Lawvere's Theorem

Tobias Reinhart   Sebastian Stengele

February 4, 2022

As famously proven by Cantor, for any set X, the cardinality of its power set $\mathcal{P}(X)$ is strictly bigger than that of X. Equivalently, for any set X there cannot exist a surjective function $f : X \to \mathcal{P}(X)$. This statement, also known as Cantor's theorem, is usually proven by assuming the existence of such a surjection f. Next, one constructs the following subset of X

$$S := \{x \in X \mid x \notin f(x)\} \tag{1}$$

This is clearly well-defined. As f was assumed to be surjective, there exist a $x_S \in X$ such that $f(x_S) = S$. Finally, one concludes that

$$x_S \in S \Leftrightarrow x_S \notin S \tag{2}$$

so the assumption must have been wrong and such a surjection cannot exist.

Identifying subsets of $A \subseteq X$ with their characteristic function $\chi_A : X \to 2$, i.e. using $\mathcal{P}(X) \cong \{\chi : X \to 2\} = 2^X$, where $2 := \{0, 1\}$, replaces f with $\chi_f : X \to 2^X$. Besides, now

$$\chi_S(x) := \text{not} \circ \chi_f(x)(x) \tag{3}$$

where $\text{not} : 2 \to 2$ such that $\text{not}(1) = 0$ and $\text{not}(0) = 1$. Now choose $x_s \in X$ as before, such that $\chi_f(x_S) = \chi_S$

$$\chi_S(x_S) = \text{not} \circ \chi_f(x_S)(x_S) = \text{not} \circ \chi_S(x_S) \tag{4}$$

Observe that this last statement yields a contradiction as not has no fixed point, i.e. there exists no $a \in 2$ with $\text{not}(a) = a$. It is now obvious, that the same proof can easily be generalized to $f' : X \to X \to Y$, as long as there exists a function $Y \to Y$ without a fixed point [1]. Proofs that obey this structure are often called diagonal arguments or proofs by diagonalization. Many important theorems can be proven in this way, for instance the undecidability of the halting problem, Gödel's incompleteness theorem, Tarski's non-definability of truth to only name a few. All these can be seen as instances of a more abstract theorem, that uses the language of cartesian closed categories to capture the essence of these diagonal arguments, the so-called Lawvere's fixed point theorem.

1 Lawvere's fixed point theorem

Definition 1 (category with finite products, [2])**.** *A category* $\mathbf{C}$ *is said to have all* finite products *if for any finite collection* $C_1, ..., C_n$ *of objects of* $\mathbf{C}$ *their product* $C_1 \times ... \times C_n$ *with projections* $p_i : C_1 \times ... \times C_n \to C_i$ *for* $i = 1, ..., n$ *exists in* $\mathbf{C}$

$$\begin{tikzcd}[column sep=small] & & & C_1 \times ... \times C_n \arrow[dlll, "p_1"'] \arrow[dll, "p_2"] \arrow[dr, "p_{n-1}"'] \arrow[drr, "p_n"] & & \\ C_1 & C_2 & \dots & & C_{n-1} & C_n \end{tikzcd} \tag{5}$$

Note that if $\mathbf{C}$ has all finite products, in particular it contains the empty product 1, which by the universal property of the product is terminal in $\mathbf{C}$. If the above condition only holds for binary products, i.e. for $n = 2$ we say that $\mathbf{C}$ has all binary products. It can be shown that if $\mathbf{C}$ has a terminal object as well as all binary products, then $\mathbf{C}$ has all finite products [2, Proposition 3.7.1].

Lemma 1 ([2, Proposition 3.7.1 & Exercise 3.7.1])**.** *Let* $\mathbf{C}$ *be a category which has all finite products, then the following* 4 *statements hold.*

1. $(\times) : \mathbf{C} \times \mathbf{C} \to \mathbf{C}$ *with* $(A, B) \mapsto A \times B$ *defines a bifunctor that on objects, maps* $(A, B) \mapsto A \times B$ *and on arrows maps* $(f : A \to C, g : B \to D)$ *to the unique arrow* $f \times g : A \times B \to C \times D$ *that make the following diagram commutative*

$$\begin{tikzcd} B \arrow[r, "g"] & D \\ A \times B \arrow[u, "p_2"] \arrow[r, "f \times g"] \arrow[d, "p_1"] & C \times D \arrow[u, "p_2"'] \arrow[d, "p_1"] \\ A \arrow[r, "f"] & B \end{tikzcd} \tag{6}$$

2. $(\times)$ *is associative up to natural isomorphism, i.e. for any three objects* A, B, C*, there exists an isomorphism*

$$\alpha_{A,B,C} : A \times (B \times C) \to (A \times B) \times C \tag{7}$$

that is natural in A*,* B *and* C*.*

3. *Up to natural isomorphism,* 1 *serves as two sided identity of* $(\times)$*, i.e. there exist two natural isomorphisms* π *and* ρ *such that for any object* C

$$\pi_C : 1 \times C \to C \quad \textit{and} \quad \rho_C : C \times 1 \to C \tag{8}$$

