1 Universal Algebra and Monads

Concerto Al Andalus

Marcel Khalifé

In this chapter, we cover most of the content on universal algebra and monads that we will need in the rest of the thesis. This material has appeared many times in the literature[0], but for completeness (and to be honest my own satisfaction) we take our time with it. In **??**, we will follow the outline of the current chapter to generalize the definitions and results to sets equipped with a notion of distance. Thus, many choices in our notations and presentation are motivated by the needs of **??**.

[0] [Wec12] and [Bau19] are two of my favorite references on universal algebra, and both [Rie17, Chapter 5] and [BW05, Chapter 3] are great references for monads (the latter calls them *triples*).

Outline: In §1.1, we define algebras, terms, and equations over a signature of finitary operation symbols. In §1.2, we explain how to construct the free algebra for a given signature and set of equations. In §1.3, we give the rules for equational logic to derive equations from other equations, and we show it sound and complete. In §1.4, we give define monads and algebraic presentations for monads. We give examples all throughout, some small ones to build intuition, and some bigger ones that will be needed later.

1.1 Algebras and Equations

Definition 1 (Signature)**.** A **signature** is a set Σ whose elements, called **operation symbols**, each come with an **arity** $n \in \mathbb{N}$. We write $\mathsf{op} : n \in \Sigma$ for a symbol op with arity n in Σ. With some abuse of notation, we also denote by Σ the functor $\Sigma : \mathbf{Set} \to \mathbf{Set}$ with the following action:[1]

$$\Sigma(A) := \coprod_{\mathsf{op}:n\in\Sigma} A^n \text{ on sets} \quad \text{and} \quad \Sigma(f) := \coprod_{\mathsf{op}:n\in\Sigma} f^n \text{ on functions.}$$

[1] The set $\Sigma(A)$ can be identified with the set containing $\mathsf{op}(a_1, \ldots, a_n)$ for all $\mathsf{op} : n \in \Sigma$ and $a_1, \ldots, a_n \in A$. Then, the function $\Sigma(f)$ sends $\mathsf{op}(a_1, \ldots, a_n)$ to $\mathsf{op}(f(a_1), \ldots, f(a_n))$.

Definition 2 (Σ-algebra)**.** A Σ**-algebra** (or just algebra) is a set A equipped with functions $[\![\mathsf{op}]\!]_A : A^n \to A$ for every $\mathsf{op} : n \in \Sigma$ called the **interpretation** of the symbol. We call A the **carrier** or **underlying** set, and when referring to an algebra, we will switch between using a single symbol $\mathbb{A}$[2] or the pair $(A, [\![-]\!]_A)$, where $[\![-]\!]_A : \Sigma(A) \to A$ is the function sending $\mathsf{op}(a_1, \ldots, a_n)$ to $[\![\mathsf{op}]\!]_A(a_1, \ldots, a_n)$ (it compactly describes the interpretations of all symbols).

[2] We will try to match the symbol for the algebra and the one for the underlying set only modifying the former with `mathbb`.

A **homomorphism** from $\mathbb{A}$ to $\mathbb{B}$ is a function $h : A \to B$ between the underlying sets of $\mathbb{A}$ and $\mathbb{B}$ that preserves the interpretation of all operation symbols in Σ, namely, for all $\mathsf{op} : n \in \Sigma$ and $a_1, \ldots, a_n \in A$,[3]

$$h(\llbracket \mathsf{op} \rrbracket_A(a_1, \ldots, a_n)) = \llbracket \mathsf{op} \rrbracket_B(h(a_1), \ldots, h(a_n)). \tag{1}$$

The identity maps $\mathrm{id}_A : A \to A$ and the composition of two homomorphisms are always homomorphisms, therefore we have a category whose objects are Σ-algebras and morphisms are Σ-algebra homomorphisms. We denote it by $\mathbf{Alg}(\Sigma)$.

This category is concrete over **Set** with the forgetful functor $U : \mathbf{Alg}(\Sigma) \to \mathbf{Set}$ which sends an algebra $\mathbb{A}$ to its carrier and a homomorphism to the underlying function between carriers.

Remark 3. In the sequel, we will rarely distinguish between the homomorphism $h : \mathbb{A} \to \mathbb{B}$ and the underlying function $h : A \to B$. Although, we may write Uh for the latter, when disambiguation is necessary.

