

The Informatic Derivative at a Compact Element

Keye Martin

Oxford University Computing Laboratory
Wolfson Building, Parks Road, Oxford OX1 3QD
kmartin@comlab.ox.ac.uk
<http://web.comlab.ox.ac.uk/oucl/work/keye.martin>

Abstract. We extend the informatic derivative to compact elements in domains. This allows one to quantitatively analyze processes which manipulate both continuous and discrete data in a *uniform* manner.

1 Introduction

One of the important aspects of the measurement formalism in domain theory is its informatic derivative: Given a map $f : D \rightarrow D$ on a domain with a measurement (D, μ) , its informatic derivative is given by

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

where the limit is taken in the μ topology at a point p which is not isolated in the μ topology on D .

As one might expect, this idea was applied to establish the first relationship between domain theory and the differential calculus [2], where it was shown that it could be used to give a purely domain theoretic account of C^1 mappings on the real line. But it has other applications which appear to be more promising. One of these is that it allows us to make sense of the idea “rate of change” with respect to a measurement. For instance, one can discuss the rate at which a mapping on a domain converges to its fixed point. This has natural applications in the study of what one might term continuous phenomena, like in numerical analysis, where it provides a uniform approach to calculating rates of convergence – while some rates of convergence may be determined using the classical derivative (an example of an informatic derivative), some of the most well-known methods for zero finding in all of numerical analysis have rates of convergence that arise as informatic derivatives which are *not* classical (in the sense of calculus).

But 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 $\text{rest}: [S] \rightarrow [S]$ which removes the first element from a nonempty list, and sends the empty list to itself, satisfies

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

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

In this paper, we offer 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 work 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 that we have identified ideas which allow us to systematically calculate both of these complexity notions in a uniform manner. But what appears more pertinent than either of these is that we are (slowly but surely) developing a very real understanding of informatic rates of change.

2 Background

2.1 Domain Theory

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

For elements x, y of a dcpo D , we write $x \ll y$ iff for every directed subset S with $y \sqsubseteq \bigsqcup S$, we have $x \sqsubseteq s$, for some $s \in S$. In the special case that this occurs for $x = y$, we call x *compact*. The set of compact elements in D is $K(D)$.

Definition 1. Let (D, \sqsubseteq) be a dcpo. We set

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

and say D is *continuous* if $\downarrow x$ is directed with supremum x for each $x \in D$. A *domain* is a continuous dcpo.

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

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

and that it *preserves directed suprema*:

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

for all directed $S \subseteq D$.

Definition 2. A subset B of a dcpo D is a *basis* for D if $B \cap \downarrow x$ contains a directed subset with supremum x , for each $x \in D$. A dcpo is *algebraic* if its compact elements form a basis. A dcpo is ω -*continuous* if it has a countable basis.

The next definition is fundamental in this paper: Splittings model the recursive part of an algorithm [4].

Definition 3. A *splitting* on a poset D is a function $s : D \rightarrow D$ with $x \sqsubseteq s(x)$ for all $x \in D$.

2.2 Examples of Domains

In this section, we give examples of domains that we will use in this paper.

Example 1. Orders on the naturals.

(i) The naturals $\mathbb{N}^\infty = \mathbb{N} \cup \{\infty\}$ in their usual order with a top element ∞ :

$$(\forall x, y \in \mathbb{N}^\infty) x \sqsubseteq y \Leftrightarrow (x \leq y \ \& \ x, y \in \mathbb{N}) \text{ or } y = \infty.$$

(ii) The naturals $\mathbb{N}^* = \mathbb{N}$ in their dual order: $x \sqsubseteq y \Leftrightarrow y \leq x$.

(iii) The naturals $\mathbb{N}^\flat = \mathbb{N}$ ordered flatly: $x \sqsubseteq y \Leftrightarrow x = y$.

