

The previous theorem justifies the following definition.

Definition 49. Let $f : D \to \mathbb{R}$ be a function on a continuous dcpo D with a measurement μ which measures D at $p \in D \setminus \min(D)$. If $d^- f_\mu(p)$ exists, $d^+ f_\mu(p)$ exists and $d^- f_\mu(p) \leq d^+ f_\mu(p)$, then we define

$$df_\mu(p) := d^+ f_\mu(p)$$

and call this number the *informatic derivative* of f at p.

By Theorem 31, the new definition and the old definition agree in the continuous case ($p \notin K(D)$). We now turn our attention to the discrete case ($p \in K(D)$).

Theorem 32. *Let $f : D \to \mathbb{R}$ be a function on an algebraic dcpo D with a measurement μ that measures D at $p \in K(D) \setminus \min(D)$. Then the following are equivalent:*

(i) *The derivative $df_\mu(p)$ exists.*
(ii) *The supremum*

$$\sup\left\{\frac{f(x) - f(p)}{\mu x - \mu p} : x \in K(D) \cap \downarrow\!p, x \neq p\right\}$$

exists and the infimum

$$\inf\left\{\frac{f(x) - f(p)}{\mu x - \mu p} : x \in K(D) \cap \downarrow\!p, x \neq p\right\}$$

exists.

In either case, the value of $d^+ f_\mu(p)$ is the supremum in (ii)*, while the value of $d^- f_\mu(p)$ is the infimum in* (ii).

Finally, the definition of derivative for selfmaps on a domain D.

Definition 50. Let $f : D \to D$ be a function on a domain (D, μ) with a map μ that measures D at $p \in D \setminus \min(D)$. If $d(\mu f)_\mu(p)$ exists, then we write

$$df_\mu(p) := d(\mu f)_\mu(p)$$

and call this number the *informatic derivative* of f at p with respect to μ. We also set $d^* f_\mu(p) := d^*(\mu f)_\mu(p)$ for $* \in \{+, -\}$.

It is easy to extend this definition for a map $f : (D, \mu) \to (E, \lambda)$, as was done for the original formulation of the derivative in the continuous case [15], but in the present paper there are no applications warranting such an abstraction.

Example 41. Derivatives of list operations.

(i) The map first : $[S] \to [S]$, $\mathrm{first}(a :: x) = [a]$, $\mathrm{first}[\,] = [\,]$. Using Theorem 32,

$$d(\mathrm{first})_\mu(x) = d^+(\mathrm{first})_\mu(x) = d^-(\mathrm{first})_\mu(x) = 0,$$

for all $x \neq [\,]$. At $x = [\,]$, $d(\mathrm{first})_\mu(x) = d^+(\mathrm{first})_\mu(x) = 1 \geq 0 = d^-(\mathrm{first})_\mu(x)$.

(ii) The map rest : $[S] \to [S]$, $\mathrm{rest}(a :: x) = x$, $\mathrm{rest}[\,] = [\,]$. Using Theorem 32,

$$d(\mathrm{rest})_\mu(x) = d^+(\mathrm{rest})_\mu(x) = d^-(\mathrm{rest})_\mu(x) = 1,$$

for all $x \neq [\,]$. At $x = [\,]$, $d(\mathrm{rest})_\mu(x) = d^+(\mathrm{rest})_\mu(x) = 1 \geq 0 = d^-(\mathrm{rest})_\mu(x)$.

There is something worth pointing out before we focus on the derivative in the discrete case. The definition of $df_\mu(p)$ splits into two cases, the continuous ($p \notin K(D)$) and the discrete ($p \in K(D)$). From this bifurcation appears a remarkable duality: In the continuous case the inequality $df^+_\mu(p) \leq df^-_\mu(p)$ always holds, but $df^-_\mu(p) \leq df^+_\mu(p)$ may not; in the discrete case the opposite is true, $df^-_\mu(p) \leq df^+_\mu(p)$ always holds, but $df^+_\mu(p) \leq df^-_\mu(p)$ may not.

The results of this section allow for only one interpretation of this phenomenon: In the continuous case, the derivative is determined by *local* properties of the function; in the discrete case, the derivative is determined by *global* properties of the function.

Measuring the length of an orbit

Throughout this section, we assume that (D, μ) is an algebraic dcpo whose compact elements $K(D)$ form a lower set $K(D) = \downarrow K(D)$. Some important examples of this are $\mathbb{N}^*$, $\mathbb{N}^\infty = \mathbb{N} \cup \{\infty\}$, $[S]$, $\mathcal{P}\omega$, Σ^∞, and $[\mathbb{N} \rightharpoonup \mathbb{N}]$. Computationally, this is not much of an assumption.

Theorem 33 (The Mean Value Theorem). *Let $f : D \to D$ be a function on (D, μ) such that $df_\mu(p)$ exists at a compact element p. Then*

$$(\mu x - \mu p) \cdot d^- f_\mu(p) \leq \mu f(x) - \mu f(p) \leq d^+ f_\mu(p) \cdot (\mu x - \mu p),$$

for all $x \sqsubseteq p$.

If a splitting r has a compact fixed point p reachable by iteration $\bigsqcup r^n(x) = p$, then the derivative of r at p can be used to provide a precise measure of the number of iterations required to get to p from an input of x. Later we will see that such quantities can play an integral role in determining the complexity of certain algorithms.

Definition 51. Let $r : D \to D$ be a splitting. An *orbit* is a sequence of iterates $(r^n x)$. An orbit is *compact* if

$$\bigsqcup_{n \geq 0} r^n(x) \in K(D).$$

The *length* of a compact orbit $(r^n x)$ is

$$|(r^n x)| := \inf\{n \geq 0 : r^{n+1}(x) = r^n(x)\}.$$

A compact orbit is *nontrivial* when $|(r^n x)| > 0$; otherwise it is a *fixed point.*

In this new language, we can say that we are interested in determining the length of nontrivial compact orbits of splittings. If $(r^n x)$ is a compact orbit, then $r^l(x)$ is a fixed point of r where $l = |(r^n x)|$. For this reason, we say that the orbit $(r^n x)$ *ends* at $p = r^l(x)$.

Lemma 5. *If a splitting $r : D \to D$ has a nontrivial compact orbit which ends at $p \in K(D)$, and $dr_\mu(p)$ exists, then $0 \leq dr_\mu(p) \leq 1$.*

Theorem 34. *Let r be a splitting with a nontrivial compact orbit $(r^n x)$ that ends at p. If $dr_\mu(p) = 0$, then $r(x) = p$. If $0 < dr_\mu(p) < 1$, then*

$$n \geq \left\lceil \frac{\log((\mu x - \mu p)/\varepsilon)}{\log(1/dr_\mu(p))} \right\rceil + 1 \Rightarrow |\mu r^n(x) - \mu p| < \varepsilon,$$

for any $\varepsilon > 0$.

By the compactness of p, there is a choice of $\varepsilon > 0$ which will ensure that $|\mu r^n(x) - \mu p| < \varepsilon \Rightarrow r^n(x) = p$, but at this level of generality we cannot give a precise description of it. It depends on μ. For lists, the value is $\varepsilon = 1$.

Example 42. Let r be a splitting on $[S]$ with $0 < dr_\mu(p) < 1$ at any fixed point p. Then for any x, there is some $k \geq 0$ such that $r^k(x) = p$ is a fixed point. By the last result, doing

$$n > \left\lceil \frac{\log(\mu x - \mu p)}{\log(1/dr_\mu(p))} \right\rceil$$

iterations implies that $r^n(x) = p$.

Let's consider an important example of this type.

Example 43. Contractive list operations. For a positive integer $x > 0$, define

$$m(x) = \begin{cases} x/2 & \text{if } x \text{ even;} \\ (x+1)/2 & \text{if } x \text{ odd.} \end{cases}$$

Consider the splittings

$$\text{left}(x) = [x(1), \cdots, x(m(\mu x) - 1)]$$

$$\text{right}(x) = [x(m(\mu x) + 1), \cdots, x(\mu x)]$$

each of which takes lists of length one or less to the empty list $[\,]$. Each has a derivative at its unique fixed point $[\,]$ as follows.

First, since both of these maps are splittings and $p = [\,]$ has measure $\mu p = 0$, each has a derivative at p – it is simply a matter of determining d^+ at $[\,]$ in each case. For this, if $x \neq [\,]$, then

$$\frac{\mu\,\text{left}(x)}{\mu x} \leq \frac{(\mu x/2) - (1/2)}{\mu x} = \frac{1}{2} \cdot \left(1 - \frac{1}{\mu x}\right) \leq \frac{1}{2}$$

$$\frac{\mu\,\text{right}(x)}{\mu x} \leq \frac{\mu x/2}{\mu x} = \frac{1}{2}$$

which means $d(\text{left})_\mu[\,] = d(\text{right})_\mu[\,] = 1/2$.

Notice that the case of ‘left’ is much more interesting than the case of ‘right.’ In the former, the value of the derivative is never attained by any of the quotients $\mu\,\text{left}/\mu$ – it is determined by a ‘limit’ process which extracts global information about the mapping left.

Already we notice a relationship to processes in numerical analysis: the case $dr_\mu(p) = 0$ is an extreme form of *superlinear* convergence (extreme since in one iteration the computation finishes), while the case $0 < dr_\mu(p) < 1$ behaves just like ordinary *linear* convergence. However, unlike numerical analysis, we can actually say something about the case $dr_\mu(p) = 1$.

To do this is nontrivial, and in what follows, we seek only to illustrate the value of the informatic derivative in the discrete case by showing that the precise number of iterations required to calculate a fixed point p by iteration of a map r can be determined when $dr_\mu(p) = 1$ – a case in which classical derivatives are notorious for yielding no information.

A compact element p that is not minimal has a natural set of *predecessors,* these are formally defined as the set of maximal elements in the dcpo $\downarrow\!p \setminus \{p\}$:

$$\text{pred}(p) = \max(\downarrow\!p \setminus \{p\}).$$

To see that this makes sense, notice that $\downarrow\!p \setminus \{p\}$ is nonempty since p is not minimal, and is closed in the μ topology, as the intersection of μ closed sets. But a μ closed set is closed under directed suprema, and so must have at least one maximal element.