Examples 4. 1. Let $\Sigma = \{\mathsf{p} : 0\}$ be the signature containing a single operation symbol p with arity 0. A Σ-algebra is a set A equipped with an interpretation of p as a function $\llbracket \mathsf{p} \rrbracket_A : A^0 \to A$. Since A^0 is the singleton $\mathbf{1}$, $\llbracket \mathsf{p} \rrbracket_A$ is just a choice of element in A,[4] so the objects of $\mathbf{Alg}(\Sigma)$ are pointed sets (sets with a distinguished element). Moreover, instantiating (1) for the symbol p, we find that a homomorphism from $\mathbb{A}$ to $\mathbb{B}$ is a function $h : A \to B$ sending the distinguished point of A to the distinguished point of B. We conclude that $\mathbf{Alg}(\Sigma)$ is the category $\mathbf{Set}_*$ of pointed sets and functions preserving the points.

2. Let $\Sigma = \{\mathsf{f} : 1\}$ be the signature containing a single unary operation symbol f. A Σ-algebra is a set A equipped with an interpretation of f as a function $\llbracket \mathsf{f} \rrbracket_A : A \to A$.

   For example, we have the Σ-algebra whose carrier is the set of integers $\mathbb{Z}$ and where f is interpreted as "adding 1", i.e. $\llbracket \mathsf{f} \rrbracket_{\mathbb{Z}}(k) = k + 1$. We also have the integers mod 2 $\mathbb{Z}_2$ where $\llbracket \mathsf{f} \rrbracket_{\mathbb{Z}_2}(k) = k + 1 (\mathrm{mod}\ 2)$.

   The fact that a function $h : A \to B$ satisfies (1) for the symbol f is equivalent to the following commutative square.

$$\begin{array}{ccc} A & \xrightarrow{h} & B \\ {\scriptstyle \llbracket \mathsf{f} \rrbracket_A}\downarrow & & \downarrow{\scriptstyle \llbracket \mathsf{f} \rrbracket_B} \\ A & \xrightarrow[h]{} & B \end{array}$$

   We conclude that $\mathbf{Alg}(\Sigma)$ is the category whose objects are endofunctions and whose morphisms are commutative squares as above.[5] There is a homomorphism $\mathsf{is\_odd}$ from $\mathbb{Z}$ to $\mathbb{Z}_2$ that sends k to $k (\mathrm{mod}\ 2)$, that is, to 0 when it is even and to 1 when it is odd.

3. Let $\Sigma = \{+ : 2\}$ be the signature containing a single binary operation symbol. A Σ-algebra is a set A equipped with an interpretation $\llbracket + \rrbracket_A : A \times A \to A$. Such

[3] Equivalently, h makes the following square commute:

$$\begin{array}{ccc} \Sigma(A) & \xrightarrow{\Sigma(f)} & \Sigma(B) \\ {\scriptstyle \llbracket - \rrbracket_A}\downarrow & & \downarrow{\scriptstyle \llbracket - \rrbracket_B} \\ A & \xrightarrow[f]{} & B \end{array} \tag{0}$$

This amounts to an equivalent and more concise definition of $\mathbf{Alg}(\Sigma)$: it is the category of algebras for the signature functor $\Sigma : \mathbf{Set} \to \mathbf{Set}$ [Awo10, Definition 10.8].

[4] For this reason, we often call 0-ary symbols **constants**.

[5] For more categorical thinkers, we can also identify $\mathbf{Alg}(\Sigma)$ with the functor category $[\mathbf{B}\mathbb{N}, \mathbf{Set}]$ from the delooping of the (additive) monoid $\mathbb{N}$ to the category of sets. Briefly, it is because a functor $\mathbf{B}\mathbb{N} \to \mathbf{Set}$ is completely determined by where it sends $1 \in \mathbb{N}$.

a structure is often called a magma, and it is part of many more well-known algebraic structures like groups, rings, monoids, etc. While every group has an underlying Σ-algebra, not every Σ-algebra underlies a group since $[\![+]\!]_A$ is not required to be associative for example. The following definitions will allow us to talk about certain classes of Σ-algebras with some properties like associativity.

