DIAGONAL ARGUMENTS AND LAWVERE’S THEOREM

DAN FRUMIN & GUILLAUME MASSAS

Abstract. Overview of the Lawvere’s fixed point theorem and some of its applications.

Category theory

Categories. A category $\mathcal{C}$ is a collection of objects $\mathcal{C}_0$ and arrows $\mathcal{C}_1$, such that each arrow $f \in \mathcal{C}_1$ has a domain and a codomain, both objects $\mathcal{C}_0$. We write $f : A \to B$ for an arrow $f \in \mathcal{C}_1$ with a domain $A \in \mathcal{C}_0$ and a codomain $B \in \mathcal{C}_0$.

Given two arrows $f : A \to B$ and $g : B \to C$, we can compose them, to obtain an arrow $g \circ f = gf : A \to C$.

The composition operation, when defined, is associative, i.e. $h(gf) = (hg)f$. We additionally require for each object $A \in \mathcal{C}_0$ an arrow $\text{id}_A : A \to A$ that is an identity element: $\text{id}_B \circ f = f \circ \text{id}_A = f$ for any $f : A \to B$.

By $\text{Hom}_\mathcal{C}(A, B)$ we denote a collection of arrows with a domain $A$ and a codomain $B$.

Example 1. Some prominent categories: $\text{Set}$, a category of sets and functions between them; $\text{Grp}$, a category of groups and group homomorphisms; a trivial category $\mathcal{1}$ consisting of one object and one identity arrow. The last example can be generalized as follows: pick a poset $(P, \leq)$, it induces a category with objects elements of $P$. The set $\text{Hom}_P(a, b)$ contains exactly one arrow if $a \leq b$, and $\text{Hom}_P(a, b) = \emptyset$ otherwise.

Finite products. We say that a category $\mathcal{C}$ has binary products if for every pari of objects $A, B \in \mathcal{C}_0$ there is an object $A \times B$ and arrows $\pi_1 : A \times B \to A, \pi_2 : A \times B \to B$ such that for any two arrows $f : X \to A, g : X \to B$ there is a unique arrow $(f, g) : X \to A \times B$ such that $\pi_1 \circ (f, g) = f$ and $\pi_2 \circ (f, g) = g$ (see the diagram below on the left).

The definition of binary products can be generalized to $n$-ary products for any finite $n$. In case $n = 0$ we speak of a terminal object $1$, with the following property (see the diagram on the right above): for each object $X$ there is a unique arrow $X \to 1$.

Example 2. In $\text{Set}$, a product $A \times B$ is just a cartesian product of two sets. The terminal object is then a one-element set $1 = \{\ast\}$.

Lawvere’s diagonal argument

Generalizing from the example of sets, we call maps $1 \to X$ global elements of $X$. In $\text{Set}$ such functions precisely correspond to members of $X$.

We can then state when some arrow $f : A \to B$ behaves like a “surjection” on global elements.
Definition 3. An arrow \( f : A \to B \) is point-surjective if for every global element \( b : 1 \to B \) there is a global element \( a : 1 \to A \) such that \( f \circ a = b \).

Equivalently, \( \text{Hom}(1, f) : \text{Hom}(1, A) \to \text{Hom}(1, B) \) is surjective.

Some categories with products support “function spaces”: objects \( B^A \), which somehow internalize arrows \( A \to B \) (in \textbf{Set}: a collection of arrows \( \text{Hom}(A, B) \) between sets is itself a set). For such a function space we can weaken the notion of point-surjectivity, requiring that an element of the preimage of some function \( g \) is only \textit{extensionally} equal to \( g \). Luckily, we can state this property without mentioning categorical exponents.

Definition 4. An arrow \( f : X \times A \to Y \) is weakly point-surjective if for every arrow \( g : X \to Y \) there is a global element \( a : 1 \to A \) such that for all \( x : 1 \to X \), \( f \circ \langle x, a \rangle = g \circ a \):

\[
\forall g \exists a \forall x (f \circ \langle x, a \rangle = g \circ a)
\]

One can think of such \( f \) as a series of functions \( f(-, a) \) such that for each \( g : X \to Y \) there is a function \( f(-, a) \) which is extensionally equal to \( g \).

Theorem 5 (Lawvere). Suppose that \( f : A \times A \to B \) is weakly point-surjective. Then every map \( t : B \to B \) has a fixed point, i.e., an element \( x : 1 \to B \) such that \( t x = x \).