Theorem 35. *Let $r : D \to D$ be a splitting on (D, μ) with a compact fixed point $p = r(p)$ such that*

$$(\forall x)\, x \sqsubseteq p \Rightarrow \bigsqcup_{n \geq 0} r^n(x) = p.$$

If $d^+ r_\mu(x) = 1$ for all $x \sqsubseteq p$ and $d^- r_\mu(x) = 1$ for all $x \sqsubseteq p$ with $x \neq p$, then for all $x \sqsubseteq p$ with $x \neq p$, there is $q \in \mathrm{pred}(p)$ such that

$$r^n(x) = p \Leftrightarrow n = \frac{\mu x - \mu p}{\mu q - \mu p}.$$

It is interesting to notice in the last result that if $d^- r_\mu(p) = 1$, then we *must* have $r(x) = x$ for all $x \sqsubseteq p$. Of course, our hypotheses on r rule this out since the fixed point p must be an attractor on $\downarrow\! p$.

Example 44. In Example 41, we saw that the map rest : $[S] \to [S]$ is an example of the sort hypothesized in Theorem 35 with $p = [\,]$. The predecessors of p are the one element lists

$$\mathrm{pred}(p) = \{[x] : x \in S\}.$$

Thus, the last theorem says that

$$\mathrm{rest}^n(x) = [\,] \Leftrightarrow n = \mu x,$$

for any $x \neq [\,]$.

Complexity

We briefly consider how the informatic derivative offers a new perspective on the complexity of algorithms.

Example 45. Linear search. To search a list x for a key k consider

$$\mathrm{search} : [S] \times S \to \{\perp, \top\}$$

given by

$$\begin{array}{lll} \mathrm{search}([\,], k) & = \perp & \\ \mathrm{search}(x, k) & = \top & \text{if first } x = k, \\ \mathrm{search}(x, k) & = \mathrm{search}(\mathrm{rest}\, x, k) & \text{otherwise.} \end{array}$$

Let $D = [S] \times S^\flat$ – the product of $[S]$ with the set S ordered flatly. We measure this domain as $\mu(x, k) = \mu x$. Let $r : D \to D$ be the splitting $r(x, k) = (\mathrm{rest}\, x, k)$.

On input (x, k) in the *worst case,* the number of comparisons n done by this algorithm is the same as the number of iterations needed to compute

$r^n(x,k) = ([\,],k)$. Since $d^+r_\mu(x) = 1$ for all x and $d^-r_\mu(x) = 1$ for all $x \neq ([\,],k)$, Theorem 35 applies to give

$$r^n(x,k) = ([\,],k) \Leftrightarrow n = \mu(x,k) = \mu x,$$

which helps us understand how the complexity of a discrete algorithm can be determined by the derivative of a splitting which models its iteration mechanism.

Example 46. Binary search. To search a sorted list x for a key k, we use

$$\mathrm{bin} : [S] \times S \to \{\perp, \top\}$$

given by

$$\begin{array}{lll} \mathrm{bin}([\,],k) & = \perp & \\ \mathrm{bin}(x,k) & = \top & \text{if } \mathrm{mid}\, x = k, \\ \mathrm{bin}(x,k) & = \mathrm{bin}(\mathrm{left}\, x,k) & \text{if } \mathrm{mid}\, x > k, \\ \mathrm{bin}(x,k) & = \mathrm{bin}(\mathrm{right}\, x,k) & \text{otherwise.} \end{array}$$

where $\mathrm{mid}\, x := x(m(\mu x))$. Again $D = [S] \times S^\flat$ and $\mu(x,k) = \mu x$. This time we consider the splitting $r : D \to D$ by

$$r(x,k) = \begin{cases} (\mathrm{left}\, x,k) & \text{if } \mathrm{mid}\, x > k; \\ (\mathrm{right}\, x,k) & \text{otherwise.} \end{cases}$$

On input (x,k) in the *worst case,* the number of comparisons n must satisfy $r^n(x,k) = ([\,],k)$. In this case, we have $dr_\mu([\,],k) = 1/2$, so by Theorem 34,

$$n \leq \left\lceil \frac{\log(\mu x)}{\log(2)} \right\rceil + 1 = \lceil \log_2(\mu x) \rceil + 1,$$

since we know that the expression on the right is a number m that satisfies $r^m(x,k) = ([\,],k)$ but that n is the *least* of all such natural numbers because it was produced by the algorithm bin.

To summarize these simple examples: we have two different algorithms which solve the same problem recursively by iterating splittings r and s, respectively, on a domain (D,μ) in an effort to compute a fixed point p. If $dr_\mu(p) < ds_\mu(p)$, then the algorithm using r is faster than the one which uses s. In the case of linear search we have $ds_\mu(p) = 1$, while for binary search we have $dr_\mu(p) = 1/2$. As we have already seen, this is identical to the way one compares zero finding methods in numerical analysis – by comparing the derivatives of mappings at fixed points.

Thoughts on the discrete derivative

Theorem 31 is crucial in that it characterizes differentiability *independent* of its continuous component. Taking only this result as motivation for the definition of derivative leaves a few distinct possibilities. For instance, if we had called the derivative the interval $[d^- f_\mu(p), d^+ f_\mu(p)]$, we might notice more clearly the tendency of continuous information to collapse at a point. Another possibility is to say that the derivative is $d^- f_\mu(p)$. The author chose $d^+ f_\mu$ because it makes the most sense from an applied perspective. As an illustration, consider the intuitions we have about it: algorithms r with $dr_\mu(p) = 0$ belong to $O(1)$, those with $0 < dr_\mu(p) < 1$ belong to $O(\log n)$, while $dr_\mu(p) = 1$ indicates a process is in $O(n)$.

At first glance, an extension of the informatic derivative to the case of discrete data (compact elements) seems like an absurd idea. To begin, we have to confront the issue of essentially defining unique limits at isolated points. But even if we assume we have this, we need the new formulation to extend the previous, which means spotting a relationship between limits in the continuous realm versus finite sequences of discrete objects. But the truth is that all of this only sounds difficult because of what we are *taught*: that the continuous and discrete are 'fundamentally different' and that one of the crucial distinctions between the two is the sensibility of the limit concept for continuous objects, as compared to the discrete case where 'limit' has no meaning. From this, we conclude that math students should spend less time attending lectures and more time coming up with new ideas.

The existence of a derivative in the discrete case means much more than it does in the continuous case. Most results on discrete derivatives do not hold in the continuous case. Just consider a quick example: let $r : D \to D$ be any continuous map with $p = r(p) \in K(D)$ and $dr_\mu(p) = 0$. If $x \sqsubseteq p$, then $r(x) = p$. Now compare this to the continuous case (like calculus on the real line), where one can only conclude that there is an $a \ll p$ such that $r^n(x) \to p$ for all x with $a \ll x \sqsubseteq p$. Again, this sharp contrast is due to the fact that discrete derivatives make use of *global* information, while continuous derivatives use only *local* information. Nevertheless, each is an instance of a common theme.

6 Forms of process evolution

6.1 Intuition

The idea in the measurement formalism is to analyze *processes*: a process is a thing that evolves in a space of informatic objects. The space of informatic objects is formally described by a domain with a measurement. By contrast, the measurement formalism allows for considerable flexibility in formalizing the notion of process. We have already seen one such notion of process: a function $f : D \to D$ that on input x produces iterates $(f^n(x))$ which converge to a fixed point $\bigsqcup f^n(x)$. This discrete form of evolution has various generalizations within the measurement formalism. We consider a few more of them in this section. The renee equation, which is a discrete extension of iteration, can be used to define recursion on general domains while still maintaining a first order view of evolution. The *trajectory* leads one to daydream about a *kinematics* of computation: for instance, the complexity of an algorithm is the amount of time it takes its trajectory in informatic space to achieve its order theoretic maximum. Lastly, we consider a third notion of process, one grounded on a 'thermodynamical' view of evolution, very different from the first two. The basic idea is this: before a process evolves, there are several possible states it may evolve to; when it finishes evolving, we gain information, but the acquisition of information is not free – how much does it cost?

Major references: [15, 22, 28, 29]

6.2 The renee equation

The renee equation is a model of recursion. After introducing this equation, we discuss two major results. The first is that every renee equation has a *unique* solution. The second is that the partial and primitive recursive functions on the naturals may be captured by taking closure under renee equations on the domains $\mathbb{N}^\infty$ and $\mathbb{N}^*$ – the naturals in their usual and opposite orders, respectively. This suggests that the information order on a domain determines a natural notion of computability, and that the renee equation yields a systematic method for determining this notion of computability. We will also see that one renee equation describing an algorithm leads to another which captures its complexity. This provides a qualitative and quantitative *first order* view of computation, one very much in line with actual program development.

Unique solvability of the equation

Recall the definition of the μ topology from Section 2.4.

Definition 52. Let $(X, +)$ be a Hausdorff space with a binary operation that is associative. If (x_n) is a sequence in X, then its *infinite sum* is

$$\sum_{n\geq 1} x_n := \lim_{n\to\infty}(x_1 + \cdots + x_n)$$

provided that the limit of the partial sums on the right exists.

Definition 53. Let $+ : D^2 \to D$ be a binary operation on a continuous dcpo. A point $x \in D$ is *idle* if there is a μ open set $\sigma(x)$ around x such that

(i) $(\sigma(x), +)$ is a semigroup, and
(ii) If (x_n) is any sequence in $\sigma(x)$ which converges to x in the μ topology, then

$$\sum_{n\geq 1} x_n \text{ exists and } \lim_{n\to\infty}\sum_{k\geq n} x_k = \lim_{n\to\infty} x_n.$$

The operation $+$ is said to be *idle* at x.

An idle point is one where the "unwinding" of a recursive definition stops. For example, $0 \in \mathbb{N}$, or the empty list.

Definition 54. Let D be a continuous dcpo. A μ continuous operation $+ : D^2 \to D$ is *iterative* if it has at least one idle point.