Definition 5 (Term)**.** Let Σ be a signature and A be a set. We denote with $T_\Sigma A$ the set of **Σ-terms** built syntactically from A and the operation symbols in Σ, i.e., the set inductively defined by

$$\frac{a \in A}{a \in T_\Sigma A} \quad \text{and} \quad \frac{\mathsf{op} : n \in \Sigma \qquad t_1, \dots, t_n \in T_\Sigma A}{\mathsf{op}(t_1, \dots, t_n) \in T_\Sigma A}.$$

We identify elements $a \in A$ with the corresponding terms $a \in T_\Sigma A$, and we also identify (as outlined in Footnote 1) elements of $\Sigma(A)$ with terms in $T_\Sigma A$ containing exactly one occurrence of an operation symbol.[6]

The assignment $A \mapsto T_\Sigma A$ can be turned into a functor $T_\Sigma : \mathbf{Set} \to \mathbf{Set}$ by inductively defining, for any function $f : A \to B$, the function $T_\Sigma f : T_\Sigma A \to T_\Sigma B$ as follows:[7]

$$\frac{a \in A}{T_\Sigma f(a) = f(a)} \quad \text{and} \quad \frac{\mathsf{op} : n \in \Sigma \qquad t_1, \dots, t_n \in T_\Sigma A}{T_\Sigma f(\mathsf{op}(t_1, \dots, t_n)) = \mathsf{op}(T_\Sigma f(t_1), \dots, T_\Sigma f(t_n))}. \tag{2}$$

Proposition 6. *We defined a functor $T_\Sigma : \mathbf{Set} \to \mathbf{Set}$, i.e. $T_\Sigma \mathrm{id}_A = \mathrm{id}_{T_\Sigma A}$ and $T_\Sigma(g \circ f) = T_\Sigma g \circ T_\Sigma f$.*

Proof. We proceed by induction for both equations. For any $a \in A$, we have $T_\Sigma \mathrm{id}_A(a) = \mathrm{id}_A(a) = a$ and

$$T_\Sigma(g \circ f)(a) = (g \circ f)(a) = T_\Sigma g(T_\Sigma f(a)).$$

For any $t = \mathsf{op}(t_1, \dots, t_n)$, we have

$$T_\Sigma \mathrm{id}_A(\mathsf{op}(t_1, \dots, t_n)) \stackrel{(2)}{=} \mathsf{op}(T_\Sigma \mathrm{id}_A(t_1), \dots, T_\Sigma \mathrm{id}_A(t_n)) \stackrel{\text{I.H.}}{=} \mathsf{op}(t_1, \dots, t_n),$$

and

$$\begin{aligned} T_\Sigma(g \circ f)(t) &= T_\Sigma(g \circ f)(\mathsf{op}(t_1, \dots, t_n)) \\ &= \mathsf{op}(T_\Sigma(g \circ f)(t_1), \dots, T_\Sigma(g \circ f)(t_n)) && \text{by (2)} \\ &= \mathsf{op}(T_\Sigma g(T_\Sigma f(t_1)), \dots, T_\Sigma g(T_\Sigma f(t_n))) && \text{I.H.} \\ &= T_\Sigma g(\mathsf{op}(T_\Sigma f(t_1), \dots, T_\Sigma f(t_n))) && \text{by (2)} \\ &= T_\Sigma g T_\Sigma f(\mathsf{op}(t_1, \dots, t_n)). && \text{by (2)} \end{aligned}$$ □

Examples 7. 1. With $\Sigma = \{\mathsf{p} : 0\}$, a Σ-term over A is either an element of A or p. The functor T_Σ is then naturally isomorphic to the functor sending A to $A + \mathbf{1}$.

2. With $\Sigma = \{\mathsf{f} : 1\}$, a Σ-term over A is either an element of A or a term $\mathsf{f}(\mathsf{f}(\cdots \mathsf{f}(a)))$ for some a and a finite number of iterations of f. The functor T_Σ is then naturally isomorphic to the functor sending A to $\mathbb{N} \times A$.

[6] Note that any constant $\mathsf{p} : 0 \in \Sigma$ belongs to all $T_\Sigma A$ by the second rule defining $T_\Sigma X$.

[7] Note that $T_\Sigma f$ acts as identity on constants.