Proof. Consider a composite \( t \circ f \circ \langle \text{id}_A, \text{id}_A \rangle : A \to B \):

\[
\begin{array}{ccc}
A \times A & \xrightarrow{f} & B \\
\Delta \uparrow & & \downarrow t \\
A & \xrightarrow{t \circ f \circ \langle \text{id}_A, \text{id}_A \rangle} & B \\
\end{array}
\]

By the assumption, there is a global element \( a : 1 \to A \) such that

\[
\forall (x : 1 \to A), (f \circ \langle x, a \rangle = t \circ f \circ \langle \text{id}_A, \text{id}_A \rangle \circ x = t(f \circ \langle x, x \rangle))
\]

In particular, for \( x = a \): \( f(a, a) = t(f(a, a)) \). Hence, \( f(a, a) \) is a fixed point of \( t \). \( \square \)

Corollary 6. Suppose that a map \( \neg : \Omega \to \Omega \) doesn’t have a fixed point. Then there is no weakly point-surjective map \( A \to \Omega^A \) for any \( A \).

Then we can obtain Cantor’s theorem in a straightforward way: since the negation map \( \neg : 2 \to 2 \) has a fixed-point, there is not surjective map \( A \to 2^A = \mathcal{P}(A) \). By substituting \( \Omega \) for 2 we obtain Cantor’s theorem in an arbitrary (non-degenerate) topos.

Russel’s Paradox and Unbounded Comprehension

Suppose there is a set-theoretic universe \( \mathcal{U} \in \textbf{Set} \), a “set of all sets”. To recover Russel’s paradox we consider a relation \( \epsilon : \mathcal{U} \times \mathcal{U} \to 2 \) where \( \epsilon(x, y) = 1 \iff x \in y \), and take the negation of the diagonal of \( \epsilon \):

\[
\begin{array}{ccc}
\mathcal{U} \times \mathcal{U} & \xrightarrow{\epsilon} & 2 \\
\Delta \uparrow & & \downarrow \neg \\
\mathcal{U} & \xrightarrow{\neg \circ \epsilon \circ \Delta} & 2 \\
\end{array}
\]

The composite \( \neg \circ \epsilon \circ \Delta \) is a map \( \mathcal{U} \to 2 \), that is, a predicate on \( \mathcal{U} \) that is \textit{true} on the sets \( x \) for which \( \neg(x \in x) \) holds; i.e., for sets that do not contain themselves. Now, for obtaining Russel’s paradox we would have to show that \( \epsilon \) is weakly-point surjective. What does it mean for \( \mathcal{U} \) specifically? It would mean that for any predicate \( \phi : \mathcal{U} \to 2 \) on sets there exists a set \( x \in \mathcal{U} \) (corresponding to a map \( x : 1 \to \mathcal{U} \)) such that the members of \( x \) are exactly such sets that satisfy \( \phi \):

\[
\exists x \in \mathcal{U} \forall y \in \mathcal{U} (y \in x \iff \phi(y))
\]
This rule is exactly the unbounded comprehension scheme for \( \mathcal{U} \)! As you can see, employing Lawvere’s analysis for this paradox pinpoints exactly to the problematic part: the unbounded comprehension schema for \( \mathcal{U} \). Restricting the comprehension schema to already-defined sets is exactly the fix that was utilized in axiomatic set theory. Notice that this analysis shows that the issue does not lie in self-reference or the size of \( \mathcal{U} \) per se. After all, the universe \( \mathcal{U} \) does not have to contain “all” sets; we can replace the word “set” in the previous paragraph by “\( \mathcal{U} \)-set” and the argument would still go through.

**Lindenbaum-Tarski Categories and incompleteness**

Consider a first-order theory \( T \). We form \( \mathcal{C}(T) \) a classifying category of \( T \) in the following way: objects of \( \mathcal{C}(T) \) are generated by a sort object \( A \) (more object if the theory is multi-sorted), and a dummy object \( 2 \), by closure under products. Thus, the objects of \( \mathcal{T} \) are of the form \( A^n \times 2^m \). A map \( \varphi : A^n \to 2 \) is an equivalence class of provably equivalent formulas \( \varphi \) of \( n \) variables. A map \( t : A^n \to A \) is a class of provably equal terms with \( n \) free variables. In particular, maps \( 1 \to 2 \) are sentences of \( \mathcal{T} \), and maps \( 1 \to A \) are definable constants/terms of \( \mathcal{T} \).

A theory is **consistent** if the collection of maps \( 1 \to 2 \) contains at least two elements \( \text{true}, \text{false} \), corresponding to statements that are provable in the theory, and statements that are refutable in the theory. A theory is **complete** if the collection of maps \( 1 \to 2 \) is exactly \( \{ \text{true}, \text{false} \} \), i.e. every sentence is either provable or refutable.