Here are a few simple examples of iterative operations.

Example 47. Data types.

(i) $([S], \cdot)$ concatenation of lists. The idle points are $\{[\,]\}$.
(ii) $(\mathbb{N}^*, +)$ addition of natural numbers. The idle points are $\{0\}$.
(iii) $(\mathbb{N}^*, \times)$ multiplication of natural numbers. The idle points are $\{0, 1\}$.
(iv) $(\{\bot, \top\}, \vee)$ Boolean 'or.' The idle points are $\{\bot, \top\}$.
(v) $(\{\bot, \top\}, \wedge)$ Boolean 'and.' The idle points are $\{\bot, \top\}$.

The μ topology on each domain above is discrete. The *fixed points* of a function $f : P \to P$ are $\mathrm{fix}(f) = \{x \in P : f(x) = x\}$.

Definition 55. A splitting $r : D \to D$ on a dcpo D is *inductive* if for all $x \in D$, $\bigsqcup r^n x \in \mathrm{fix}(r)$.

Definition 56. Let D be a dcpo and $(E, +)$ be a domain with an iterative operation. A function $\delta : D \to E$ *varies* with an inductive map $r : D \to D$ provided that

(i) For all $x \in \mathrm{fix}(r)$, $\delta(x)$ is idle in E, and
(ii) For all $x \in D$, $\delta(r^n x) \to \delta(\bigsqcup r^n x)$ in the μ topology on E.

The function δ interprets the recursive part r of an algorithm in the domain $(E, +)$. A fixed point of r is mapped to an idle point in E: A point where recursion stops.

Definition 57. Let D be a dcpo and $(E, +)$ be a domain with an iterative operation. A *renee equation* on $D \to E$ is one of the form

$$\varphi = \delta + \varphi \circ r$$

where $\delta : D \to E$ varies with an inductive map $r : D \to D$.

Theorem 36 (Canonicity). *The renee equation*

$$\varphi = \delta + \varphi \circ r$$

has a unique solution which varies with r and agrees with δ on $\mathrm{fix}(r)$.

Please stop and read the last theorem again. Thank you. The importance of φ varying with r is that it enables a verification principle [15]. Here are a few basic instances of the renee equation.

Example 48. The factorial function

$$\mathrm{fac} : \mathbb{N} \to \mathbb{N}$$

is given by

$$\begin{aligned} \mathrm{fac}\,0 &= 1 \\ \mathrm{fac}\,n &= n \times \mathrm{fac}(n-1). \end{aligned}$$

Let $D = \mathbb{N}^*$ and $E = (\mathbb{N}^*, \times)$. Define $\delta : D \to E$ by

$$\delta(n) = \begin{cases} 1 \text{ if } n = 0, \\ n \text{ otherwise.} \end{cases}$$

and $\mathrm{pred} : D \to D$ by $\mathrm{pred}(n) = n - 1$, if $n > 0$, and $\mathrm{pred}(0) = 0$. The unique solution of

$$\varphi = \delta \times \varphi \circ \mathrm{pred}$$

which satisfies $\varphi(0) = 1$ is the factorial function.

Example 49. The length of a list

$$\mathrm{len} : [S] \to \mathbb{N}$$

is given by

$$\begin{aligned} \mathrm{len}\,[\,] \quad &= 0 \\ \mathrm{len}\,a :: x &= 1 + \mathrm{len}\,x. \end{aligned}$$

Let $D = [S]$ and $E = (\mathbb{N}^*, +)$. Define $\delta : D \to E$ by

$$\delta(x) = \begin{cases} 0 \text{ if } x = [\,], \\ 1 \text{ otherwise.} \end{cases}$$

and $\mathrm{rest} : D \to D$ by $\mathrm{rest}(a :: x) = x$ and $\mathrm{rest}([\,]) = [\,]$. The unique solution of

$$\varphi = \delta + \varphi \circ \mathrm{rest}$$

which satisfies $\varphi([\,]) = 0$ is the length function.

Example 50. The merging of two sorted lists of integers

$$\mathrm{merge} : [\mathrm{int}] \times [\mathrm{int}] \to [\mathrm{int}]$$

is given by the following ML code

```
fun merge( [ ], ys )          = ys : int list
  | merge( xs, [ ] )          = xs
  | merge( x :: xs, y :: ys ) = if x ≤ y then
                                   x :: merge( xs, y :: ys )
                                else
                                   y :: merge( x :: xs, ys );
```

Let $D = [\text{int}] \times [\text{int}]$ and $E = ([\text{int}], \cdot)$. Define $\delta : D \to E$ by

$$\begin{aligned}
\delta(x, [\,]) &= x\\
\delta([\,], y) &= y\\
\delta(x, y) &= [\min(\text{first}\, x, \text{first}\, y)], \text{ otherwise.}
\end{aligned}$$

and $\pi : D \to D$ by

$$\begin{aligned}
\pi(x, [\,]) &= ([\,], [\,])\\
\pi([\,], y) &= ([\,], [\,])\\
\pi(x, y) &= (\text{rest}\, x, y), \text{ if first}\, x \le \text{first}\, y;\\
\pi(x, y) &= (x, \text{rest}\, y), \text{ otherwise.}
\end{aligned}$$

The unique solution of

$$\varphi = \delta \cdot \varphi \circ \pi$$

satisfying $\varphi([\,], [\,]) = [\,]$ is merge.

The last example is interesting because solving the equation yields a new iterative operation on [int]. We shall make use of this fact in the next example to solve an equation for sorting. In this way, we see that algorithms can be built up by solving sequences of renee equations.

Example 51. The prototypical bubblesort of a list of integers

$$\text{sort} : [\text{int}] \to [\text{int}]$$

is given by

$$\begin{aligned}
\text{sort}\, [\,] &= [\,]\\
\text{sort}\, x &= \text{merge}(\ [\text{first}\, x],\ \text{sort rest}\, x\ )
\end{aligned}$$

Let $D = [\text{int}]$ and $E = ([\text{int}], +)$ where

$$+ : [\text{int}]^2 \to [\text{int}]$$

$$(x, y) \mapsto \text{merge}(x, y)$$

is the merge operation of Example 50. Define $\delta : D \to E$ by

$$\delta(x) = \begin{cases} [\,] & \text{if } x = [\,]\\ [\text{first}\, x] & \text{otherwise} \end{cases}$$

and let rest : [int] → [int] be the usual splitting. The unique solution of

$$\varphi = \delta + \varphi \circ \text{rest}$$

satisfying $\varphi[\,] = [\,]$ is sort.

Computability from the information order

In this section, we will see that the primitive and partial recursive functions can both be captured using the renee equation: each arises as a canonical notion of computability derivable from a given information order.

Definition 58. Let $\mathbb{N}_\perp$ denote the set $\mathbb{N} \cup \{\perp\}$, where $\perp$ is an element that does not belong to $\mathbb{N}$.

For instance, one could take $\perp = \{\mathbb{N}\}$, should the need arise.

Definition 59. A *partial function* on the naturals is a function

$$f : \mathbb{N}^n \to \mathbb{N}_\perp,$$

where $n \geq 1$. We say that f is *undefined* at x exactly when $f(x) = \perp$.

Thinking of f as an algorithm, $f(x) = \perp$ means that the program f crashed when we sent it input x.

Definition 60. The *composition* of a partial map $f : \mathbb{N}^n \to \mathbb{N}_\perp$ with partial mappings $g_i : \mathbb{N}^k \to \mathbb{N}_\perp$, $1 \leq i \leq n$, is the partial map

$$f(g_1, \cdots, g_n) : \mathbb{N}^k \to \mathbb{N}_\perp$$

$$f(g_1, \cdots, g_n)(x) = \begin{cases} f(g_1(x), \cdots, g_n(x)) & \text{if } (\forall i)\ g_i(x) \neq \perp; \\ \perp & \text{otherwise.} \end{cases}$$

That is, if in the process of trying to run the program f, the computation of one of its inputs fails, then the entire computation fails.

Definition 61. A partial map $f : \mathbb{N}^{n+1} \to \mathbb{N}_\perp$ is defined by *primitive recursion* from $g : \mathbb{N}^n \to \mathbb{N}_\perp$ and $h : \mathbb{N}^{n+2} \to \mathbb{N}_\perp$ if

$$f(\bar{x}, y) = \begin{cases} g(\bar{x}) & \text{if } y = 0; \\ h(\bar{x}, y-1, f(\bar{x}, y-1)) & \text{otherwise.} \end{cases}$$

where we have written $\bar{x} \in \mathbb{N}^n$.

Computationally, primitive recursion is a counting loop.

Definition 62. The class of *primitive recursive functions* on the naturals is the smallest collection of functions $f : \mathbb{N}^n \to \mathbb{N}$ which contains the zero function, the successor, the projections, and is closed under composition and primitive recursion.

The analogue of a ‘while’ loop is provided by minimization.

Definition 63. The *minimization* of a partial function $f : \mathbb{N}^{n+1} \to \mathbb{N}_\perp$ is the partial function

$$\mu f : \mathbb{N}^n \to \mathbb{N}_\perp$$

$$\mu f(x) = \min\{y \in \mathbb{N} : (\forall z < y)\; f(x,z) \neq \perp \;\&\; f(x,y) = 0\}$$

with the convention that $\mu f(x) = \perp$ if no such y exists.

Definition 64. The class of *partial recursive functions* on the naturals is the smallest collection of partial maps $f : \mathbb{N}^n \to \mathbb{N}_\perp$ which contains the zero function, the successor, the projections, and is closed under composition, primitive recursion, and minimization.

Let D be a domain which as a set satisfies $\mathbb{N} \subseteq D \subseteq \mathbb{N} \cup \{\infty\}$.

Definition 65. The sequence of domains $(D^n)_{n \geq 1}$ is given inductively by

$$\begin{aligned} D^1 &= D, \\ D^{n+1} &= D^n \times D^1,\; n > 0. \end{aligned}$$

We extend a few simple initial functions to D.

Definition 66. The initial functions.