3. With $\Sigma = \{+:2\}$, a Σ-term is either an element of A or any expression formed by "adding" elements of A together like $a+b$, $a+(b+c)$, $((a+a)+c)+(b+c)$ and so on when $a, b, c \in A$.[8]

[8] We write $+$ infix as is very common. The parentheses are formal symbols to help delimit which $+$ is taken first. They are necessary because the interpretation of $+$ is not necessarily associative so $a+(b+c)$ and $(a+b)+c$ can be interpreted differently in some Σ-algebras.

As we said above, any element in A is a term in $T_\Sigma A$, we will denote this embedding with $\eta_A^\Sigma : A \to T_\Sigma A$, in particular, we will write $\eta_A^\Sigma(a)$ to emphasize that we are dealing with the term a and not the element of A. For instance, the base case of the definition of $T_\Sigma f$ in (2) becomes

$$\frac{a \in A}{T_\Sigma f(\eta_A^\Sigma(a)) = \eta_B^\Sigma(f(a))}.$$

This is exactly what it means for the family of maps $\eta_A^\Sigma : A \to T_\Sigma A$ to be natural in A,[9] in other words that $\eta^\Sigma : \mathrm{id}_{\mathbf{Set}} \Rightarrow T_\Sigma$ is a natural transformation. We can mention now that it will be part of some additional structure on the functor T_Σ (a monad).

[9] As a commutative square:

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ {\scriptstyle \eta_A^\Sigma}\downarrow & & \downarrow{\scriptstyle \eta_B^\Sigma} \\ T_\Sigma A & \xrightarrow[T_\Sigma f]{} & T_\Sigma B \end{array} \qquad (3)$$

For an arbitrary signature Σ, we can think of $T_\Sigma A$ as the set of rooted trees whose leaves are labelled with elements of A and whose nodes with n children are labelled with n-ary operation symbols in Σ. This makes the action of a function $T_\Sigma f$ fairly straightforward: it applies f to the labels of all the leaves.

This point of view is particularly helpful when describing the **flattening** of terms: there is a natural way to see a Σ-term over Σ-terms over A as a Σ-term over A. This is carried out by the map $\mu_A^\Sigma : T_\Sigma T_\Sigma A \to T_\Sigma A$ which takes a tree T whose leaves are labelled with trees $T_1, \ldots, T_n$ to the tree T where instead of the leaf labelled T_i, there is the root of T_i with all its children and their children and so on (we "glue" the tree T_i at the leaf labelled T_i). Figure 1.1 shows an example for $\Sigma = \{+:2\}$. More formally, μ_A^Σ is defined inductively by:

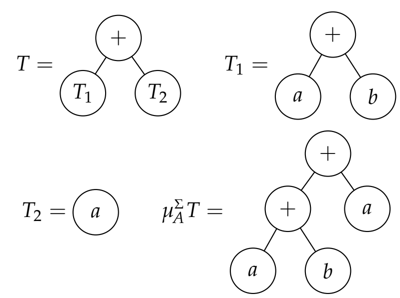


Figure 1.1: Flattening of a term

$$\mu_A^\Sigma(\eta_{T_\Sigma A}^\Sigma(t)) = t \text{ and } \mu_A^\Sigma(\mathsf{op}(t_1, \ldots, t_n)) = \mathsf{op}(\mu_A^\Sigma(t_1), \ldots, \mu_A^\Sigma(t_n)). \qquad (4)$$

The use of the word "natural" above is not benign, μ^Σ is actually a natural transformation.

Proposition 8. *The family of maps $\mu_A^\Sigma : T_\Sigma T_\Sigma A \to T_\Sigma A$ is natural in A.*

Proof. We need to prove that for any function $f : A \to B$, $T_\Sigma f \circ \mu_A^\Sigma = \mu_B^\Sigma \circ T_\Sigma T_\Sigma f$.[10] It makes sense intuitively, we should get the same result when we apply f to all the leaves before or after flattening. Formally, we use induction.