Example 2. The *interval domain* is the collection of compact intervals of the real line

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

ordered under reverse inclusion

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

is an ω -continuous dcpo. The supremum of a directed set $S \subseteq \mathbf{IR}$ is $\bigcap S$, while the approximation relation is characterized by $I \ll J \Leftrightarrow J \subseteq \text{int}(I)$. A countable basis for \mathbf{IR} is given by $\{[p, q] : p, q \in \mathbb{Q} \ \& \ p \leq q\}$.

Our final example is the domain $[S]$ of finite lists over a poset (S, \leq) .

Definition 4. A *list* over S is a function $x : \{1, \dots, n\} \rightarrow S$, for $n \geq 0$. The *length* of a list x is $|\text{dom } x|$. The set of all (finite) lists over S is $[S]$. A list x is *sorted* if x is monotone as a map between posets.

A list x can be written as $[x(1), \dots, x(n)]$, where the *empty list* (the list of length 0) is written $[\]$. We 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 $[1, 2, 3]$ is written $1 :: [2, 3]$.

Definition 5. 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 $\text{scale}(K) : \{1, \dots, |K|\} \rightarrow K$ given by

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

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

Definition 6. For $x, y \in [S]$, x is a *sublist* of y iff there is a convex subset $K \subseteq \{1, \dots, \text{length } y\}$ such that $y \circ \text{scale } K = x$.

Example 3. 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 . However, $[1, 4, 5, 6], [1, 3]$ and $[2, 4]$ are *not* sublists of L .

Lemma 1 (Martin [2]). *The finite lists $[S]$ over a set S , ordered under reverse convex containment,*

$$x \sqsubseteq y \Leftrightarrow y \text{ is a sublist of } x,$$

form an algebraic dcpo with $[S] = K([S])$. Thus, $[S]$ is ω -continuous iff S is countable.

The order on $[S]$ is based on computational progress: Intuitively, it is easier to solve a problem on input $[\]$ than for any other input x , hence $x \sqsubseteq [\]$.

2.3 Content and the μ Topology

It is during the study of measurement [2] that one encounters for the first time the μ topology on a domain. Let $[0, \infty)^*$ denote the domain of nonnegative reals in the order opposite to their natural one.

Definition 7. A Scott continuous map $\mu : D \rightarrow [0, \infty)^*$ on a domain *measures* the elements in a set $X \subseteq D$ when for all $x \in X$, if (x_n) is a sequence with $x_n \ll x$ then

$$\lim_{n \rightarrow \infty} \mu x_n = \mu x \Rightarrow \bigsqcup_{n \in \mathbb{N}} x_n = x,$$

and this supremum is directed. We write this as $\mu \rightarrow \sigma_X$.

The terminology used in [2] and [5] is different, but the ideas are identical. The next result is fundamental and we use it often (implicitly).

Proposition 1 (Martin [2]). *Let $\mu : D \rightarrow [0, \infty)^*$ be a map that measures $X \subseteq D$. Then for all $x \in D$ and $y \in X$, we have $x \sqsubseteq y$ and $\mu x = \mu y \Rightarrow x = y$.*

For the sake of formality, we mention the following.

Definition 8. A *measurement* is a map $\mu : D \rightarrow [0, \infty)^*$ which measures the set $\ker \mu := \{x \in D : \mu x = 0\}$.

If x is an approximation of r , that is, $x \ll r$, then $|\mu x - \mu r|$ is a measure of how closely x approximates r , while μx is a measure of the uncertainty in x . If we obtain an improved approximation y of r , $x \sqsubseteq y \ll r$, then y should be more certain than x . Hence $\mu x \geq \mu y$.

Definition 9. The μ topology on a continuous dcpo D has as a basis all sets of the form $\uparrow x \cap \downarrow y$ where $x, y \in D$. It is denoted μ_D .