(i) Addition of naturals $+ : D^2 \to D$ given by

$$(x,y) \mapsto \begin{cases} x+y & \text{if } x,y \in \mathbb{N}; \\ \infty & \text{otherwise.} \end{cases}$$

(ii) Multiplication of naturals $\times : D^2 \to D$ given by

$$(x,y) \mapsto \begin{cases} x \times y & \text{if } x,y \in \mathbb{N}; \\ \infty & \text{otherwise.} \end{cases}$$

(iii) The predicate $\leq: D^2 \to D$ given by

$$(x,y) \mapsto \begin{cases} x \leq y & \text{if } x,y \in \mathbb{N}; \\ \infty & \text{otherwise.} \end{cases}$$

(iv) The projections $\pi_i^n : D^n \to D$, for $n \geq 1$ and $1 \leq i \leq n$, given by

$$(x_1, \cdots, x_n) \mapsto \begin{cases} x_i \text{ if } (x_1, \cdots, x_n) \in \mathbb{N}^n; \\ \infty \text{ otherwise.} \end{cases}$$

A map $r : D^n \to D^n$ may be written in terms of its *coordinates* $r_i : D^n \to D$, for $1 \leq i \leq n$, as $r = (r_1, \cdots, r_n)$.

Definition 67. Let $\mathcal{C}(D)$ be the smallest class of functions $f : D^n \to D$ with the following properties:

(i) $\mathcal{C}(D)$ contains $+$, $\times$, $\leq$, and π_i^n, for $n \geq 1$ and $1 \leq i \leq n$,
(ii) $\mathcal{C}(D)$ is closed under substitution: If $f : D^n \to D$ is in $\mathcal{C}(D)$ and $g_i : D^k \to D$ is in $\mathcal{C}(D)$, for $1 \leq i \leq n$, then

$$f(g_1, \cdots, g_n) : D^k \to D \text{ is in } \mathcal{C}(D),$$

and
(iii) $\mathcal{C}(D)$ is closed under iteration: If $\delta : D^n \to D$ and $+ : D^2 \to D$ are in $\mathcal{C}(D)$, and $r : D^n \to D^n$ is a map whose coordinates are in $\mathcal{C}(D)$, then

$$\varphi = \delta + \varphi \circ r \in \mathcal{C}(D)$$

whenever this is a renee equation on $D^n \to D$.

$\mathcal{C}(D)$ contains maps of type $D^n \to D$. To obtain functions on the naturals, we simply restrict them to $\mathbb{N}^n$. In general, we obtain partial maps on the naturals, depending on whether or not D contains ∞.

Definition 68. The restriction of a mapping $f : D^n \to D$ to $\mathbb{N}^n$ is

$$|f| : \mathbb{N}^n \to \mathbb{N}_\perp$$

$$|f|(x) = \begin{cases} f(x) & \text{if } f(x) \in \mathbb{N}; \\ \perp & \text{otherwise.} \end{cases}$$

Let $\mathbb{N}^\infty$ denote the domain of naturals in their usual order with ∞ as a top element, $\mathbb{N}^*$ denote the domain of naturals in their dual order and $\mathbb{N}^\flat$ denote the domain of naturals ordered flatly: $x \sqsubseteq y \equiv x = y$.

The information order on a domain determines a notion of computability *because* it determines our ability to iterate.

Theorem 37.

(i) $|\mathcal{C}(\mathbb{N}^\infty)|$ *is the class of partial recursive functions.*
(ii) $|\mathcal{C}(\mathbb{N}^*)|$ *is the class of primitive recursive functions.*
(iii) $|\mathcal{C}(\mathbb{N}^\flat)|$ *is the smallest class of functions containing the initial functions which is closed under substitution.*

A renee equation for algorithmic complexity

One renee equation describing an algorithm leads to another describing its complexity. If we have an algorithm $\varphi = \delta + \varphi \circ r$, then in order to calculate $\varphi(x)$, we must calculate $\delta(x)$, $r(x)$, $\varphi(rx)$ and $\delta(x) + \varphi(rx)$. Thus, the *cost* $c_\varphi(x)$ of calculating $\varphi(x)$ is the sum of the four costs associated with computing $\delta(x)$, $r(x)$, $\varphi(rx)$ and $\delta(x) + \varphi(rx)$. In symbols,

$$c_\varphi(x) = c_\delta(x) + c_r(x) + c_\varphi(rx) + c_+(\delta(x), \varphi(rx)).$$

When the functions (c_δ, c_+, c_r) actually describe the complexity of an algorithm, the equation above can be solved uniquely for c_φ.

Proposition 6. *Let $\varphi = \delta + \varphi \circ r$ be a renee equation on $D \to E$. If $c_\delta : D \to \mathbb{N}^*$, $c_r : D \to \mathbb{N}^*$ and $c_+ : E^2 \to \mathbb{N}^*$ are functions such that for all $x \in D$,*

$$\lim_{n\to\infty} c_\delta(r^n x) = \lim_{n\to\infty} c_r(r^n x) = \lim_{n\to\infty} c_+(\delta(r^n x), \varphi r(r^n x)) = 0,$$

then

$$c_\varphi = c_\delta + c_r + c_+(\delta, \varphi \circ r) + c_\varphi(r)$$

is a renee equation on $D \to (\mathbb{N}^*, +)$.

Thus, one renee equation describing an algorithm leads to another describing its complexity. Let's briefly consider a quick example just to check that the ideas work the way they should, we calculate the complexity of a sorting algorithm.

Example 52. Recall the prototypical bubblesort of a list of integers

$$\text{sort} : [\text{int}] \to [\text{int}]$$

is given by

$$\begin{aligned} &\text{sort}\,[\,] = [\,] \\ &\text{sort}\,x \;\; = \text{merge}(\,[\text{first}\,x],\ \text{sort}\,\text{rest}\,x\,) \end{aligned}$$

Let $D = [\text{int}]$ and $E = ([\text{int}], +)$ where

$$+ : [\text{int}]^2 \to [\text{int}]$$

$$(x, y) \mapsto \text{merge}(x, y)$$

is the merge operation mentioned previously. Define $\delta : D \to E$ by

$$\delta(x) = \begin{cases} [\,] & \text{if } x = [\,] \\ [\text{first}\,x] & \text{otherwise} \end{cases}$$

and let $r : D \to D$ be the splitting $rx = \text{rest}\,x$. The unique solution of

$$\varphi = \delta + \varphi \circ r$$

satisfying $\varphi[\,] = [\,]$ is sort.

For the worst case analysis of sort $= \delta + \text{sort} \circ r$ the number of comparisons performed by r and δ on input x is zero. Hence,

$$c_r(x) = c_\delta(x) = 0,$$

while the cost of merging two lists x and y can be as great as $\mu x + \mu y$, so

$$c_+(x, y) = \mu x + \mu y.$$

By Prop. 6, we have a renee equation

$$c_{\text{sort}} = c_+(\delta, \text{sort} \circ r) + c_{\text{sort}}(r)$$

which *should* measure the complexity of bubblesort. But does it? By Theorem 36,

$$c_{\text{sort}}[\,] = 0,$$

while for any other list x, we have

$$\begin{aligned} c_{\text{sort}}(x) &= c_+(\delta(x), \text{sort}(rx)) + c_{\text{sort}}(rx) \\ &= \mu\,\delta(x) + \mu\,\text{sort}(rx) + c_{\text{sort}}(rx) \\ &= 1 + (\mu x - 1) + c_{\text{sort}}(rx) \\ &= \mu x + c_{\text{sort}}(rx). \end{aligned}$$

However, the function $f(x) = [\mu x(\mu x + 1)]/2$ varies with r, agrees with δ on $\text{fix}(r)$, and satisfies the equation above, so by the uniqueness in Theorem 36, we have

$$c_{\text{sort}}(x) = \frac{\mu x(\mu x + 1)}{2},$$

for all x.

One can go further with these ideas. In [18], the renee equation and measurement combine to provide a practical formal model of what a classical 'search' method $\varphi = \delta + \varphi \circ r$ is. A particular highlight of the approach is that it does not force one to distinguish between discrete notions of searching, such as linear and binary searching of lists, and continuous notions of searching, such as zero finding methods like the bisection. The complexity c_φ of such methods is then shown to be determined by the number of iterations it takes r to get 'close enough' to a fixed point. Thus, c_φ can also be calculated using the informatic derivative at a compact element.

6.3 Trajectories

Iterating an operator $f : D \to D$ yields a sequence $x, f(x), f^2(x), \ldots, f^n(x)$. Each $f^n(x)$ can be thought of as occuring at time n. It is natural to then wonder if an element $f^t(x)$ exists where $t \in [0, \infty)$. We would then have a trajectory $x : [0, \infty) \to D$ which describes the effect that f has had on x after t units of time. We could then take derivatives of x with respect to time and use them to learn things about a process. For instance, maybe the complexity of a process would amount to the point in time t when $x(s) \sqsubseteq x(t)$ for all s i.e. the "absolute maximum" of x. Maybe we could graph trajectories on the $t{-}\sqsubseteq$ axis to learn things about processes that we didn't know before. Maybe we should try this.

Kinematics

Proofs for this section can be found in [22].

Definition 69. A *variable* on a dcpo is a measurement $v : D \to [0, \infty)^*$ such that for all $x, y \in D$, we have $x \sqsubseteq y$ & $vx = vy \Rightarrow x = y$.

Definition 70. A *curve* on a domain D is a function $x : \mathrm{dom}(x) \to D$ where $\mathrm{dom}(x)$ is a nontrivial interval of the real line.

Each curve x determines a value of v at time t, which is the number $vx(t)$.

Definition 71. For a curve x and variable v on a dcpo,

$$\dot{x}_v(t) := \lim_{s \to t} \frac{vx(s) - vx(t)}{s - t}.$$

We then define

$$\ddot{x}_v := \frac{d\dot{x}_v}{dt}$$

and so on for higher order.

Because $\dot{x}_v : [0, \infty) \to \mathbb{R}$ is an ordinary function, higher order derivatives are calculated as usual – its the first derivative that requires theory.