**Undefinability of \( \text{sat} \).** Suppose that the satisfiabilty predicate is definable in \( T \):

\[
\vdash \text{sat}(a, \uparrow \varphi^\uparrow) \leftrightarrow \varphi(a)
\]

for all \( \varphi, a \).

In categorical terms, we have a Gödel encoding, \( \uparrow - \downarrow : \text{Hom}(A^n, 2) \to \text{Hom}(1, A) \), and a formula \( \text{sat} : A \times A \to 2 \), such that for any \( \varphi : A \to 2 \), and for all \( a : 1 \to A \), \( \text{sat}(a, \uparrow \varphi^\uparrow) = \varphi(a) \). But this is exactly the condition for weak point-surjectivity! Hence, every function \( 2 \to 2 \) has a fixed point, and we are in an inconsistent theory.

**Undefinability of truth.** We say that truth is definable in a theory, if there is a formula \( T \), such that

\[
\vdash T(\uparrow \varphi^\uparrow) \leftrightarrow \varphi
\]

So it is very much like \( \text{sat} \), but only for sentences. Categorically, we can say that \( T : A \to 2 \) is a truth predicate, if \( \text{Hom}(1, T) : \text{Hom}(1, A) \to \text{Hom}(1, 2) \) is a retract of \( \uparrow - \downarrow : \text{Hom}(1, 2) \to \text{Hom}(1, A) \); or, \( T \circ \uparrow \varphi^\uparrow = \varphi \). So, suppose that \( T \) has a truth predicate, and suppose further that it supports “substitution”:

\[
\text{T} \vdash \text{subst}(a, \uparrow \varphi^\uparrow) = \uparrow \varphi(a)^\uparrow
\]

In that case, we can define \( \text{sat} \) as the composite \( T \circ \text{subst} \).

**Incompleteness.** A provability predicate is a predicate \( P \) such that

\[
\text{T} \vdash P(\uparrow \varphi^\uparrow) \quad \text{iff} \quad \text{T} \vdash \varphi
\]

In categorical terms, \( P \circ \uparrow \varphi^\uparrow = \varphi \) given that both \( P \circ \uparrow \varphi^\uparrow \) and \( \varphi \) take value in \( \{ \text{true}, \text{false} \} \). But if \( T \) is complete, then the provability predicate is also a truth predicate.
Assemblies and the halting problem

Consider the category $\text{Asm}$ of assemblies. The objects are pairs $(X, \models_X)$ where $X \in \text{Set}$ and $\models_X \subseteq \mathbb{N} \times X$ such that for each $x \in X$ there is at least one number $n \models_X x$. Elements $m$ such that $m \models_X x$ are called realizers of $x$ and we say that $m$ realizes $x$. A map $f : (X, \models_X) \to (Y, \models_Y)$ is a morphism of assemblies if there is a partial computable function $\phi$ such that whenever $n \models_X x$, $\phi(n)$ terminates and $\phi(n) \models_Y f(y)$. We say that $\phi$ tracks or realizes $f$. The products in $\text{Asm}$ are given by surjective pairings. There is a natural numbers object $\mathbb{N}$ in $\text{Asm}$ given by $(\mathbb{N}, \models_{\mathbb{N}})$ where $n \models_{\mathbb{N}} m$ iff $n = m$.

**Proposition 7.** The morphisms $\mathbb{N} \to \mathbb{N}$ are exactly (total) computable functions.

**Definition 8.** $\text{Asm}$ has all finite types. For instance, the object 2 is given by $(\{0,1\}, \models_2)$ where $i \models_2 j$ iff $i = j$.

Suppose that the halting problem is decidable. We define a morphism $\text{halt} : \mathbb{N} \times \mathbb{N} \to 2$ such that $\text{halt}(n, m) = 1$ iff the partial computable function $\{n\}(\_ : \mathbb{N} \to \mathbb{N}$ terminates on the input $m$. For $\text{halt}$ to be weak point-surjective we must show that for any morphism $f : \mathbb{N} \to 2$ there is a number $n$ such that $\text{halt}(n, m) = f(m)$ for all $m$, i.e. $\{n\}(m)$ terminates iff $f(m) = 1$. How do we construct such $n$? Well, $f$ is tracked by some computable $\phi$, so $n$ is just the Gödel code of an algorithm/function that runs $\phi(m)$ on input $m$ and terminates if the output of $\phi(m)$ is 1, and diverges otherwise.