[10] As a commutative square:

$$\begin{array}{ccc} T_\Sigma T_\Sigma A & \xrightarrow{T_\Sigma T_\Sigma f} & T_\Sigma T_\Sigma B \\ {\scriptstyle \mu_A^\Sigma}\downarrow & & \downarrow{\scriptstyle \mu_B^\Sigma} \\ T_\Sigma A & \xrightarrow[T_\Sigma f]{} & T_\Sigma B \end{array} \qquad (5)$$

For the base case (i.e. terms in the image of $\eta_{T_\Sigma A}^\Sigma$), we have

$$\begin{aligned} \mu_B^\Sigma(T_\Sigma T_\Sigma f(\eta_{T_\Sigma A}^\Sigma(t))) &= \mu_B^\Sigma(\eta_{T_\Sigma B}^\Sigma(T_\Sigma f(t))) && \text{by (3)} \\ &= T_\Sigma f(t) && \text{by (4)} \\ &= T_\Sigma f(\mu_A^\Sigma(\eta_{T_\Sigma A}^\Sigma(t))). && \text{by (4)} \end{aligned}$$

For the inductive step, we have

$$\begin{aligned} \mu_B^\Sigma(T_\Sigma T_\Sigma f(\mathsf{op}(t_1, \ldots, t_n))) &= \mu_B^\Sigma(\mathsf{op}(T_\Sigma T_\Sigma f(t_1), \ldots, T_\Sigma T_\Sigma f(t_n))) && \text{by (2)} \\ &= \mathsf{op}(\mu_B^\Sigma(T_\Sigma T_\Sigma f(t_1)), \ldots, \mu_B^\Sigma(T_\Sigma T_\Sigma f(t_n))) && \text{by (4)} \end{aligned}$$

$$
\begin{aligned}
&= \mathsf{op}(T_\Sigma f(\mu^\Sigma_A(t_1)), \dots, T_\Sigma f(\mu^\Sigma_A(t_n))) && \text{I.H.}\\
&= T_\Sigma f(\mathsf{op}(\mu^\Sigma_A(t_1), \dots, \mu^\Sigma_A(t_n))) && \text{by (2)}\\
&= T_\Sigma f(\mu^\Sigma_A(\mathsf{op}(t_1, \dots, t_n))) && \text{by (4)} \quad \square
\end{aligned}
$$

By definition, we have that $\mu^\Sigma \cdot \eta^\Sigma T_\Sigma$ is the identity transformation $\mathbb{1}_{T_\Sigma} : T_\Sigma \Rightarrow T_\Sigma$.[11] In words, we say that seeing a term trivially as a term over terms then flattening it yields back the original term. Another similar property is that we see all the variables in a term trivially as terms and flatten the resulting term over terms, the result is the original term. Formally:

[11] We write $\cdot$ to denote the vertical composition of natural transformations and juxtaposition (e.g. $F\phi$ or ϕF to denote the action of functors on natural transformations), namely, the component of $\mu^\Sigma \cdot \eta^\Sigma T_\Sigma$ at A is $\mu^\Sigma_A \circ \eta^\Sigma_{T_\Sigma A}$ which is $\mathrm{id}_{T_\Sigma A}$ by (4).

Lemma 9. *For any set A, $\mu^\Sigma_A \circ T_\Sigma \eta^\Sigma_A = \mathrm{id}_{T_\Sigma A}$, hence $\mu^\Sigma \cdot T_\Sigma \eta^\Sigma = \mathbb{1}_{T_\Sigma}$.*

Proof. We proceed by induction. For the base case, we have

$$\mu^\Sigma_A(T_\Sigma \eta^\Sigma_A(\eta^\Sigma_A(a))) \overset{(3)}{=} \mu^\Sigma_A(\eta^\Sigma_{T_\Sigma A}(\eta^\Sigma_A(a))) \overset{(4)}{=} \eta^\Sigma_A(a).$$

For the inductive step, if $t = \mathsf{op}(t_1, \dots, t_n)$, we have

$$
\begin{aligned}
\mu^\Sigma_A(T_\Sigma \eta^\Sigma_A(t)) &= \mu^\Sigma_A(T_\Sigma \eta^\Sigma_A(\mathsf{op}(t_1, \dots, t_n)))\\
&= \mu^\Sigma_A(\mathsf{op}(T_\Sigma \eta^\Sigma_A(t_1), \dots, T_\Sigma \eta^\Sigma_A(t_n))) && \text{by (2)}\\
&= \mathsf{op}(\mu^\Sigma_A(T_\Sigma \eta^\Sigma_A(t_1)), \dots, \mu^\Sigma_A(T_\Sigma \eta^\Sigma_A(t_n))) && \text{by (4)}\\
&= \mathsf{op}(t_1, \dots, t_n) = t && \text{I.H.} \quad \square
\end{aligned}
$$