Proposition 7. *Let x be a curve with $\dot{x}_v$ defined on (a, b).*

(i) *x is monotone increasing on $[a, b]$ iff $\dot{x}_v \leq 0$ on (a, b) and $x[a, b]$ is a chain.*
(ii) *x is monotone decreasing on $[a, b]$ iff $\dot{x}_v \geq 0$ on (a, b) and $x[a, b]$ is a chain.*
(iii) *x is constant on $[a, b]$ iff $\dot{x}_v = 0$ on (a, b) and $x[a, b]$ is a chain.*

Notice that the sign of $\dot{x}_v$ is an indicator of how *uncertainty* behaves: If $\dot{x}_v \leq 0$, then uncertainty is decreasing, so we are moving up in the order.

Definition 72. A curve x has a *relative maximum* at an interior point $t \in \mathrm{dom}(x)$ if there is an open set U_t containing t such that $x(s) \sqsubseteq x(t)$ for all $s \in U_t$. *Relative minimum* is defined dually, and these two give rise to *relative extremum.*

Notice that a *qualitative* relative maximum is a point in time where the *quantitative* uncertainty is a local minimum.

Lemma 6. *If a curve x has a relative extremum at interior point $t \in \mathrm{dom}(x)$, then for all variables v, either $\dot{x}_v(t) = 0$, or it does not exist.*

A nice illustration of why the qualitative idea $\sqsubseteq$ is important: if a curve has a derivative with respect to *just one* variable v, then its set of extreme points is contained in the set $\{t : \dot{x}_v(t) = 0\}$. This is quite valuable: we are free to choose the variable which makes the calculation as simple as possible.

Once we have the extreme points there is also a systematic way in the informatic setting to determine which (if any) are maxima or minima: the second derivative test, whose formalization requires one to acknowledge the qualitative structure on which it is implicitly founded.

Definition 73. A curve x is a *trajectory* if for all $t \in \mathrm{dom}(x)$ there is an open set U_t containing t such that

$$x(s) \sqsubseteq x(t) \text{ or } x(t) \sqsubseteq x(s)$$

for all $s \in U_t$.

Thus, a *trajectory* is a curve x with underlying qualitative structure; it is called C^2_v when $\ddot{x}_v$ is continuous, with respect to variable v.

Proposition 8. *Let x be a C^2_v trajectory. If $\dot{x}_v(t) = 0$ and $\ddot{x}_v(t) \neq 0$ for some interior point $t \in \mathrm{dom}(x)$, then x has a relative extremum at t.*

(i) *If $\ddot{x}_v(t) > 0$, then $x(t)$ is a relative maximum.*
(ii) *If $\ddot{x}_v(t) < 0$, then $x(t)$ is a relative minimum.*

In this work we will be mostly concerned with the strongest form of extrema on domains:

Definition 74. A curve x has an *absolute maximum* at $t \in \mathrm{dom}(x)$ if

$$x(s) \sqsubseteq x(t)$$

for all $s \in \mathrm{dom}(x)$. Absolute minimum is defined similarly.

Here is a simple but surprisingly useful way of establishing the existence of absolute extrema.

Proposition 9. *Let v be a variable on D and $x : [a,b] \to D$ a curve whose image is a chain. If $vx : [a,b] \to \mathbb{R}$ is Euclidean continuous, then*

(i) *The map x is continuous from the Euclidean to the Scott topology, and*
(ii) *The map x assumes an absolute maximum and an absolute minimum on $[a,b]$. In particular, its absolute maximum is*

$$x(t^*) = \bigsqcup_{t \in [a,b]} x(t)$$

for some $t^* \in [a,b]$, with a similar expression for the absolute minimum.

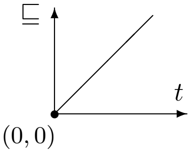
A valuable property of absolute maxima: If $x(t^*)$ is an absolute maximum, then for all variables v,

$$vx(t^*) = \inf\{vx(t) : t \in \mathrm{dom}(x)\}.$$

That is, an absolute maximum is a point on a curve which simultaneously minimizes *all* variables.

Linear searching

Suppose a list has $n > 0$ elements. Linear search begins with the first element in the list and proceeds to the next and so on until the key is located. At time t (after t comparisons), all elements with indices from 1 to t have been searched. Thus, a trajectory representing the information we have gained is $x(t) = t$ for $t \in [0, n]$. The natural space of informatic objects is $D = [0, n]$ whose natural measure of uncertainty is $vx = n - x$.



Next is a better example – one where the kinematics of computation will help us visualize a computation.

Binary searching

This algorithm causes a trajectory on $(\mathbf{I}\mathbb{R}, v)$ with $v[a, b] = b - a$. For a continuous $f : \mathbb{R} \to \mathbb{R}$, let $\mathrm{split}_f : \mathbf{I}\mathbb{R} \to \mathbf{I}\mathbb{R}$ be the bisection method on the interval domain defined by

$$\mathrm{split}_f[a, b] := \begin{cases} \mathrm{left}[a, b] & \text{if } f(a) \cdot f((a + b)/2) \leq 0; \\ \mathrm{right}[a, b] & \text{otherwise.} \end{cases}$$

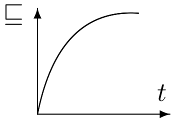
A given $x \in \mathbf{I}\mathbb{R}$ leads to a *trajectory* $x : [0, \infty) \to \mathbf{I}\mathbb{R}$ defined on natural numbers by

$$x(n) = \mathrm{split}_f^n(x)$$

and then extended to all intermediate times $n < t < n + 1$ by declaring $x(t)$ to be the *unique* element satisfying

$$x(n) \sqsubseteq x(t) \sqsubseteq x(n + 1) \quad \text{and} \quad \mu x(t) = \frac{vx}{2^t}.$$

By definition, the trajectory of binary search is also increasing. But graphing it is more subtle. It looks like this:



But why? Using the *kinematics of computation*, since $vx(t) = e^{-(\ln 2)t} \cdot vx(0)$, we have

$$\dot{x}_v(t) = (-\ln 2)vx(t) < 0$$

reflecting the fact that $x : [0, \infty) \to \mathbf{I}\mathbb{R}$ is *increasing.* In addition, $\ddot{x}_v(t) > 0$, so the graph is concave *down.* Notice that as $t \to \infty$, the trajectory should tend toward the answer as its velocity tends to zero.

Trajectories of classical search algorithms tend to increase with time. All of the curves basically look the same, so what's the point? It makes the dream of a "kinematics of computation" seem out of reach. But then, what is the point in dreaming of things that are *within* reach? Those aren't dreams, they're just things you plan to do.

Quantum searching

Grover's algorithm [9] for searching is the only known quantum algorithm whose complexity is *provably better* than its classical counterpart. It searches a list L of length n (a power of two) for an element k known to occur in L precisely m times with $n > m \geq 1$. The register begins in the pure state

$$|\psi\rangle = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} |i\rangle$$

and after j iterations of the Grover operator G

$$G^j|\psi\rangle = \frac{\sin(2j\theta + \theta)}{\sqrt{m}} \sum_{L(i)=k} |i\rangle + \frac{\cos(2j\theta + \theta)}{\sqrt{n-m}} \sum_{L(i)\neq k} |i\rangle$$

where $\sin^2\theta = m/n$. The probability that a measurement yields i after j iterations is

$$\sin^2(2j\theta + \theta)/m \text{ if } L(i) = k$$

and

$$\cos^2(2j\theta + \theta)/(n - m) \text{ if } L(i) \neq k.$$

To get the answer, we measure the state of the register in the basis $\{|i\rangle : 1 \leq i \leq n\}$; if we perform this measurement after j iterations of G, when the state of the register is $G^j|\psi\rangle$, our knowledge about the result is represented by the vector

$$x(j) = \left(\frac{\sin^2(2j\theta + \theta)}{m}, \ldots, \frac{\sin^2(2j\theta + \theta)}{m}, \right. \\ \left. \frac{\cos^2(2j\theta + \theta)}{n-m}, \ldots, \frac{\cos^2(2j\theta + \theta)}{n-m} \right)$$

The crucial step now is to *imagine* t iterations,

$$x(t) = \left(\frac{\sin^2(2t\theta + \theta)}{m}, \ldots, \frac{\sin^2(2t\theta + \theta)}{m}, \right. \\ \left. \frac{\cos^2(2t\theta + \theta)}{n-m}, \ldots, \frac{\cos^2(2t\theta + \theta)}{n-m} \right)$$

Thus, x is a curve on the domain $\mathit{\Delta}^n$ of classical states in its *implicative order* (Section 2.2)

$$x \sqsubseteq y \equiv (\forall i)\ x_i < y_i \Rightarrow x_i = x^+$$

where x^+ refers to the largest probability in x. Thus, only a maximum probability is allowed to increase as we move up in the information order on $\mathit{\Delta}^n$. If the maximum probability refers to a solution of the search problem, then moving up in this order ensures that we are getting closer to the answer.

We will now use this trajectory to analyze Grover's algorithm using the kinematics of computation. Here are some crucial things our analysis will yield:

(a) The complexity of the algorithm,
(b) A qualitative property the algorithm possesses called *antimonotonicity.* Without knowledge of this aspect, an experimental implementation would almost certainly fail (for reasons that will be clear later).
(c) An explanation of the algorithm as being an attempt to calculate a *classical proposition.*

Precisely now, the classical state $x(t)$ is a vector of probabilities that do not increase for $t \in \mathrm{dom}(x) = [a, b]$, $a = 0$ and $b = \pi/2\theta - 1$. The image of $x : [a, b] \to \mathit{\Lambda}^n$ is a chain in the implicative order, which is simplest to see by noting that it has the form

$$x = (f, \ldots, f, g, \ldots, g)$$

so that $g(s) \geq g(t) \Rightarrow x(s) \sqsubseteq x(t)$; otherwise, $x(t) \sqsubseteq x(s)$. We can now determine the exact nature of the motion represented by x using kinematics. Because $x : [a, b] \to D$ is a curve on a domain D whose image is a chain and whose time derivative $\dot{x}_v(t)$ exists with respect to a variable v on $\mathit{\Delta}^n$, we know that

(i) The curve x has an absolute maximum on $[a, b]$: There is $t^* \in [a, b]$ such that

$$x(t^*) = \bigsqcup_{t \in [a,b]} x(t),$$

and
(ii) Either $t^* = a$, $t^* = b$ or $\dot{x}_v(t^*) = 0$.