To clarify the relation between the two ideas, given a measurement $\mu \rightarrow \sigma_D$, consider the elements ε -close to $x \in D$, for $\varepsilon > 0$, given by

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

Regardless of the measurement we use, these sets are *always* a basis for the μ topology.

Theorem 1 (Martin [2]). *For a Scott continuous mapping $\mu : D \rightarrow [0, \infty)^*$, $\mu \rightarrow \sigma_D$ iff $\{\mu_\varepsilon(x) : x \in D \ \& \ \varepsilon > 0\}$ is a basis for the μ topology on 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*. In the next definition, the limit is taken with respect to the μ topology.

Definition 10. Let D be a domain with a map $\mu : D \rightarrow [0, \infty)^*$ that measures $X \subseteq D$. If $f : D \rightarrow \mathbb{R}$ is a map and $p \in X$ is not a compact element of D , then

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

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

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*: It is not difficult to show that $\{p\}$ is μ open iff p is compact. Notice too the importance of strict monotonicity of μ in Prop. 1: Without it, we could not define the derivative.

Definition 11. Let $f : D \rightarrow D$ be a function on a domain (D, μ) with a map μ that measures D at $p \in D \setminus K(D)$. If

$$df_\mu(p) := d(\mu f)_\mu(p)$$

exists, then we call this number *the informatic derivative* of f at p with respect to μ .

Our first example of this comes from calculus, and in fact provided the first relationship between domain theory and the differential calculus [2].

Theorem 2 (Martin [2]). *Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be a continuous map on the real line with $p \in \mathbb{R}$. If $f'(p)$ exists, then*

$$df_\mu[p] = |f'(p)|$$

where $\bar{f}(x) = f(x)$ is the canonical extension of f to \mathbf{IR} 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. Here is an example of a process with an informatic derivative that is *not* classical.

Example 4. The Bisection Method. To a continuous map $f : [a, b] \rightarrow \mathbb{R}$, we may assign the splitting

$$\text{split}_f : \mathbf{IR} \rightarrow \mathbf{IR}$$

$$\text{split}_f(x) = \begin{cases} \text{left}(x) & \text{if } \text{left}(x) \in C(f); \\ \text{right}(x) & \text{otherwise;} \end{cases}$$

where $C(f) = \{[a, b] \in \mathbf{IR} : f(a) \cdot f(b) \leq 0\}$, $\text{left}[x, y] = [x, (x + y)/2]$ and $\text{right}[x, y] = [(x + y)/2, y]$. At any $[r] \in \text{fix}(\text{split}_f)$, its rate of convergence is given by

$$d(\text{split}_f)[r] = 1/2.$$

Notice that this informatic derivative is *not* a classical derivative.

Other examples appear in [2], [6] and [5]. To be brief, let us say this: The *smaller* the derivative of a map on a domain at a fixed point, the quicker the convergence.

3 The Derivative at an Isolated Point

The reason that the informatic derivative requires 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 12. Let $f : D \rightarrow \mathbb{R}$ 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., $p \notin \min(D)$.

Trying to conceptualize these quantities isn’t such a good idea, so we resort to theorem proving.

Theorem 3. Let $f : D \rightarrow \mathbb{R}$ 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 $df_\mu(p) = d^+ f_\mu(p) = d^- f_\mu(p)$.

Proof. First suppose that $df_\mu(p)$ exists. Then by definition, for each $\varepsilon > 0$ there is a $\delta > 0$ such that

$$(\forall x) x \sqsubseteq p \ \& \ 0 < |\mu x - \mu p| < \delta \Rightarrow \left| \frac{f(x) - f(p)}{\mu x - \mu p} - df_\mu(p) \right| < \varepsilon.$$

Because μ measures D at p and p is not compact,

$$\mu_\delta(p) := \{x : x \sqsubseteq p \ \& \ |\mu x - \mu p| < \delta\} \in \text{approx}_\mu(p).$$