4. *The diagonal* Δ *with component at* C $\Delta_C : C \to C \times C$*, being uniquely defined by commutativity of the following diagramm*

$$
\begin{tikzcd}
 & C \arrow[dl, "\mathrm{id}_C"'] \arrow[d, "\Delta_C"] \arrow[dr, "\mathrm{id}_C"] & \\
C & C \times C \arrow[l, "p_1"] \arrow[r, "p_2"'] & C
\end{tikzcd}
\tag{9}
$$

defines a natural transformation between the identity functor and the functor mapping objects as $C \mapsto C \times C$ *and arrows as* $(f : C \to D) \mapsto (f \times f : C \times C \to D \times D)$*.*

Note that item 3 in particular implies that $1 \cong 1 \times 1$. Thus also 1×1 is terminal in $\mathbf{C}$ and it follows that

$$
\begin{aligned}
\pi_1 &= \rho_1 \\
\pi_1^{-1} &= \rho_1^{-1} = \Delta_1
\end{aligned}
\tag{10}
$$

Definition 2 (cartesian closed category, [2])**.** *A category* $\mathbf{C}$ *is called* cartesian closed *if*

1. $\mathbf{C}$ *has all finite products*
2. *for each object* B *of* $\mathbf{C}$ *the functor* $- \times B : \mathbf{C} \to \mathbf{C}$ *mapping objects* $C \mapsto C \times B$ *and arrows* $f \mapsto f \times \mathrm{id}_B$ *has a right adjoint* $-^B : \mathbf{C} \to \mathbf{C}$*, i.e. for any objects* A, B

$$
\mathrm{Hom}_{\mathbf{C}}(A \times B, C) \cong \mathrm{Hom}_{\mathbf{C}}(A, C^B) \tag{11}
$$

natural both in A *and* C*.*

Given an arrow $f : A \times B \to C$ we will denote its transpose as $\lambda_B(f) : A \to C^B$. Conversely, given $g : A \to C^B$, its transpose will be denoted as $\epsilon_B(g) : A \times B \to C$.

The counit of the adjunction $\epsilon_{B,A} := \epsilon_B(\mathrm{id}_{A^B}) \in \mathrm{Hom}_{\mathbf{C}}(A^B \times B, A)$ is called evaluation map. If we now take any $f : A \times B \to C$, and $\lambda_B(f) : A \to C^B$, naturality of eq. (11) implies that the following diagram commutes

$$
\begin{tikzcd}
A \times B \arrow[d, "\lambda_B(f) \times \mathrm{id}_B"'] \arrow[dr, "f"] & \\
C^B \times B \arrow[r, "\epsilon_{B,C}"'] & C
\end{tikzcd}
\tag{12}
$$

The pair $(C^B, \epsilon_{B,C})$ can thus be characterized by the following universal property: for any $f : A \times B \to C$ there exists a unique $\lambda_B(f) : A \to C^B$ such that

$f = \epsilon_{B,C} \circ (\lambda_B(f) \times \mathrm{id}_B)$. Considering any category with all binary products and any two of its objects A, B, an object A^B for which there exists such an evaluation map $\epsilon_{B,A} : A^B \times B \to A$ such that the pair $(A^B, \epsilon_{B,A})$ satisfies the above condition is called exponential object. It can be shown that the second requirement in definition 2 can equivalently be replaced by the condition that for any two objects A, B in $\mathbf{C}$, there exists an exponential object A^B [3, 4].

Definition 3 (point surjectivity, [5, 6])**.** *Let* $\mathbf{C}$ *be a category and* 1 *be terminal in* $\mathbf{C}$*, an arrow* $f : A \to B$ *is called* point surjective *if for every arrow* $b : 1 \to B$ *there exists an arrow* $a : 1 \to A$ *that satisfies* $f \circ a = b$*.*

$$\begin{tikzcd} 1 \arrow[d, "a"'] \arrow[dr, "b"] & \\ A \arrow[r, "f"'] & B \end{tikzcd} \tag{13}$$

If C *additionally has all binary products, an arrow* $f : A \times B \to C$ *is called* weakly point surjective*, if for every* $g : B \to C$ *there exists an arrow* $x_g : 1 \to A$ *such that for every* $b : 1 \to C$*,* $f \circ (x_g \times b) \circ \Delta_1 = g \circ b$*, i.e. the following diagram commutes.*

$$\begin{tikzcd} 1 \arrow[d, "\Delta_1"'] \arrow[dr, "b"] & & \\ 1 \times 1 \arrow[d, "x_g \times b"'] & B \arrow[dr, "g"] & \\ A \times B \arrow[rr, "f"] & & C \end{tikzcd} \tag{14}$$

Let $\mathbf{C}$ *be cartesian closed, an arrow* $f : A \to C^B$ *is called* weakly point surjective *if its transpose* $\epsilon_B(f)$ *is weakly point surjective.*

Remark. *Note that using eq. (12) together with the functoriality of* $\times$ *and eq. (11), we can rewrite the condition for weak point surjectivity of* $f : A \to C^B$ *as*

$$\epsilon_{B,C} \circ ((f \circ x_g) \times b) \circ \Delta_1 = g \circ b \tag{15}$$

This is how weak point surjectivity is presented in [5].