Part of the power of this simple approach is that we are free to choose any v we like. To illustrate, a tempting choice might be entropy $v = H$, but then solving $\dot{x}_v = 0$ means solving the equation

$$-m\dot{f}(1 + \log f) - (n - m)\dot{g}(1 + \log g) = 0$$

and we also have to determine the points where $\dot{x}_v$ is undefined, the set $\{t : g(t) = 0\}$. However, if we use

$$v = 1 - \sqrt{x^+},$$

we only have to solve a single elementary equation

$$\cos(2t\theta + \theta) = 0$$

for t, allowing us to conclude that the maximum must occur at $t = a$, $t = b$, or at points in

$$\{t : \dot{x}_v(t) = 0\} = \{b/2\}.$$

The absolute maximum of x is

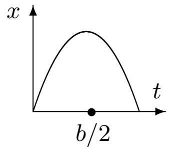
$$x(b/2) = (1/m, \ldots, 1/m, 0, \ldots, 0)$$

because for the other points we find a minimum of

$$x(a) = x(b) = \perp = (1/n, \ldots, 1/n).$$

The value of knowing the absolute maximum is that it allows us to calculate the complexity of the algorithm: it is $O(b/2)$, the amount of time required to move to a state from which the likelihood of obtaining a correct result by measurement is maximized. This gives $O(\sqrt{n/m})$ using $\theta \geq \sin\theta \geq \sqrt{m/n}$ and then $b/2 \leq (\pi/4)\sqrt{n/m} - 1/2$.

From $\dot{x}_v(t) \leq 0$ on $[a, b/2]$ and $\dot{x}_v(t) \geq 0$ on $[b/2, b]$, we can also graph x:



This is the 'antimonotonicity' of Grover's algorithm: if $j = b/2$ iterations will solve the problem accurately, $2j$ iterations will mostly unsolve it! This means that our usual way of reasoning about iterative procedures like numerical methods, as in "we must do at least j iterations," no longer applies. We must say "do exactly j iterations; no more, no less." As is now clear, precise estimates like these have to be obtained before experimental realization is possible.

Finally, as explained in more detail in [23], we can view Grover's algorithm as an attempt to calculate as closely as possible the classical proposition

$$x(b/2) = (1/m, \ldots, 1/m, 0, \ldots, 0) \in \mathrm{Ir}(\Delta^n) = \left\{x : \bigwedge \uparrow\!x \cap \max(\Delta^n) = x\right\}.$$

It does so by generating *approximations*

$$x(t) \ll x(b/2)$$

for all $t \neq b/2$.

Amplitude damping

Let $\mathcal{H}$ be the state space for a two dimensional quantum system. Two parties communicate with each other as follows. First, they agree up front on a fixed basis of $\mathcal{H}$, say $\{|\psi\rangle, |\phi\rangle\}$, which can be expressed in some basis $\{|0\rangle, |1\rangle\}$ as

$$|\psi\rangle = a|0\rangle + b|1\rangle \quad \& \quad |\phi\rangle = c|0\rangle + d|1\rangle$$

where the amplitudes a, b, c, d are all complex. The state $|\psi\rangle$ is taken to mean '0', while the state $|\phi\rangle$ is taken to mean '1'. The first party, the sender, attempts to send one of these two qubits $|*\rangle \in \{|\psi\rangle, |\phi\rangle\}$ to the second party, the receiver. The second party receives *some* qubit and performs a measurement in the agreed upon basis. The result of this measurement is one of the qubits $\{|\psi\rangle, |\phi\rangle\}$, which is then interpreted as meaning either a '0' or a '1'.

We say *some qubit* because as $|*\rangle$ travels, it suffers an unwanted interaction with its environment, whose effect on density operators can be described as

$$\varepsilon(\rho) = E_0 \rho E_0^\dagger + E_1 \rho E_1^\dagger$$

where the operation elements are given by

$$E_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\lambda} \end{pmatrix} \quad \& \quad E_1 = \begin{pmatrix} 0 & \sqrt{\lambda} \\ 0 & 0 \end{pmatrix}$$

This effect is known as *amplitude damping* and the parameter $\lambda \in [0, 1]$ can be thought of as the probability of losing a photon. Thus, the receiver does not necessarily acquire the qubit $|*\rangle$, but instead receives some degradation of it, describable by the density operator $\varepsilon(|*\rangle\langle*|)$.

The probability that '0' is received when '0' is sent is

$$\alpha = P(0|0) = -2|a|^4 p(\lambda) + |a|^2(\lambda + 2p(\lambda)) + 1 - \lambda$$

while the probability that '0' is received when '1' is sent is

$$\beta = P(0|1) = 2|a|^4 p(\lambda) + |a|^2(\lambda - 2p(\lambda))$$

where $p(\lambda) = -1 + \lambda + \sqrt{1-\lambda} \geq 0$. Thus, each choice of basis defines a classical binary channel (α, β). Notice that the probabilities α and β only depend on $|a|^2$ because $|c|^2 = |a|^2$ and $|b|^2 = |d|^2 = 1 - |a|^2$ by the orthogonality of $|\psi\rangle$ and $|\phi\rangle$, and because the initial expressions for α and β turn out to only depend on modulus squared terms. Because the basis is fixed, $|a|^2 \in [0, 1]$ is a constant and we obtain a function $x : [0, 1] \to \mathbb{N}$ of λ given by

$$x(\lambda) = (\alpha(\lambda), \beta(\lambda))$$

where we recall that $\mathbb{N} \simeq \mathbf{I}[0, 1]$ is the domain of binary channels. Its domain theoretic nature was first established in [29]:

Proposition 10. *The trajectory $x : [0,1] \to \mathbb{N}$ is Scott continuous.*

One valuable aspect of x being Scott continuous is that we can now make precise the connection between quantum information's intuitive use of the word 'noise' and information theory's precise account of it: the quantity $\mathrm{C}(x(\lambda))$ decreases as λ increases i.e. the amount of information that the two parties can communicate decreases as the the probability of losing a photon increases. In the extreme cases,

$$x(0) = (1,0) \ \ \& \ \ x(1) = (|a|^2, |a|^2)$$

yielding respective capacities of 1 and 0. There is a more fundamental idea at work in this example and in many others like it: we have learned about capacity by only examining how the probabilities in the noise matrix change, and this more than justifies the domain theoretic approach. Imagine what would happen if we actually tried to calculate $\mathrm{C}(x(\lambda))$ explicitly: we would have to substitute $\alpha(\lambda) = -2|a|^4p(\lambda) + |a|^2(\lambda + 2p(\lambda)) + 1 - \lambda$ for a and $\beta(\lambda) = 2|a|^4p(\lambda) + |a|^2(\lambda - 2p(\lambda))$ for b into the formula

$$\mathrm{C}(a,b) = \log_2\left(2^{\frac{\bar{a}H(b)-\bar{b}H(a)}{a-b}} + 2^{\frac{bH(a)-aH(b)}{a-b}}\right)$$

and then seek to show that the resulting quantity decreases as λ increases.

Decoherence over time

One interesting aspect of amplitude damping is that it is *not unital.* Any unital qubit channel will lead to a trajectory defined on some nontrivial interval since all classical channels derived from them are binary symmetric and the binary symmetric channels form a chain in $\mathbb{N}$. An interesting example in this last regard is phase damping as a function of time, whose effect on the pure state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ with density operator

$$\rho = |\psi\rangle\langle\psi| = \begin{pmatrix} |\alpha|^2 & \alpha\beta^* \\ \alpha^*\beta & |\beta|^2 \end{pmatrix}$$

after t units of time is

$$\rho(t) = \begin{pmatrix} |\alpha|^2 & \alpha\beta^* e^{-t/t_d} \\ e^{-t/t_d}\alpha^*\beta & |\beta|^2 \end{pmatrix}$$

where t_d is a constant known as the *decoherence time.* If the qubit decoheres for t units of time, then a '0' may no longer be a '0' and a '1' may no longer be a '1'. Specifically, the probability that a '0' is still a '0' is

$$P(0|0) = |a|^4 + 2e^{-t/t_d}|a|^2|b|^2 + |b|^4$$

while the probability that a '1' changes into a '0' is

$$P(0|1) = 1 - P(0|0).$$

This gives rise to a binary symmetric channel.

6.4 Vectors

We can think of domains as a qualitative way of reasoning about informative objects, and measurement as a way of determining the amount of information in an object. But neither set of ideas attempts to *directly* answer the question "What is information?" In this section, we offer one possible answer to this question which has pragmatic value and is of interest to computer science.

To begin, we assume that the words 'complexity' and 'information' are just that – words. We start from a clean slate, forgetting the various connotations these words have in the sciences, and simply begin talking about them intuitively. 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 (information) you have about its solution: the less you know about the solution, the more work you should expect to do. An inequality like the one above could be valuable.

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 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. By (1), our intuitive development of 'complexity' turns out to be capable of deriving lower bounds on the complexity of algorithms such as sorting and searching. 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 the *majorization relation* from Section 2.2. Muirhead [34] discovered majorization in 1903, and in the last 100 years his relation has found impressive applications in areas such as economics, computer science, physics and pure mathematics [2][14]. We will see that the complexity c_D is determined by its value on this subset.

The limits (1) and (2) comprise what we call *the universal limit,* because it is taken over the class of *all* domains. The pair $(\sigma,\leq)$ can also be derived on a *fixed* domain (D,μ) provided 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 identifies the essential mathematical structure required to execute classical information theory [41] over the class of semantic domains, that the *fixed point theorem* springs forth: as with recursive programs, the semantics of *information* can also be specified by a least fixed point:

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

where Φ is the copying operator and $\bot$ is the complexity c_D, i.e., the least fixed point of domain theory connects complexity in computer science to entropy in physics. We thus learn that one can use domains to define the complexity of objects in such a way that information becomes a concept derived from complexity in a precise and systematic manner: as a least fixed point.