Thus, the sets used in the definition of $d^+ f_\mu(p)$ and $d^- f_\mu(p)$ are nonempty since for the set $V = \mu_\delta(p)$, we can write

$$(\forall x \in V) f(x) - f(p) \geq c^+ \cdot (\mu x - \mu p) \ \& \ f(x) - f(p) \leq c^- \cdot (\mu x - \mu p),$$

where $c^+ = df_\mu(p) - \varepsilon$ and $c^- = df_\mu(p) + \varepsilon$. But now if we can show that $d^+ f_\mu(p)$ and $d^- f_\mu(p)$ exist, the argument just given applies to give

$$d^- f_\mu(p) \leq df_\mu(p) + \varepsilon \ \text{and} \ df_\mu(p) - \varepsilon \leq d^+ f_\mu(p),$$

for each $\varepsilon > 0$, and hence that $d^- f_\mu(p) \leq df_\mu(p) \leq d^+ f_\mu(p)$.

With this last remark in mind, suppose we have numbers c^+ , c^- and μ open sets $V^+, V^- \in \text{approx}_\mu(p)$ such that

$$(\forall x \in V^+) f(x) - f(p) \geq c^+ \cdot (\mu x - \mu p)$$

and

$$(\forall x \in V^-) f(x) - f(p) \leq c^- \cdot (\mu x - \mu p).$$

Because p is not compact, we have $V = V^+ \cap V^- \in \text{approx}_\mu(p)$ and

$$c^+ \leq \frac{f(x) - f(p)}{\mu x - \mu p} \leq c^-$$

for all $x \in V \setminus \{p\}$. Letting $x \rightarrow p$ in the μ topology, we have $c^+ \leq df_\mu(p) \leq c^-$. Then $d^- f_\mu(p)$ and $d^+ f_\mu(p)$ both exist and satisfy $d^+ f_\mu(p) \leq df_\mu(p) \leq d^- f_\mu(p)$. Now our earlier remark applies, leaving $d^+ f_\mu(p) = df_\mu(p) = d^- f_\mu(p)$.

For the other direction, suppose $d^- f_\mu(p)$ and $d^+ f_\mu(p)$ both exist and satisfy $d^- f_\mu(p) \leq d^+ f_\mu(p)$. First we show that $d^- f_\mu(p) = d^+ f_\mu(p)$. Let $\varepsilon > 0$. Then there are numbers c^+, c^- and μ open sets $V^+, V^- \in \text{approx}_\mu(p)$ such that

$$d^+ f_\mu(p) - \varepsilon/2 < c^+ \leq d^+ f_\mu(p) \ \text{and} \ (\forall x \in V^+) f(x) - f(p) \geq c^+ \cdot (\mu x - \mu p)$$

and

$$d^- f_\mu(p) \leq c^- < d^- f_\mu(p) + \varepsilon/2 \ \text{and} \ (\forall x \in V^-) f(x) - f(p) \leq c^- \cdot (\mu x - \mu p).$$

Let $V = V^+ \cap V^- \in \text{approx}_\mu(p)$. Then the inequality

$$d^+ f_\mu(p) - \varepsilon/2 < c^+ \leq \frac{f(x) - f(p)}{\mu x - \mu p} \leq c^- < d^- f_\mu(p) + \varepsilon/2$$

holds for all x in the *nonempty* set $V \setminus \{p\}$. Using this inequality once, we obtain $d^+ f_\mu(p) < d^- f_\mu(p) + \varepsilon$ for each $\varepsilon > 0$ and hence $d^+ f_\mu(p) \leq d^- f_\mu(p)$, which gives $d^- f_\mu(p) = d^+ f_\mu(p)$. But going back to the inequality a second time reveals that the number $d := d^+ f_\mu(p) = d^- f_\mu(p)$ has the following property: For all $\varepsilon > 0$ there is a μ open set $V \in \text{approx}_\mu(p)$ such that

$$(\forall x \in V) x \neq p \implies d - \varepsilon < \frac{f(x) - f(p)}{\mu x - \mu p} < d + \varepsilon.$$

This is precisely the statement that $df_\mu(p)$ exists and is equal to d . □

The previous theorem justifies the following definition.