Example 1. *Consider the category* ***Set*** *with terminal object* 1 *any one element set and products given by the set theoretic cartesian product. In* ***Set*** *morphisms* $a : 1 \to A$ *can uniquely be identified with elements of* A*, via their only value,* $a(1) \in A$*. Thus a function* $f : A \to B$ *is point surjective if and only if for all*

$b \in B$ there exists an $a \in A$ with $f(a) = b$, i.e. if and only if f is surjective. A function $f : A \times B \to C$ is weakly point surjective if for any $g : B \to C$ there exists an $x_g \in A$ such that for any $b \in B$, $f(x_g, b) = g(b)$. By functional extensionality, this is equivalent to $f(x_g, -) = g$, so also in this case f, or more precisely its transpose, is surjective as a function $f : A \to C^B$.

Example 2. *Consider the category **Grp** of groups and group homomorphisms. The trivial, one-element group 1 is terminal in **Grp**. The product of groups is given as the direct product. Note that any group homomorphism $f : A \to B$ necessarily maps the identity of A to the identity of B. Hence the trivial group 1 is also initial in **Grp** and therefore any morphism in **Grp** is point surjective. For the same reason every morphism $f : A \times B \to C$ is weakly point surjective.*

Lemma 2. *Let $f : A \to C^B$ be point surjective. Then f is weakly point surjective.*

Proof. For f to be weakly point surjective, we need to show that for all $g : B \to C$ there exists x_g such that for all $b : 1 \to B$ the following holds:

$$g \circ b = \epsilon_B(f) \circ (x_g \times b) \circ \Delta_1 \tag{16}$$

This is essentially diagram eq. (14) with f replaced by $\epsilon_B(f)$. Using the following identity for the counit (which follows from eq. (12))

$$\epsilon_B(f) = \epsilon_{B,C} \circ (f \times \mathrm{id}_B) \tag{17}$$

the surjectivity of $f : A \to C^B$ (choosing x_g accordingly)

$$f \circ x_g = \lambda_B(g \circ \pi_B) : 1 \to C^B \tag{18}$$

and eq. (12) ,we can rewrite (eq. (16)):

$$\begin{aligned} \epsilon_B(f) \circ (x_g \times b) \circ \Delta_1 &= \epsilon_{B,C} \circ (f \times \mathrm{id}_B) \circ (x_g \times b) \circ \Delta_1 \\ &= \epsilon_{B,C} \circ (\lambda_B(g \circ \pi_B) \times b) \circ \Delta_1 \\ &= \epsilon_{B,C} \circ (\lambda_B(g \circ \pi_B) \times \mathrm{id}_B) \circ (\mathrm{id}_1 \times b) \circ \Delta_1 \\ &= g \circ \pi_B \circ (\mathrm{id}_1 \times b) \circ \Delta_1 \\ &= g \circ b \circ \pi_1 \circ \Delta_1 \\ &= g \circ b \end{aligned} \tag{19}$$

Where we additionally used the naturality of π_B and the fact that $\pi_1^{-1} = \Delta_1$. $\square$

Definition 4 (fixed point, [5])**.** *Let $\mathbf{C}$ be a category with terminal object 1. An arrow $f : A \to A$ is said to have a* fixed point *if there exists an arrow $a : 1 \to A$ such that $f \circ a = a$.*

Theorem 1 (Lawvere's fixed point theorem, [5, Theorem 1.1])**.** *Let $\mathbf{C}$ be a cartesian closed category and $f : A \to B^A$ be weakly point surjective, then every arrow $g : B \to B$ has a fixed point.*

Proof. Consider the transpose of f, $\epsilon_A(f) : A \times A \to B$. Take any $g : B \to B$ and construct $h := g \circ \epsilon_A(f) \circ \Delta_A : A \to B$

$$A \xrightarrow{\Delta_A} A \times A \xrightarrow{\epsilon_A(f)} B \xrightarrow{g} B \tag{20}$$

As f was assumed to be weakly point surjective there exists an arrow $x_h : 1 \to A$ such that for every $a : 1 \to A$ it holds that

$$g \circ \epsilon_A(f) \circ \Delta_A \circ a = \epsilon_A(f) \circ (x_h \times a) \circ \Delta_1 \tag{21}$$

Now chose $a = x_h$. Naturality of Δ implies that the following diagram commutes

$$\begin{tikzcd} 1 \arrow[r, "\Delta_1"] \arrow[d, "x_h"] & 1 \times 1 \arrow[d, "x_h \times x_h"] \\ A \arrow[r, "\Delta_A"] & A \times A \end{tikzcd} \tag{22}$$