Processes

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.'

Definition 75. A pair of elements $x, y \in D$ are *orthogonal* if $\mu(\uparrow x \cap \uparrow y) \subseteq \{0\}$. This is written $x \perp y$.

The word 'domain' in this section means a continuous dcpo D with a least element $\perp$ and a map μ that measures all of D. By replacing μ with $\mu/\mu\perp$ if necessary, we can and will assume that $\mu\perp = 1$. Finally, we will assume that

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

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

Example 53.

(i) $\mathbf{I}[0,1]$ with the length measurement μ is a domain.

(ii) Let $p \in \Delta^n$ be a classical state with all $p_k > 0$ and Σ^∞ the strings over the alphabet $\Sigma = \{0, \ldots, n-1\}$. Define $\mu : \Sigma^\infty \to [0, \infty)^*$ by $\mu\perp = 1$ and $\mu i = p_{i+1}$ for $i \in \Sigma$, 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).

An immediate corollary is the case $p = (1/2, 1/2) \in \Delta^2$ and $\Sigma = \{0, 1\} = 2$, the binary strings with the usual measurement: $(2^\infty, 1/2^{|\cdot|})$ is a domain. This is the basis for the study of binary codes. The fact that it is a domain *implies* the vital *Kraft inequality* of classical information theory.

Theorem 38 (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 54. The Kraft inequality. We apply the last example with

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

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

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

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 (D, μ) gives rise to a notion of *process*: a set of outcomes which are (a) different and (b) achievable in finite time.

Definition 76. 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.

Complexity (quantitative)

There is a natural function $-\log \mu : P^n(D) \to (0, \infty)^n$ which takes a process $p \in P^n(D)$ to the positive vector

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

By considering processes on the domain of binary strings $(2^\infty, \mu)$, 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.

Proposition 11. *For all $x \in \Delta^n$,*

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

In particular, the function c is symmetric.

So we can restrict our attention to monotone decreasing states Λ^n.

Definition 77. The *expectation* of $p \in P^n(D)$ is $\langle p \rangle : \Lambda^n \to [0, \infty)^*$ given by

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

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:

Definition 78. 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 $c(x) = h(\text{sort}^-(x))$ for all $x \in \Delta^n$. The *Shannon entropy* $\sigma : \Delta^n \to [0,\infty)$

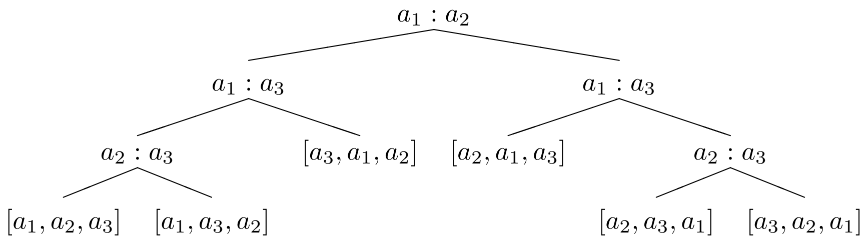
$$\sigma x := -\sum_{i=1}^{n} x_i \log x_i$$

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

Proposition 12. *If (D,μ) is a domain, then the complexity $h_D : (\Lambda^n, \leq) \to [0,\infty)^*$ is Scott continuous and $h_D \geq \sigma$ where σ is entropy and $\leq$ is majorization.*

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:

Example 55. Lower bounds on algorithmic complexity. Consider the problem of sorting lists of n objects by comparisons. Any algorithm which achieves this has a binary decision tree. For example, for lists with three elements, a_1, a_2, 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

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

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

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

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 $O(\log n)$.

We have used *domain theoretic structure* as the basis for a new approach to counting the number of leaves in a binary tree. Just as different domains can give rise to different notions of computability (Section 6.2), different domains can also give rise to different complexity classes, for the simple reason that changing the order changes the notion of process. An example of this is $(L, \mu) \subseteq (2^\infty, \mu)$ which models linear search (Example 57).

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 79. 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. 12.

Lemma 7. *For each domain (D, μ), $\leq\ \subseteq\ \leq_D$.*

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 80. 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. One example of a symbolic domain is (Σ^∞, μ) for an alphabet Σ with $\otimes$ being concatenation.

Example 56. The $\otimes$ on $\mathbf{I}[0,1]$ is

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

$(\mathbf{I}[0,1], \otimes)$ is a monoid with $\perp \otimes x = x \otimes \perp = x$ and the measurement μ is a homomorphism! We can calculate zeroes of real-valued functions by repeatedly $\otimes$-ing left$(\perp) = [0, 1/2]$ and right$(\perp) = [1/2, 1]$, i.e., the bisection method.

We can $\otimes$ processes too: If $p : \{1, \ldots, n\} \to D$ and $q : \{1, \ldots, m\} \to D$ are processes, then $p \otimes q : \{1, \ldots, nm\} \to D$ is a 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. We can characterize qualitative complexity on symbolic domains:

Theorem 39. *Let $(D, \otimes, \mu)$ be a symbolic domain. If there is a binary process $p : \{1, 2\} \to D$, then the relation $\leq_D \, = \, \leq$.*

The universal limit

We now see 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 40. *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.

Corollary 6. *Shannon entropy $\sigma : (\Lambda^n, \leq) \to [0, \infty)^*$ is Scott continuous.*

By Theorem 40, 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 39, 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.

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 57. 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 $(\mathbf{I}[0,1],\mu)$, $h_n(x) = -\sum_{i=1}^{n} x_i \log x_i$, Shannon entropy.

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 41. *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.

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' elements using $\otimes$, h and σ can be very different: Searching costs $O(n)$ on L, so h_L and σ are not of the same order. We need a slightly better estimate.

Definition 81. If $(D,\otimes,\mu)$ is a symbolic domain, then the integer

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

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

By orthogonality, $c(k) \geq \log k$ always holds, so to calculate the algebraic index we need only prove $c(k) \leq \log k$. The value of the index for us is that:

Corollary 7. *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.

There are results in [28] which explain why the algebraic index is a natural idea, but these use the *Gibbs map* and *partition function* from thermodynamics, which we do not have the space to discuss. But, it is simple to see that the algebraic index of $\mathbf{I}[0,1]$ is 2, the algebraic index of Σ^∞ is $|\Sigma|$ and in general, if there is a process $p \in P^n(D)$ on a symbolic domain with $(\mu p_1, \ldots, \mu p_n) = \perp \in \Lambda^n$ for some n, then D has an algebraic index $k \leq n$.

The fixed point theorem

Let Λ be the set of *all* monotone decreasing states and let $\otimes : \Lambda \times \Lambda \to \Lambda$ be

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

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}.

Definition 82. The copying operator $! : X \to X$ on a set X with a tensor $\otimes$ is

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

for all $x \in X$.

If $p \in P^n(D)$ is a process whose possible outputs are distributed as $x \in \Lambda^n$, then two independent copies of p considered together as a single process $!p$ will have outputs distributed according to $!x$. Now let $[\Lambda \to [0,\infty)^*]$ be the dcpo with the pointwise order $f \sqsubseteq g \equiv (\forall x)\, f(x) \geq g(x)$.

Theorem 42. *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.

This iterative process is very sensitive to where one begins. 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 proves that σ is genuinely a *least* fixed point.

The fixed point theorem can be used to derive Shannon's noiseless coding theorem [28]. In the proof of Theorem 42, we can regard Λ a continuous dcpo by viewing it as a disjoint union of domains. But 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 8. *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. In closing, we should mention that Φ might also provide a systematic approach to defining information fix(Φ) from complexity h in situations more general than symbolic domains.

The quantum case

The fixed point theorem also holds for quantum states where one replaces σ by von Neumann entropy, and $\otimes$ on domains by the algebraic tensor $\otimes$ of operators. (The domain theoretic $\otimes$ can also be mapped homomorphically onto the tensor of quantum states in such a way that domain theoretic orthogonality implies orthogonality in Hilbert space.) Several new connections emerge between computer science and quantum mechanics whose proofs combine new results with work dating as far back as Schrödinger [39] in 1936. The bridge that connects them is domain theory and measurement. One such result proves that reducing entanglement by a technique called local operations and classical communication is equivalent to simultaneously reducing the average case complexity of all binary trees, a major application of Theorem 39 that we could not include in this paper due to space limitations. These and related results are in [28].

7 Provocation

...and accordingly all experience hath shewn,
that mankind are more disposed to suffer, while evils are sufferable,
than to right themselves by abolishing the forms to which they are accustomed.
– Thomas Jefferson, The Declaration of Independence.

What is a domain?