**Obtaining fixed points**

**Retractions & the $Y$-combinator.** An epimorphism $r : E \to B$ is said to be split, if there is a map $s : B \to E$ in the opposite direction such that $r \circ s = \text{id}_B$. This is equivalent to saying that $\text{Hom}(A, r) : \text{Hom}(A, E) \to \text{Hom}(A, B)$ is surjective for all $A$. Clearly, any split epimorphism is point-surjective, the choice for the witness for the existential quantifier is given by $s$. (However, not every epimorphism is point-surjective, and not every point-surjective map is epi)

Consider the category $\text{CPO}_\bot$ of direct-complete partial orders with $\bot$. It is a cartesian closed category with a reflexive element $U$; that is an object $U \neq 1$ such that there is a retraction $r : U \to U^\bot$. Such a domain $U$ provides a model for untyped $\lambda$-calculus; furthermore, a complete class of models of $\lambda$-calculus arises in such a way: see section 5.5 in Barendregt’s book.

Anyway, what follows is that every map $t : U \to U$ has a fixed point; this fixed point is exactly the one given by the $Y$-combinator!

By computation, a fixed point of $t$ is given by $\tau \circ \Delta \circ s((t \circ \tau \circ \Delta))$. Mixing syntax and semantics informally we have $\tau \circ \Delta = ab$ and $s(x \mapsto g(x)) = \lambda x.g(x)$, so the fixed point is

$$(s((t \circ \tau \circ \Delta)))(s((t \circ \tau \circ \Delta))) = (\lambda x.(t \circ \tau \circ (x, x)))(\lambda x.(t \circ \tau \circ (x, x))) = (\lambda x.(t(xx)))(\lambda x.(t(xx)))$$

which is exactly $Y(t)$.

**Enumerations of r.e. sets.** Consider an assembly $\Sigma \in \text{Asm}$ defined as an underlying set $\{\top, \bot\}$ with the realizability relation

$$n \models \top \iff \{n\}(n)$$

Such $\Sigma$ is called a r.e. subobject classifier or a r.e. dominance.

Morphisms $X : \mathbb{N} \to \Sigma$ are recursively-enumerable sets. Given a map $X : \mathbb{N} \to \Sigma$ tracked by $\phi$ we define a set $\overline{X} = \{ x \in \mathbb{N} \mid X(x) = \top \}$. To check that $n \in \overline{X}$ we attempt to compute $\{\phi(n)\}(\phi(n))$. If $\{\phi(n)\}(\phi(n))$ terminates, then $n \in \overline{X}$. Similarly, given a r.e. set $Y$ we put $\overline{Y}(n) = \top \iff \{n \in Y\}$; $\overline{Y}$ is then tracked by a computable function that sends $n$ to the Gödel code of the decision procedure $x \mapsto [n \in X]$.

The exponent $\Sigma^\mathbb{N}$ is then the collection of r.e. sets. We know that there is an enumeration of r.e. sets, thus a weakly point-surjective $W : \mathbb{N} \to \Sigma^\mathbb{N}$. Hence, by Lawvere’s theorem every map $\Sigma \to \Sigma$ has a fixed point. It immediately follows that negation is not definable on $\Sigma$ and hence r.e. sets are not closed under complements.
Note that $\Sigma^N \simeq \Sigma^{N \times N} \simeq \Sigma^{N^N}$, so every map $\Sigma^N \to \Sigma^N$ has a fixed point as well. We can identify the exponent $\Sigma^N$ with an assembly $(RE, W)$ where $RE$ is the set of r.e. subsets of $N$ and $W(A) = \{ e \mid A = W_e \}$ for an enumeration $\{ W_i \}$ of r.e. sets.

A map $F : (RE, W) \to (RE, W)$ is an enumeration operator: $F(W_e) = W_{\phi(e)}$, for some computable $\phi$. The Lawvere’s argument states that every such operator has a fixed point: $W_k = W_{\phi(k)}$. Consider a computable $\phi$ which for every $n$ outputs the r.e. index of a r.e. set that is just a singleton $\{ n \}$, that is $W_{\phi(n)} = \{ n \}$. By the existence of a fixed point we have a number $k$ such that $W_k = \{ k \}$.

References


Appendix

We would like to make the following additional remark.

A finer analysis of the argument might reveal the following fact: it is not necessary to take the diagonal map $\Delta : A \to A \times A$. One can easily take any other map $(\text{id}_A, k)$ for a “good” $k : A \to A$ (say, if $k$ is an isomorphism). Then the fixed-point for a map $t : B \to B$ can be constructed from

$$t(f(x, k(x))) = f(x, b)$$

If $k$ is an isomorphism, then we can find such $x$ that $k(x) = b$. Then we obtain the fixed point in a similar manner.