Trees also make the depth of a term a visual concept. A term $t \in T_\Sigma A$ is said to be of **depth** $d \in \mathbb{N}$ if the tree representing it has depth d.[12] We give an inductive definition:

[12] i.e. the longest path from the root to a leaf has d edges. In Figure 1.1, the depth of T and T_1 is 1, the depth of T_2 is 0 and the depth of $\mu^\Sigma_A T$ is 2.

$$\mathrm{depth}(a) = 0 \text{ and } \mathrm{depth}(\mathsf{op}(t_1, \dots, t_n)) = 1 + \max\{\mathrm{depth}(t_1), \dots, \mathrm{depth}(t_n)\}.$$

A term of depth 0 is a term in the image of η^Σ_A. A term of depth 1 is an element of $\Sigma(A)$ seen as a term (recall Footnote 1).

In any Σ-algebra $\mathbb{A}$, the interpretations of operation symbols give us an element of A for each element of $\Sigma(A)$. Using, the inductive definition of $T_\Sigma a$, we can extend these interpretations to all terms in $T_\Sigma A$: abusing notation, we define the function $[\![-]\!]_A : T_\Sigma A \to A$ by[13]

[13] For categorical thinkers, $T_\Sigma A$ is essentially defined to be the initial algebra for the endofunctor $\Sigma + A$: **Set** $\to$ **Set** sending X to $\Sigma(X) + A$. Any Σ-algebra $(A, [\![-]\!]_A)$ defines another algebra for that functor $[[\![-]\!]_A, \mathrm{id}_A] : \Sigma(A) + A \to A$. Then, the extension of $[\![-]\!]_A$ to terms is the unique algebra morphism drawn below.

$$
\begin{array}{ccc}
\Sigma(T_\Sigma A) + A & \dashrightarrow & \Sigma(A) + A\\
\downarrow & & \downarrow {\scriptstyle [[\![-]\!]_A, \mathrm{id}_A]}\\
T_\Sigma A & \dashrightarrow & A
\end{array}
$$

$$\frac{a \in A}{[\![a]\!]_A = a} \quad \text{and} \quad \frac{\mathsf{op} : n \in \Sigma \qquad t_1, \dots, t_n \in T_\Sigma A}{[\![\mathsf{op}(t_1, \dots, t_n)]\!]_A = [\![\mathsf{op}]\!]_A([\![t_1]\!]_A, \dots, [\![t_n]\!]_A)}. \tag{6}$$

This allows to further extend the interpretation $[\![-]\!]_A$ to all terms $T_\Sigma X$ over some set of variables X, provided we have an assignment of variables $\iota : X \to A$, by precomposing with $T_\Sigma \iota$. We denote this interpretation with $[\![-]\!]^\iota_A$:

$$[\![-]\!]^\iota_A = T_\Sigma X \xrightarrow{T_\Sigma \iota} T_\Sigma A \xrightarrow{[\![-]\!]_A} A. \tag{7}$$

Example 10. In the signature $\Sigma = \{\mathsf{f} : 1\}$ and over the variables $X = \{x\}$, we have (amongst others) the terms $t = \mathsf{ff}x$ and $s = \mathsf{fff}x$. If we compute the interpretation of t and s in $\mathbb{Z}$ and $\mathbb{Z}_2$,[14] we obtain

[14] Recall their Σ-algebra structure given in Examples 4.

$$[\![t]\!]^\iota_{\mathbb{Z}} = \iota(x) + 2 \quad [\![s]\!]^\iota_{\mathbb{Z}} = \iota(x) + 3 \quad [\![t]\!]^\iota_{\mathbb{Z}_2} = \iota(x) \quad [\![s]\!]^\iota_{\mathbb{Z}_2} = \iota(x) + 1 (\text{mod } 2),$$

for any assignment $\iota : X \to \mathbb{Z}$ (resp. $\iota : X \to \mathbb{Z}_2$).