Thus, in total we find that

$$g \circ \epsilon_A(f) \circ \Delta_A \circ x_h = \epsilon_A(f) \circ (x_h \times x_h) \circ \Delta_1 = \epsilon_A(f) \circ \Delta_A \circ x_h \tag{23}$$

which simply states that $\epsilon_A(f) \circ \Delta_A \circ x_h : 1 \to B$ is a fixed point of g. □

Corollary 1 ([5]). *Let* $\mathbf{C}$ *be a category that has all finite products and* $f : A \times A \to B$ *be weakly point surjective, then every* $g : B \to B$ *has a fixed point.*

Proof. Follows immediately from the proof of the previous theorem. □

2 Applications of Lawvere's fixed point theorem

2.1 Cantor's theorem

Cantor's theorem states that for any set X there cannot exist a surjective function $f : X \to \mathcal{P}(X)$ from X to its power set. This can be seen as an instance of Lawvere's fixed point theorem by considering the category **Set** of sets and functions between them.

Lemma 3 ([3]). *The category **Set** is cartesian closed with*

1. *product* $\times$ *the cartesian product of sets*
2. *for sets* A, B*, their exponential provided by the function set* $A^B := \{f : B \to A\}$ *together with the conventional function evaluation* $\epsilon_{B,A}(f, b) := f(b)$

Theorem 2 (Cantor, [5], [7, Theorem 1]). *Let* X *be a set and* $f : X \to \mathcal{P}(X)$*, then* f *is not surjective.*

Proof. Consider the category **Set**. By the previous lemma **Set** is cartesian closed. Let $1 := \{0\}$ and $2 := \{0, 1\}$. Note that 1 is terminal in **Set**. We can identify $\mathcal{P}(X)$ with 2^X, by the usual bijection that maps subsets $Y \subseteq X$ to their characteristic function $\chi_Y : X \to 2$. To be more precise, $\mathcal{P}(X) \cong 2^X$, naturally in X. This is precisely the representability of the contravariant powerset functor. Thus a function from X to its powerset can be seen as $f : X \to 2^X$. Now assume there exists such a function f that is surjective. Let $\bar{y} : 1 \to 2^X$ and $y := \bar{y}(0)$. As f is surjective there exists a $x \in X$ such that $f(x) = y$, so defining $\bar{x} : 1 \to X$ by $\bar{x}(0) := x$ shows that $f \circ \bar{x} = \bar{y}$. Hence, f is point surjective and thus by (lemma 2) in particular weakly point surjective. But then Lawvere's theorem applies, stating that every function $2 \to 2$ has a fixed point. But $\mathrm{not} : 2 \to 2$, defined by $\mathrm{not}(0) = 1$, $\mathrm{not}(1) = 0$ clearly has no fixed point. Hence no such surjective $f : X \to 2^X$ can exist. $\square$

2.2 The halting problem

The halting problem states that there cannot be a computer program/an algorithm H that when given the description of any other computer program T as well as an input X of T decides whether T halts on input X or loops forever. Also this can be proven by applying Lawvere's fixed point theorem to a suitable category. To that end we consider the category of assemblies **Asm**.

Definition 5 (Category of assemblies, [8, 9])**.** *Fix some notion of computability of functions $\mathbb{N}^k \to \mathbb{N}^l$. The category of assemblies **Asm** consists of*

1. *objects $(X, \sim_X)$ where X is some set and $\sim_X$ is a relation $\sim_X \subseteq \mathbb{N} \times X$*
2. *morphisms $f : (X, \sim_X) \to (Y, \sim_Y)$ where f is a function from X to Y such that there exists a (partial) computable function $\phi_f : \mathbb{N} \to \mathbb{N}$ that for all $x \in X$ and all $n \in \mathbb{N}$ with $n \sim_X x$ satisfies*

$$n \sim_X x \Rightarrow n \in \mathrm{dom}(\Phi_f) \quad \textit{and} \quad \phi_f(n) \sim_Y f(x) \tag{24}$$

We say that ϕ_f realizes f and whenever $n \sim_x x$, that n realizes x. For any $S \subseteq \mathbb{N}$, **Asm** contains an object $S' := (S, \sim_{S'})$, where for any $n \in S$, $m \sim_{S'} n$ by definition if and only $n = m$. Hence, **Asm** in particular contains $2'$ and $\mathbb{N}'$. Besides, **Asm** has all finite products. Fix some computable bijection with computable inverse

$$P : \mathbb{N} \times \mathbb{N} \to \mathbb{N} \tag{25}$$

Given any two objects $(X, \sim_X)$ and $(Y, \sim_Y)$, their binary product has the cartesian product $X \times Y$ as underlying set, and

$$P(n, m) \sim_{X \times Y} (x, y) :\Leftrightarrow n \sim_X x \text{ and } m \sim_Y y \tag{26}$$