Definition 13. Let $f : D \rightarrow \mathbb{R}$ be a function on a domain 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 3, 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 4. Let $f : D \rightarrow \mathbb{R}$ be a function on an algebraic domain 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).

Proof. (i) \implies (ii): By assumption, $d^+ f_\mu(p)$ and $d^- f_\mu(p)$ both exist. We first show that the supremum exists. Let $x \in K(D)$ with $x \sqsubseteq p$ and $x \neq p$. Then

$$f(x) - f(p) \geq c \cdot (\mu x - \mu p)$$

for all $t \in V = \{x, p\} \in \text{approx}_\mu(p)$, where $c = (f(x) - f(p))/(\mu x - \mu p)$. Since $d^+ f_\mu(p)$ exists, we have $c \leq d^+ f_\mu(p)$. But then the set of all such c is bounded from above by $d^+ f_\mu(p)$. Thus, its supremum satisfies

$$\sup \left\{ \frac{f(x) - f(p)}{\mu x - \mu p} : x \in K(D) \cap \downarrow p, x \neq p \right\} \leq d^+ f_\mu(p).$$

Similarly, since $d^- f_\mu(p)$ exists, the same argument gives $d^- f_\mu(p) \leq c$, which implies that the set of all such c has an infimum which satisfies

$$d^- f_\mu(p) \leq \inf \left\{ \frac{f(x) - f(p)}{\mu x - \mu p} : x \in K(D) \cap \downarrow p, x \neq p \right\}.$$

(ii) \Rightarrow (i): To show that $df_\mu(p)$ exists, we must show that $d^+ f_\mu(p)$ exists, $d^- f_\mu(p)$ exists and that $d^- f_\mu(p) \leq d^+ f_\mu(p)$. We establish this by proving that $d^+ f_\mu(p)$ is the supremum in (ii), while $d^- f_\mu(p)$ is the infimum in (ii).

First the case of $d^+ f_\mu(p)$. To start, recall the definition of $d^+ f_\mu(p)$,

$$d^+ f_\mu(p) := \sup \{c : (\exists V \in \text{approx}_\mu(p)) (\forall x \in V) f(x) - f(p) \geq c \cdot (\mu x - \mu p)\}$$

Because D is algebraic and p is a compact element that is not minimal, the set of all such c is nonempty. To see why, let $y \sqsubseteq p$ be an element with $y \neq p$, which exists since p is not minimal. By algebraicity of D , y can be approximated by some compact element $x \ll y \sqsubseteq p$. Then as in the proof of (i) \Rightarrow (ii), we have a μ open set $V = \{x, p\} \in \text{approx}_\mu(p)$ such that

$$(\forall t \in V) f(t) - f(p) \geq c \cdot (\mu t - \mu p),$$

where $c = (f(x) - f(p))/(\mu x - \mu p)$.

Now that we know this set is nonempty, let $V \in \text{approx}_\mu(p)$ and c be any constant such that $f(x) - f(p) \geq c \cdot (\mu x - \mu p)$ for all $x \in V$. Let $y \in V$ be some element with $y \neq p$, which exists since $V \neq \{p\}$. By the algebraicity of D , y is the supremum of compact elements, and since V is μ open, there is some compact element $x \ll y$ with $x \in V$. Then $x \sqsubseteq p$ since $V \subseteq \downarrow p$, and $x \neq p$, since $y \neq p$. But this means

$$\frac{f(x) - f(p)}{\mu x - \mu p} \geq c,$$

and since we know that the set of all such elements has a supremum,

$$c \leq \sup \left\{ \frac{f(x) - f(p)}{\mu x - \mu p} : x \in K(D) \cap \downarrow p, x \neq p \right\},$$

which means that the set of all such c is bounded from above, and thus that $d^+ f_\mu(p)$ exists and satisfies

$$d^+ f_\mu(p) \leq \sup \left\{ \frac{f(x) - f(p)}{\mu x - \mu p} : x \in K(D) \cap \downarrow p, x \neq p \right\}.$$

The other inequality was established in the proof of (i) \Rightarrow (ii), assuming only the existence of $d^+ f_\mu(p)$. The dual argument handles the case of $d^- f_\mu(p)$. \square

Finally, the definition of derivative for selfmaps on a domain D .