By definition, a homomorphism preserves the interpretation of operation symbols. We can prove by induction that it also preserves the interpretation of arbitrary terms. Namely, if $h : \mathbb{A} \to \mathbb{B}$ is a homomorphism, then the following square commutes.[15]

$$\begin{array}{ccc} T_\Sigma A & \xrightarrow{T_\Sigma h} & T_\Sigma B \\ {\scriptstyle [\![-]\!]_A}\downarrow & & \downarrow{\scriptstyle [\![-]\!]_B} \\ A & \xrightarrow[h]{} & B \end{array} \tag{8}$$

[15] *Quick proof.* If $t = a \in A$, then both paths send it to $h(a)$. If $t = \mathsf{op}(t_1, \dots, t_n)$, then

$$\begin{aligned} h([\![t]\!]_A) &= h([\![\mathsf{op}]\!]_A([\![t_1]\!]_A, \dots, [\![t_n]\!]_A)) \\ &= [\![\mathsf{op}]\!]_B(h([\![t_1]\!]_A), \dots, h([\![t_n]\!]_A)) \\ &= [\![\mathsf{op}]\!]_B([\![T_\Sigma h(t_1)]\!]_B, \dots, [\![T_\Sigma h(t_n)]\!]_B) \\ &= [\![\mathsf{op}(T_\Sigma h(t_1), \dots, T_\Sigma h(t_n))]\!]_B \\ &= [\![T_\Sigma h(t)]\!]. \end{aligned}$$

The converse is (almost trivially) true, if (8) commutes, then we can quickly see (0) commutes by embedding $\Sigma(A)$ into $T_\Sigma A$ and $\Sigma(B)$ into $T_\Sigma B$. It follows readily that for all homomorphisms $h : \mathbb{A} \to \mathbb{B}$ and all assignments $\iota : X \to A$,

$$h \circ [\![-]\!]^\iota_A = [\![-]\!]^{h \circ \iota}_B. \tag{9}$$

Definition 11 (Equation)**.** An **equation** over a signature Σ is a triple comprising a set X of variables called the **context**, and a pair of terms $s, t \in T_\Sigma X$. We write these as $X \vdash s = t$.

A Σ-algebra $\mathbb{A}$ **satisfies** an equation $X \vdash s = t$ if for any assignment of variables $\iota : X \to A$, $[\![s]\!]^\iota_A = [\![t]\!]^\iota_A$. We use ϕ and ψ to refer to equations, and we write $\mathbb{A} \vDash \phi$ when $\mathbb{A}$ satisfies ϕ. We also write $\mathbb{A} \vDash^\iota \phi$ when the equality $[\![s]\!]^\iota_A = [\![t]\!]^\iota_A$ holds for a particular assignment $\iota : X \to A$ and not necessarily for all assignments.

Example 12 (Associativity)**.** Let $\Sigma = \{+:2\}$, $X = \{x, y, z\}$, $s = x + (y + z)$ and $t = (x + y) + z$. The equation $\phi = X \vdash s = t$[16] asserts that the interpretation of $+$ is associative. Indeed, suppose $\mathbb{A} \vDash \phi$, we need to show that for any $a, b, c \in A$,

[16] Alternatively, we may write ϕ omitting brackets:

$$x, y, z \vdash x + (y + z) = (x + y) + z.$$

$$[\![+]\!]_A(a, [\![+]\!]_A(b, c)) = [\![+]\!]_A([\![+]\!]_A(a, b), c). \tag{10}$$

Observe that the L.H.S. is the interpretation of s under the assignment $\iota : X \to A$ sending x to a, y to b and z to c, that is, we have $[\![+]\!]_A(a, [\![+]\!]_A(b, c)) = [\![s]\!]^\iota_A$. Under the same assignment, the interpretation of t is the R.H.S. By hypothesis, $[\![s]\!]^\iota_A = [\![t]\!]^\iota_A$, so we conclude (10) holds.

Examples 13. Without going into this much details, there are many other simple examples of equations.

- $x, y \vdash x + y = y + x$ states that the binary operation $+$ is commutative.
- $x \vdash x + x = x$ states that the binary operation $+$ is idempotent.
- $x \vdash \mathsf{f}x = \mathsf{ff}x$ states that the unary operation f is idempotent.
- $x \vdash \mathsf{p} = x$ states that the constant p is equal to all elements in the algebra (this means the algebra is a singleton).

Since interpretations are preserved by homomorphisms, it is expected that satisfaction is also preserved.