Lemma 4. *For any $S \subseteq \mathbb{N}$, $\mathrm{Hom}_{\mathbf{Asm}}(\mathbb{N}', S')$ is the set of total computable functions from $\mathbb{N}$ to S.*

Proof. Take $f \in \mathrm{Hom}_{\mathbf{Asm}}(\mathbb{N}', S')$. Clearly f is total. Take any $n \in \mathbb{N}$. As $n \sim_{\mathbb{N}'} n$, $f(n) \sim_{S'} \phi_f(n)$ and thus by definition of $\sim_{S'}$, $f(n) = \phi_f(n)$, so in fact $f = \phi_f$ and thus f is in particular computable. Vice versa, any total computable function $g : \mathbb{N} \to S$ clearly defines an arrow $\mathrm{Hom}_{\mathbf{Asm}}(\mathbb{N}', S')$, simply by $\phi_g =: g$. $\square$

Lemma 5. *The arrows in* $\mathrm{Hom}_{\mathbf{Asm}}(\mathbb{N}' \times \mathbb{N}', S')$ *are exactly the total computable functions from* $\mathbb{N} \times \mathbb{N}$ *to* S*.*

Proof. The proof follows essentially the same arguments as the proof of the previous lemma. Take any $f \in \mathrm{Hom}_{\mathbf{Asm}}(\mathbb{N}' \times \mathbb{N}', S')$ and any $(x, y) \in \mathbb{N} \times \mathbb{N}$. Now $P(x, y)$ realizes (x, y) and hence $\phi_f \circ P(x, y) = f(x, y)$. As both ϕ_f and P are computable, so is their composite f. Conversely, any computable $g : \mathbb{N} \times \mathbb{N} \to S$ defines an arrow in $\mathrm{Hom}_{\mathbf{Asm}}(\mathbb{N}' \times \mathbb{N}', S')$, ϕ_g is simply given as $\phi_g := g \circ P^{-1}$. $\square$

Theorem 3 (Halting)**.** *Let e be some Gödel numbering of partial computable functions from $\mathbb{N}$ to $\mathbb{N}$, i.e. a not necessarily total surjection $e : \mathbb{N} \to \{f : \mathbb{N} \to \mathbb{N} \mid f \text{ is partial computable}\}$. The halting function w.r.t. e is defined as*

$$\begin{aligned} &\mathrm{halt}_e : \mathbb{N} \times \mathbb{N} \to 2 \\ &\mathrm{halt}_e(m, n) := \begin{cases} 1 & \textit{if } m \in \mathrm{dom}(e) \textit{ and } n \in \mathrm{dom}(e(m)) \\ 0 & \textit{else} \end{cases} \end{aligned} \tag{27}$$

This function is not total computable.

Proof. Assume halt_e is total computable, then according to lemma 5 halt_e is an arrow in **Asm**, $\mathrm{halt}_e : \mathbb{N}' \times \mathbb{N}' \to 2'$. This arrow is weakly point surjective, i.e. for any total computable function $g : \mathbb{N} \to 2$, we find a $m_g \in \mathbb{N}$ such that $\mathrm{halt}_e(m_g, -) = g$. Given $g : \mathbb{N} \to 2$, m_g can be constructed as follows: first define $\alpha : 2 \to 2$ by $\mathrm{dom}(\alpha) := \{1\}$, $\alpha(1) = 1$. α clearly is partial computable and hence so is $\alpha \circ g$. Now consider any $m_g \in e^{-1}(\alpha \circ g) \in \mathbb{N}$. As e is surjective such a m_g exists. Take any $n \in \mathbb{N}$, then

$$\begin{aligned} \mathrm{halt}_e(m_g, n) = 1 &\Leftrightarrow n \in \mathrm{dom}(\alpha \circ g) \\ &\Leftrightarrow g(n) \in \mathrm{dom}(\alpha) \\ &\Leftrightarrow g(n) = 1 \end{aligned} \tag{28}$$

so indeed $\mathrm{halt}_e(m_g, -) = g$. Therefore, Lawvere's theorem applies to the given situation, implying that every morphism $2' \to 2'$, i.e. every total computable function $2 \to 2$ has a fixed point. But this is clearly not true for $\mathrm{not} : 2 \to 2$, so the assumption that halt_e is total computable must have been wrong. $\square$

In [7] there is also a proof of Halting by Lawvere's fixed point theorem, that however works by considering a slightly different category.