Definition 14. Let $f : D \rightarrow D$ 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 $f : (D, \mu) \rightarrow (E, \lambda)$, as was done for the original formulation of the derivative in the continuous case [2], but in the present paper there are no applications warranting such an abstraction.

Example 5. Derivatives of list operations.

(i) The map $\text{first} : [S] \rightarrow [S]$, $\text{first}(a :: x) = [a]$, $\text{first}[] = []$. Using Theorem 4,

$$d(\text{first})_\mu(x) = d^+(\text{first})_\mu(x) = d^-(\text{first})_\mu(x) = 0,$$

for all $x \neq []$. At $x = []$, $d(\text{first})_\mu(x) = d^+(\text{first})_\mu(x) = 1 \geq 0 = d^-(\text{first})_\mu(x)$.

(ii) The map $\text{rest} : [S] \rightarrow [S]$, $\text{rest}(a :: x) = x$, $\text{rest}[] = []$. Using Theorem 4,

$$d(\text{rest})_\mu(x) = d^+(\text{rest})_\mu(x) = d^-(\text{rest})_\mu(x) = 1,$$

for all $x \neq []$. At $x = []$, $d(\text{rest})_\mu(x) = d^+(\text{rest})_\mu(x) = 1 \geq 0 = d^-(\text{rest})_\mu(x)$.

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.

4 Measuring the Length of an Orbit

Throughout this section, we assume that (D, μ) is an algebraic domain whose compact elements form a lower set. Some important examples of this are \mathbb{N}^* , \mathbb{N}^∞ , $[S]$, $\mathcal{P}\omega$, Σ^∞ , and $[\mathbb{N} \rightarrow \mathbb{N}_\perp]$. Computationally, this is not much of an assumption.

Theorem 5 (The Mean Value Theorem). Let $f : D \rightarrow D$ 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 $x \sqsubseteq p$.

Proof. Immediate by Theorem 4, in view of our assumption that the compact elements of D form a lower set. \square

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 15. Let $r : D \rightarrow D$ 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 2. *If a splitting $r : D \rightarrow D$ has a nontrivial compact orbit which ends at $p \in K(D)$, and $dr_\mu(p)$ exists, then $0 \leq dr_\mu(p) \leq 1$.*

Proof. By assumption, there is an $x \neq p$ with $\bigsqcup r^n(x) = p$. Since $dr_\mu(p)$ exists, by Theorem 4, we have

$$dr_\mu(p) = \sup \left\{ \frac{\mu r(x) - \mu(p)}{\mu x - \mu p} : x \in K(D) \cap \downarrow p, x \neq p \right\}.$$

Because r is a splitting, $\mu r(x) \leq \mu x$ for all x , which means that the supremum on the right is bounded above by 1. Hence $dr_\mu(p) \leq 1$.

To see that $dr_\mu(p) \geq 0$, we use the fact that $x \neq p$ and $rx \in K(D) \cap \downarrow p$, to deduce

$$0 \leq \frac{\mu r(x) - \mu p}{\mu x - \mu p} \leq dr_\mu(p),$$

which finishes the proof. \square

Theorem 6. *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 $\varepsilon > 0$.