The 'domains' of classical and quantum states are dcpo's with a definite notion of approximation, but they are *not* continuous. Their notion of approximation is

$$x \ll y \equiv (\forall \text{ directed } S)\ y = \bigsqcup S \Rightarrow (\exists s \in S)\, x \sqsubseteq s$$

On a *continuous dcpo*, the relation above is *equivalent* to the usual notion of approximation. In general, they are not equal, and the canonical examples are $(\varDelta^n, \sqsubseteq)$ in the Bayesian order and $(\varOmega^n, \sqsubseteq)$ in the spectral order. We forgot to mention this in Section 4 because we wanted to brainwash the reader, to convince them that the 'domain' illusion was real. Of course, in fairness to the author, we never said that we knew what a domain was exactly, just that they existed and that we would see lots of examples of them. Another possible example of a domain, the domain of *infinite* dimensional quantum states, is given in [33]. As a final example of something that is probably a domain, let us consider *the circle.*

I once had a prominent domain theorist tell me when I was a student that the circle could not be partially ordered in a natural way. I didn't believe it then and I don't believe it now. But now I have a reason:

The circle

If we have two pure states $|\psi\rangle$ and $|\phi\rangle$ written in a basis $|i\rangle$ of n dimensional Hilbert space $\mathcal{H}^n$,

$$|\psi\rangle = \sum_{i=1}^{n} a_i|i\rangle \quad |\phi\rangle = \sum_{i=1}^{n} b_i|i\rangle$$

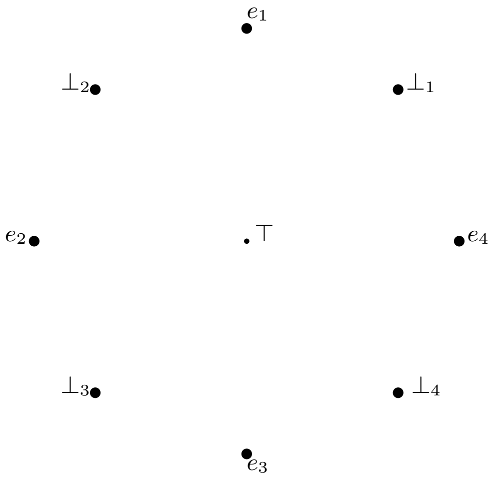
where the $a_i, b_i \in \mathbb{C}$ are complex, how can we order them so that (generally speaking) $|\psi\rangle \sqsubseteq |\phi\rangle$ means that the result of measuring the system in state ϕ is more predictable than the result of measuring the system in the state ψ? If we had an order $\sqsubseteq$ on classical probability distributions $\varDelta^n$, and another order $\sqsubseteq$ on phases $S^1 \cup \{0\}$, we could answer the question in what looks to be a natural way:

$$|\psi\rangle \sqsubseteq |\phi\rangle \equiv (|a_1|^2, \ldots, |a_n|^2) \sqsubseteq (|b_1|^2, \ldots, |b_n|^2)\ \&\ (\forall i)\, \text{phase}(a_i) \sqsubseteq \text{phase}(b_i).$$

Many orders on $\varDelta^n$ are known. So the entire question is reduced to the ordering of phases.

It's just a phase

The phase of a complex number is either zero or a point on the circle, so the problem of ordering phases is really just the question of how to order the circle.



One way to order phases is to order the circle so that the arc from any $\perp_i$ to an adjacent e_j is isomorphic to $([0,1],\leq)$, and that the center of the circle $\top = (0,0)$ is above everything. Dynamically, if we start at e_4 and begin traversing the circle counterclockwise, then we move down until reaching $\perp_1$, at which point we begin moving up until e_1, down until $\perp_2$, up until e_2, down until $\perp_3$, up until e_3, down until $\perp_4$, and then up until returning to e_4. Notice that this is the kind of domain that Grover's algorithm, when viewed as acting on a two dimensional subpsace, seems to 'move' in.

Another way to order phases is to use the discrete order: $x \sqsubseteq y$ iff $x = y$ or $y = (0,0)$. This is very satisfying in that it does not leave one worried about the meaning of the order in the case where the classical distributions stay constant but the phases are allowed to vary.

Example 58. The reason that $\top = (0,0)$ is above everything is so that relations like the following are satisfied:

$$\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \sqsubseteq |0\rangle$$

$$\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \sqsubseteq |1\rangle$$

What is a measurement?

Though the more general formulation of approximation for domains like Ω^n is certainly meaningful, there are things that are missing. The definition of 'domain' that we are looking for should allow one to do things like: prove sobriety of the Scott topology and give a satisfying definition of measurement. Yes, I realize that we defined measurement for a dcpo in Section 2.3, but I never said that we found the definition entirely convincing. The definition of measurement has more impact on a *continuous dcpo* as evidenced by results like Theorems 2 and 3.

Related to this question are two more pressing issues: (a) systematic methods for deriving higher order measurements from simpler measurements, and (b) techniques for proving that a given function is a measurement. For instance, to illustrate (a), if D is a domain and $F(D)$ is some higher order domain, like a powerdomain or an exponential object, can a measurement on D be used to *simply* construct one on $F(D)$? The question (b) is particularly urgent in physics and information theory: generally speaking, proving that functions like entropy and capacity are measurements is about as much fun as being a domain theorist in search of a decent job[4].

References

1. Abramsky, S. and Jung, A. (1994) Domain theory. In In S. Abramsky, D. M. Gabbay, T. S. E. Maibaum (editors), *Handbook of Logic in Computer Science* **III**, Oxford University Press.
2. P. M. Alberti and A. Uhlmann. *Stochasticity and partial order: doubly stochastic maps and unitary mixing.* Dordrecht, Boston, 1982.
3. R. P. Brent, *Algorithms for minimization without derivatives.* Prentice-Hall, 1973.
4. F. L. Chernous'ko, *An optimal algorithm for finding the roots of an approximately computed function.* Zh. vychisl. Mat. i mat. Fiz. **8**, 4, p. 705–724, 1968, English translation.
5. B. Coecke and K. Martin. *A partial order on classical and quantum states.* Oxford University Computing Laboratory Research Report, August 2002.
6. Edalat, A. and Heckmann, R. (1998) A computational model for metric spaces. *Theoretical Computer Science* **193**, 53–73.
7. K. Falconer, *Fractal geometry.* John Wiley and Sons, 1990.
8. O. Gross and S. M. Johnson, *Sequential minimax search for a zero of a convex function.* Mathematical Tables and Other Aids to Computation, Vol. 13, Issue 65, p. 44–51, 1959.
9. L. K. Grover. *Quantum mechanics helps in searching for a needle in a haystack.* Physical Review Letters, **78**:325, 1997.
10. S.W. Hawking and G.F.R. Ellis. *The large scale structure of space-time.* Cambridge Monographs on Mathematical Physics. Cambridge University Press, 1973.
11. Hutchinson, J. E. (1981) Fractals and self-similarity. *Indiana University Mathematics Journal* **30**, 713–747.
12. M. Kowalski, K. Sikorski and F. Stenger, *Selected topics in approximation and computation.* Oxford University Press, 1995.
13. L. G. Kraft. *A device for quantizing, grouping and coding amplitude modulated pulses.* M.S. Thesis, Electrical Engineering Department, MIT, 1949.

[4] Any such domain theorist should send a CV and some recent papers to `keye.martin@nrl.navy.mil` immediately.

14. A. W. Marshall and I. Olkin. *Inequalities: Theory of majorization and its applications.* Academic Press Inc., 1979.
15. K. Martin (2000) A foundation for computation. Ph.D. Thesis, Tulane University, Department of Mathematics.
16. K. Martin. The measurement process in domain theory. Lecture Notes In Computer Science, Vol. 1853, Springer-Verlag, 2000.
17. K. Martin (2001) Unique fixed points in domain theory. Electronic Notes in Theoretical Computer Science.
18. K. Martin (2001) A renee equation for algorithmic complexity, Lecture Notes in Computer Science, Springer-Verlag, Volume 2215.
19. K. Martin. Powerdomains and zero finding. Electronic Notes in Theoretical Computer Science, Vol. 59(3), 2001.
20. K. Martin. The informatic derivative at a compact element. Lecture Notes in Computer Science, Vol. 2303, Springer-Verlag, 2002.
21. K. Martin, *B-sides.* Oxford University Computing Lab, Research Report, 2003.
22. K. Martin. Epistemic motion in quantum searching. Oxford University Computing Laboratory, Research Report, 2003.
23. K. Martin. *A continuous domain of classical states.* Oxford University Computing Laboratory, Research Report, 2003.
24. K. Martin (2004) Fractals and domain theory. Mathematical Structures in Computer Science, Volume 14, Issue 6, p. 833–851, Cambridge University Press.
25. K. Martin and J. Ouaknine. *Informatic vs. classical differentiation on the real line.* Electronic Notes in Theoretical Computer Science, Vol. 73, 2004.
26. K. Martin and P. Panangaden. *A domain of spacetime intervals in general relativity.* Communications in Mathematical Physics, 267(3):563–586, November 2006.
27. K. Martin. *Compactness of the space of causal curves.* Journal of Classical and Quantum Gravity, 2006.
28. K. Martin. *Entropy as a fixed point.* Theoretical Computer Science, 2006.
29. K. Martin (2007) Topology in information theory in topology. Theoretical Computer Science, to appear.
30. K. Martin. *A domain theoretic model of qubit channels.* To appear, 2008.
31. K. Martin. *The maximum entropy state.* Logical Methods in Computer Science, to appear.
32. K. Martin and P. Panangaden. In preparation. 2008.
33. J. Mashburn. *A spectral order for infinite dimensional quantum spaces.* Electronic Notes in Theoretical Computer Science, Volume 173, 2007.
34. 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.
35. M. Nielsen and I. Chuang, Quantum computation and quantum information. Cambridge University Press, 2000.
36. Roger Penrose. Gravitational collapse and space-time singularities. *Phys. Rev. Lett.*, 14:57–59, 1965.
37. Roger Penrose. *Techniques of differential topology in relativity.* Society for Industrial and Applied Mathematics, 1972.
38. H. L. Royden, *Real analysis.* Third Edition, Macmillan Publishing, 1988.

39. E. Schrödinger. Proceedings of the Cambridge Philosophical Society **32**, 446 (1936).
40. D. Scott. *Outline of a mathematical theory of computation.* Technical Monograph PRG-2, Oxford University Computing Laboratory, November 1970.
41. C. E. Shannon. *A mathematical theory of communication.* Bell Systems Technical Journal 27, 379–423 and 623–656, 1948.
42. R. Sorkin. Spacetime and causal sets. In J. D'Olivo et. al., editor, *Relativity and Gravitation: Classical and Quantum.* World Scientific, 1991.
43. D. Spreen (2001) On some constructions in quantitative domain theory. Extended Abstract. `http://www.informatik.uni-siegen.de/ spreen/`
44. R.M. Wald. *General relativity.* The University of Chicago Press, 1984.
45. P. Waszkiewicz. *Quantitative continuous domains.* PhD Thesis, University of Birmingham, 2002.
46. M. Yamaguti, M. Hata and J. Kigami, *Mathematics of fractals.* Translations of Mathematical Monographs, American Math Society, vol. 167, 1997.