Remark. *Note that by essentially the same arguments, there cannot exist a weakly point surjective, total computable function $U : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$. For such a* *function U, weak point surjectivity states that for any total computable function $g : \mathbb{N} \to \mathbb{N}$ there exists a natural number m_g such that $g = U(m_g, -)$. Considering Turing machines as underlying model of computation, by the existence of universal Turing machines, such a U clearly exists. However, by Lawvere's theorem it has to be partial.*

Remark. *Note that the above proof, showing that for any Gödel numbering e of computable functions halt_e cannot be total computable, is independent of the particular model of computation, as long as*

1. *the chosen model of computation yields a well-defined category* **Asm***,*
2. *allows for the existence of the Gödel numbering e,*
3. *allows for the definition of α*
4. *and contains at least one fixed point free, total computatable function $2 \to 2$*

2.3 Fixed-point combinator

In λ-calculus a fixed-point combinator [10] is a term `fix` which yields a fixed point for any term n when applied to it, that is:

$$n(\texttt{fix}\, n) = \texttt{fix}\, n \tag{29}$$

Such a term can be used to define recursion in the following way: Suppose we want to define the function $\texttt{mul\_14} : n \mapsto 14n$ recursively, that is:

```
def mul_14(n: int) -> int:
    if n == 0:
        return 0
    else:
        return 14 + mul_14(n-1)

mul_14(3) = 14 + mul_14(2)
          = 14 + 14 + mul_14(1)
          = 14 + 14 + 14 + mul_14(0)
          = 42
```

This construction is however not possible in λ-calculus, as there is no way of referencing a function in its own definition. Using a fixed-point combinator, however, allows for the following construction:

```
def gen_mul_14(f: (int -> int)) -> (int -> int):
    def g(n: int) -> int:
        if n == 0:
            return 0
        else:
```

```
            return 14 + f(n-1)
    return g

mul_14 = fix(gen_mul_14)

mul_14(3) = fix(gen_mul_14)(3)
          = gen_mul_14(fix(gen_mul_14))(3)
          = 14 + fix(gen_mul_14)(2)
          = 14 + 14 + fix(gen_mul_14)(1)
          = 14 + 14 + 14 + fix(gen_mul_14)(0)
          = 42 + gen_mul_14(fix(gen_mul_14))(0)
          = 42 + 0
          = 42
```

The existence of a fixed point for any term n is given by Lawvere's theorem [11, 12].

The idea is to consider a cartesian closed category with a (non-trivial) reflexive object, that is, an object $\Lambda \neq 1$ together with two morphisms $app : \Lambda \to \Lambda^\Lambda$ and $lam : \Lambda^\Lambda \to \Lambda$ with $app \circ lam = \mathrm{id}$. Such categories can be used to describe λ-calculus[12–14]. For any equivalence class of lambda terms (where the equivalence is given by α-, β- and η-reduction) there is a morphism $\Lambda \to \Lambda$ in the category, such that all equations and transformations can be reproduced in the category. Note that app is a split epimorphism and hence point surjective. That is, for any $b : 1 \to \Lambda^\Lambda$ there is $a = lam \circ b$ such that $app \circ a = app \circ lam \circ b = b$. Applying Lawvere's theorem yields the fact that any morphism $\Lambda \to \Lambda$ has a fixed point. So in particular, any lambda term has a fixed point.

Informally, a concrete example of `fix` can be found by reproducing the proof of Lawvere's theorem.

It can be shown that $\epsilon_\Lambda(app) \circ \Delta_\Lambda = \lambda x.x(x)$ [9]. Then, any term $t : \Lambda \to \Lambda$ has a fixed point. Using the construction from the proof of theorem 1 we find an explicit fixed point $h(h)$, where $h = t \circ \epsilon_\Lambda(app) \circ \Delta_\Lambda$ and composition is defined as $f \circ g = \lambda x.f(g(x))$.

$$\Lambda \xrightarrow{\Delta_\Lambda} \Lambda \times \Lambda \xrightarrow{\epsilon_\Lambda(app)} \Lambda \xrightarrow{t} \Lambda \tag{30}$$

where explicitly:

$$\begin{aligned} h &= t \circ \epsilon_\Lambda(app) \circ \Delta_\Lambda \\ &= t \circ (\lambda x.x(x)) \\ &= \lambda y.t((\lambda x.x(x))(y)) \\ &= \lambda y.t(y(y)) \\ &= \lambda x.t(x(x)) \end{aligned} \tag{31}$$

then

$$h(h) = (\lambda x.t(x(x)))(h) = t(h(h)) \tag{32}$$

By λ abstraction we can then define $\texttt{fix} := \lambda t.h(h) = \lambda t.(\lambda x.t(x(x)))(\lambda x.t(x(x)))$. This instance of `fix` is called the Y-combinator.

2.4 Gödel's first incompleteness theorem

Gödel's incompleteness theorem states that for any formal system T in which enough elementary elementary arithmetic can be carried out, there is a sentence S in T such that neither S nor $\neg S$ can be proven in S. This theorem, as well as Tarski's undefineblility theorem, which states that there cannot be a definable map *truth* mapping Gödel numbers of formulas to their truth value, can be seen as consequences of Lawvere's theorem. The main idea is that both *truth* and the notion of provability, together with other assumptions, yield a point surjective morphism while consistency, i.e. the property that no sentence S in T should be equivalent to its own inverse $S \cong \neg S$, leads to a morphism without a fixed point.