Proof. First the case $dr_\mu(p) = 0$. Because $rx \sqsubseteq p$, we can apply the monotonicity of μ followed by the mean value theorem (5) to derive

$$0 \leq \mu r(x) - \mu p \leq 0 \cdot (\mu x - \mu p).$$

Then since μ measures the domain at p and $\mu r(x) = \mu p$, we must have $r(x) = p$.

For the other case, we simply apply the mean value theorem to obtain

$$0 \leq \mu r^n(x) - \mu p \leq (dr_\mu(p))^n (\mu x - \mu p),$$

and then the rest follows from arithmetic. □

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 6. 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 7. 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

$$\text{left}(x) = [x(1), \dots, x(m(\mu x) - 1)]$$

$$\text{right}(x) = [x(m(\mu x) + 1), \dots, x(\mu x)]$$

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.

We exploit the nature of the discrete: Given a compact element p that is not minimal, it 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 7. *Let $r : D \rightarrow D$ 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 $q \in \text{pred}(p)$ such that

$$r^n(x) = p \Leftrightarrow n = \frac{\mu x - \mu p}{\mu q - \mu p}.$$

Proof. Let $x \sqsubseteq p$ with $x \neq p$. Let $m \geq 0$ be the least integer with $r^{m+1}(x) = p$. Set $q = r^m(x)$. This element belongs to $\text{pred}(p)$. To see this, first note that q belongs to the dcpo $\downarrow p \setminus \{p\}$, and so there is $z \in \text{pred}(p)$ with $q \sqsubseteq z$. By assumption, $d^+r_\mu(z) = d^-r_\mu(z) = 1$, so applying the mean value theorem at z yields

$$\mu r(q) - \mu r(z) = \mu q - \mu z.$$

But $\bigsqcup r^n(z) = p$ and $z \in \text{pred}(p)$, so $r(z) = p$. And since $r(q) = p$ by definition, the expression on the left is zero, leaving $\mu q = \mu z$. Hence $q = z \in \text{pred}(p)$ by strict monotonicity of μ at the point z . Now we find the length of this orbit.

By the mean value theorem applied at q , we have $\mu r(x) - \mu r(q) = \mu x - \mu q$ for $x \sqsubseteq q$. By induction, we obtain

$$\mu r^{k+1}(x) - \mu r(q) = k \cdot \mu r(q) + \mu x - (k+1) \cdot \mu q$$

for $k \geq 0$ whenever $r^k(x) \sqsubseteq q$. Because $r(q) = p$, we can see that $r^n(x) = p$ iff $n = k+1$, where $\mu r^{k+1}(x) - \mu r(q) = 0$. Setting the equation above to zero and

solving for k yields

$$k = \frac{\mu x - \mu q}{\mu q - \mu p},$$

which means that $n = k + 1 = (\mu x - \mu p)/(\mu q - \mu p)$. □

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 8. In Example 5, we saw that the map $\text{rest} : [S] \rightarrow [S]$ is an example of the sort hypothesized in Theorem 7 with $p = []$. The predecessors of p are the one element lists

$$\text{pred}(p) = \{[x] : x \in S\}.$$

Thus, the last theorem says that

$$\text{rest}^n(x) = [] \Leftrightarrow n = \mu x,$$

for any $x \neq []$.

5 Complexity

Due to space limitations, we cannot go into great detail about how it is that algorithms can be represented in a manner favorable to the theory presented here. However, the interested reader can consult [2] and [4]. Nevertheless, we can still give the reader a good feel for how it is that our study of the informatic derivative gives us a new perspective on the complexity of algorithms.

Example 9. Linear search. To search a list x for a key k consider

$$\text{search} : [S] \times S \rightarrow \{\perp, \top\}$$

given by

$$\begin{aligned} \text{search}([], k) &= \perp \\ \text{search}(x, k) &= \top && \text{if first } x = k, \\ \text{search}(x, k) &= \text{search}(\text{rest } x, k) && \text{otherwise.} \end{aligned}$$

Let $D = [S] \times S^b$ – the product of $[S]$ with the set S ordered flatly. We measure this domain as $\mu(x, k) = \mu x$. Let $r : D \rightarrow D$ be the splitting $r(x, k) = (\text{rest } x, k)$.