Definition 6 ([5])**.** *The Lindenbaum category $\mathbf{C}$ of a (single sorted first-order) theory T is given by:*

1. *Objects $1, 2$ and A and finite products thereof.*
2. *Morphisms are equivalence classes of formulas or terms of the theory, where equivalence is given by logical equivalence or equality proven in the theory.*
3. *Composition of morphisms is defined as substitution.*

That is, morphisms $1 \to A$ are (equivalence classes of) constant terms in T, morphisms $A \to 2$ are formulas with one free variable and morphisms $1 \to 2$ are sentences of T. Composing $a : 1 \to A$ with a formula $f : A \to 2$ yields a sentence $fa : 1 \to 2$.

In particular, $\mathrm{Hom}_{\mathbf{C}}(1, 2)$ contains two morphisms $true, false$ corresponding to the classes of provable sentences and sentences for which the negation is provable, respectively.

Remark. *Models of the theory can be seen as functors $\mathbf{C} \to \mathbf{Set}$.*

A theory T is said to be consistent if there is no sentence ϕ such that both ϕ and $\neg\phi$ are provable in the theory. In our context this can be phrased as: if the theory is consistent, then there is a morphism $not : 2 \to 2$ (playing the role of $\neg$) such that $not(\phi) \neq \phi$ for all morphisms $\phi : 1 \to 2$.

For, if there would not be such a morphism, $\neg$ could not be defined or there would be sentences $\phi = \neg\phi$ rendering the theory inconsistent.

In such a category, we say that *satisfaction is definable* if there exists a morphism $sat : A \times A \to 2$ such that for every formula $f : A \to 2$ there is a constant $c : 1 \to A$ such that for every constant $a : 1 \to A$ the following diagram commutes:

$$\begin{tikzcd} 1 \arrow[r, "a"] \arrow[d, "\langle a,c \rangle"'] & A \arrow[d, "f"] \\ A \times A \arrow[r, "sat"'] & 2 \end{tikzcd} \tag{33}$$

In words, c is some kind of Gödel number for the formula f and the condition reads for all constant terms $a : 1 \to A$ $sat(a, c)$ (i.e. the formula indexed by c applied to a) is provably equivalent to the sentence $f(a)$.

Then,

Corollary 2. *If satisfaction is definable, then the theory is not consistent.*

Proof. As sat is a point-surjective morphism $A \times A \to 2$, by theorem 1 all morphisms $2 \to 2$ will have a fixed point. Thus, there cannot exist $not : 2 \to 2$ without a fixed point as defined above. □

This result is the base for the following two theorems. In words, it says that if we can find a formula sat with two free parameters which "decodes" constants to formulas, then the theory is not consistent.

Definition 7 ([15])**.** *A* Gödel numbering *is an injective map:*

$$\ulcorner - \urcorner : \mathrm{Hom}_{\mathbf{C}}(A^n, 2) \to \mathrm{Hom}_{\mathbf{C}}(1, A) \tag{34}$$

defined for all $n \in \mathbb{N}$*.*

In particular, it is defined for $\mathrm{Hom}_{\mathbf{C}}(1, 2) = \mathrm{Hom}_{\mathbf{C}}(A^0, 2)$. It encodes any formula as a constant term.

Remark. *This definition is more general than what is usually done. Namely, formulas can not only be encoded as natural numbers but as any constant term the theory provides.*

We say that *substitution is definable* if there exists a morphism $subst : A \times A \to A$ such that:

$$\forall (f : A \to 2), (a : 1 \to A) : \ subst(a, \ulcorner f \urcorner) = \ulcorner f(a) \urcorner \tag{35}$$

This morphism substitutes a in the formula encoded by $\ulcorner f \urcorner$ and returns the Gödel number of the resulting sentence $f(a)$.

We say that *truth is definable* if there exists a formula $truth : A \to 2$ such that its image under $\mathrm{Hom}_{\mathbf{C}}(1, -)$:

$$truth \circ - : \mathrm{Hom}_{\mathbf{C}}(1, A) \to \mathrm{Hom}_{\mathbf{C}}(1, 2) \tag{36}$$

is a retract of $\ulcorner - \urcorner$. That is:

$$(truth \circ -) \circ \ulcorner - \urcorner = Id_{\mathrm{Hom}_{\mathbf{C}}(1,2)}. \tag{37}$$

More concretely, for any sentence $\phi \in \mathrm{Hom}_{\mathbf{C}}(1, 2)$,

$$((truth \circ -) \circ \ulcorner - \urcorner)(\phi) = truth \circ \ulcorner \phi \urcorner = \phi \tag{38}$$

In words, $truth$ maps the Gödel number of any sentence ϕ to the sentence ϕ itself (or to some equivalent sentences).

Using those definition we can phrase Tarski's undefinability theorem:

Theorem 4. *In a consistent theory where substitution is definable (relative to $\ulcorner - \urcorner$), then truth is not definable (relative to $\ulcorner - \urcorner$).*

Hence, if substitution is definable, there is no way (in the theory itself) to map any Gödel number to its corresponding sentence or its truth value, respectively.

Proof. If both substitution and truth are definable, then the following diagram commutes for all $a :\to A$:

$$\begin{tikzcd}
1 \arrow[r, "a"] \arrow[d, "\langle a, \ulcorner \phi \urcorner \rangle"'] \arrow[dr, "\ulcorner \phi(a) \urcorner"] & A \arrow[dr, "\phi"] & \\
A \times A \arrow[r, "subst"'] & A \arrow[r, "truth"'] & 2
\end{tikzcd} \tag{39}$$

which shows that $truth \circ subst$ is a definition of satisfaction (eq. (33)) and yields a contradiction by corollary 2.

□

To finally arrive at Gödel's incompleteness theorem, two ingredients are still missing: provability and completeness.

We say that *provability is representable* in a theory if and only id there is a unary formula $pr : A \to 2$ such that for any sentence $\phi : 1 \to 2$

$$Pr \circ \ulcorner \phi \urcorner = true \Leftrightarrow \phi = true \tag{40}$$

That is, $\ulcorner \phi \urcorner$ is provable if and only if ϕ is true.

Furthermore, a theory is called *complete* if $\mathrm{Hom}_{\mathbf{C}}(1, 2) = \{true, false\}$, that is, if every sentence in T is either provably true or provably false.

Then, we get a version of Gödel's incompleteness theorem:

Theorem 5. *A theory where substitution is definable and provability is representable (relative to $\ulcorner - \urcorner$) is not complete if it is consistent.*

Proof. Assume the theory is complete (i.e that $\mathrm{Hom}_{\mathbf{C}}(1, 2) = \{true, false\}$) and consistent. Consistency implies $false \neq true$, then, by the definition of Pr for any sentence $\phi \in \mathrm{Hom}_{\mathbf{C}}(1, 2)$:

$$\phi = true \Rightarrow Pr \circ \ulcorner \phi \urcorner = true \tag{41}$$

$$\phi \neq true \Rightarrow Pr \circ \ulcorner \phi \urcorner \neq true \tag{42}$$

Then, by using completeness, the second line reads:

$$\phi = false \Rightarrow Pr \circ \ulcorner \phi \urcorner = false \tag{43}$$

Hence, for all $\phi \in \mathrm{Hom}_{\mathbf{C}}(1, 2)$:

$$Pr \circ \ulcorner \phi \urcorner = \phi. \tag{44}$$

Which shows that Pr is a truth definition (eq. (38)), yielding a contradiction by theorem 4. □

References

[1] A. Bauer, *On a proof of Cantor's theorem*, `http://math.andrej.com/2007/04/08/on-a-proof-of-cantors-theorem/`, Apr. 2007.

[2] S. MacLane, *Categories for the working mathematician*, Graduate Texts in Mathematics, Vol. 5 (Springer-Verlag, New York, 1971), pp. ix+262.

[3] nLab authors, *Cartesian closed category*, `http://nlab-pages.s3.us-east-2.amazonaws.com/nlab/show/cartesian+closed+category`.

[4] nLab authors, *Exponential object*, `http://nlab-pages.s3.us-east-2.amazonaws.com/nlab/show/exponential+object`.

[5] F. W. Lawvere, "Diagonal arguments and cartesian closed categories", in Category theory, homology theory and their applications II (1969), pp. 134–145.

[6] nLab authors, *Lawvere's fixed point theorem*, `http://nlab-pages.s3.us-east-2.amazonaws.com/nlab/show/Lawvere's+fixed+point+theorem`.

[7] N. S. Yanofsky, *A universal approach to self-referential paradoxes, incompleteness and fixed points*, 2003, arXiv:`math/0305282 [math.LO]`.

[8] W. P. Stekelenburg, *Categories of assemblies for realizability*, 2013, arXiv:`1307.0663 [math.LO]`.

[9] G. M. Dan Frumin, *Diagonal arguments and Lawvere's theorem*, `https://cs.ru.nl/~dfrumin/pdf/diagonal_argument.pdf`.

[10] nLab authors, *Fixed-point combinator*, `http://nlab-pages.s3.us-east-2.amazonaws.com/nlab/show/fixed-point+combinator`.

[11] *Mathematics Stack Exchange: Y combinator as an application of Lawvere's fixed point theorem*, `https://math.stackexchange.com/questions/2823016/y-combinator-as-an-application-of-lawveres-fixed-point-theorem`.

[12] nLab authors, *Domain theory*, `http://nlab-pages.s3.us-east-2.amazonaws.com/nlab/show/domain+theory`.

[13] D. Scott, "Relating theories of the λ-calculus", in To H.B. Curry: essays on combinatory logic, lambda-calculus and formalism (1980), pp. 403–450.

[14] H. Barendregt, *The lambda calculus: its syntax and semantics* (North-Holland Publishing Company, 1981).

[15] R. Cardona, *Lawvere's fixed point theorem (talk)*, `https://www.albany.edu/~rc782885/talk/2019-10-october-atia/`, 2019.