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 7 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 10. Binary search. To search a sorted list x for a key k , we use the partially defined

$$\text{bin} : [S] \times S \rightarrow \{\perp, \top\}$$

is given by

$$\begin{aligned} \text{bin}([], k) &= \perp \\ \text{bin}(x, k) &= \top && \text{if mid } x = k, \\ \text{bin}(x, k) &= \text{bin}(\text{left } x, k) && \text{if mid } x > k, \\ \text{bin}(x, k) &= \text{bin}(\text{right } x, k) && \text{otherwise.} \end{aligned}$$

Again $D = [S] \times S^b$ and $\mu(x, k) = \mu x$. This time we consider the splitting $r : D \rightarrow D$ by

$$r(x, k) = \begin{cases} (\text{left } x, k) & \text{if mid } x > k; \\ (\text{right } x, k) & \text{otherwise.} \end{cases}$$

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 6,

$$n \leq \left\lceil \frac{\log(\mu x)}{\log(2)} \right\rceil + 1 = \lceil \log_2(\mu x) \rceil + 1,$$

since we know that the number on the right will *guarantee* that $r^n(x, k) = ([], k)$.

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 mentioned, this is identical to the way one compares zero finding methods in numerical analysis – by comparing the derivatives of mappings at fixed points.

6 Presentation

First, we have worked with total mappings, though the derivative works just as well on partial mappings [2]. The reason for this choice was our emphasis on the discrete setting: In it, the partial maps encountered in practice tend to have domains which are μ closed as well as μ open (hence domains in their own right).

Another issue concerns the very definition of df_μ itself. Theorem 3 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 (as in probability theory) 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 established about it in this paper: 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 $O(n)$.

7 Conclusion

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 world versus finite sequences of discrete objects. But what we are *taught* is that the continuous and discrete are ‘fundamentally different,’ and that the essence of this distinction is to be found in the sensibility of the limit concept for continuous objects, as compared to the discrete case where ‘limit’ has no meaning.

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 $r : D \rightarrow D$ 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 $r^n(x) \rightarrow p$ 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.

Finally, let us put this work in its proper perspective. It is evidence that the idea of an “informatic rate of change” is real – it is equally meaningful for understanding processes which manipulate information, the nature of the information is irrelevant (continuous/discrete) in this regard. We hope to inspire others to take a closer look at the idea. One feels certain that better approaches are possible, that much more can be said, and that many new and exciting applications await us.

References

1. S. Abramsky and A. Jung, *Domain theory*. In S. Abramsky, D. M. Gabbay, T. S. E. Maibaum, editors, *Handbook of Logic in Computer Science*, vol. III. Oxford University Press, 1994.
2. K. Martin, *A foundation for computation*. Ph.D. Thesis, Department of Mathematics, Tulane University, 2000.
<http://web.comlab.ox.ac.uk/oucl/work/keye.martin>
3. K. Martin, *A principle of induction*. Lecture Notes in Computer Science, vol. 2142, Springer-Verlag, 2001.
4. K. Martin, *A renee equation for algorithmic complexity*. Lecture Notes in Computer Science, vol. 2215, Springer-Verlag, 2001.
5. K. Martin, *The measurement process in domain theory*. Proceedings of the 27th International Colloquium on Automata, Languages and Programming (ICALP), Lecture Notes in Computer Science, vol. 1853, Springer-Verlag, 2000.
6. K. Martin, *Powerdomains and zero finding*. Electronic Notes in Theoretical Computer Science, vol 59.3, 2001, to appear.
7. K. Martin, *Unique fixed points in domain theory*. Proceedings of MFPS XVII, Electronic Notes in Theoretical Computer Science, vol. 45, 2001.