Lemma 14. *Let ϕ be a equation with context X. If $h : \mathbb{A} \to \mathbb{B}$ is a homomorphism and $\mathbb{A} \vDash^{\iota} \phi$ for an assignment $\iota : X \to A$, then $\mathbb{B} \vDash^{h\circ\iota} \phi$.*

Proof. Let ϕ be the equation $X \vdash s = t$, we have

$$\begin{aligned} \mathbb{A} \vDash^{\iota} \phi &\iff [\![s]\!]^{\iota}_A = [\![t]\!]^{\iota}_A && \text{definition of } \vDash \\ &\implies h([\![s]\!]^{\iota}_A) = h([\![t]\!]^{\iota}_A) && \\ &\implies [\![s]\!]^{h\circ\iota}_B = [\![t]\!]^{h\circ\iota}_B && \text{by (9)} \\ &\iff \mathbb{B} \vDash^{h\circ\iota} \phi. && \text{definition of } \vDash \end{aligned}$$

□

What is more surprising is that flattening interacts well with intepreting in the following sense.

Lemma 15. *For any Σ-algebra $\mathbb{A}$, the following square commutes.*[17]

$$\begin{array}{ccc} T_\Sigma T_\Sigma A & \xrightarrow{T_\Sigma [\![-]\!]_A} & T_\Sigma A \\ {\scriptstyle \mu^\Sigma_A}\downarrow & & \downarrow{\scriptstyle [\![-]\!]_A} \\ T_\Sigma A & \xrightarrow[{[\![-]\!]_A}]{} & A \end{array} \tag{11}$$

[17] In words, given a term in $T_\Sigma T_\Sigma A$, you obtain the same result if you interpret its flattening in $\mathbb{A}$, or if you interpret the term obtained by first interpreting all the "inner" terms.

Proof. We proceed by induction. For the base case, we have

$$[\![\mu^\Sigma_A(\eta^\Sigma_A(t))]\!]_A \overset{(4)}{=} [\![t]\!]_A \overset{(6)}{=} [\![\eta^\Sigma_A([\![t]\!]_A)]\!]_A \overset{(3)}{=} [\![T_\Sigma[\![-]\!]_A(\eta^\Sigma_A(t))]\!].$$

For the inductive step, if $t = \mathsf{op}(t_1, \dots, t_n)$, then

$$\begin{aligned} [\![\mu^\Sigma_A(t)]\!]_A &= [\![\mathsf{op}(\mu^\Sigma_A(t_1), \dots, \mu^\Sigma_A(t_n))]\!]_A && \text{by (4)} \\ &= [\![\mathsf{op}]\!]_A\left([\![\mu^\Sigma_A(t_1)]\!]_A, \dots, [\![\mu^\Sigma_A(t_n)]\!]_A\right) && \text{by (6)} \\ &= [\![\mathsf{op}]\!]_A\left([\![T_\Sigma[\![-]\!]_A(t_1)]\!]_A, \dots, [\![T_\Sigma[\![-]\!]_A(t_n)]\!]_A\right) && \text{I.H.} \\ &= [\![\mathsf{op}(T_\Sigma[\![-]\!]_A(t_1), \dots, T_\Sigma[\![-]\!]_A(t_n))]\!]_A && \text{by (6)} \\ &= [\![T_\Sigma[\![-]\!]_A(\mathsf{op}(t_1, \dots, t_n))]\!]_A && \text{by (2)} \\ &= [\![T_\Sigma[\![-]\!]_A(t)]\!]_A. \end{aligned}$$

□

Remark 16. To see Lemma 15 in another way, notice that (11) looks a lot like (8), but the map on the left is not the interpretation on an algebra. Except it is! Indeed, we can give a trivial interpretation of $\mathsf{op} : n \in \Sigma$ on the set $T_\Sigma A$ by $[\![\mathsf{op}]\!]_{T_\Sigma A}(t_1, \dots, t_n) = \mathsf{op}(t_1, \dots, t_n)$. Then, we can verify by induction[18] that $[\![-]\!]_{T_\Sigma A} : T_\Sigma T_\Sigma A \to T_\Sigma A$ is equal to μ^Σ_A. We conclude that Lemma 15 says that for any algebra, $[\![-]\!]_A$ is a homomorphism from $(T_\Sigma A, [\![-]\!]_{T_\Sigma A})$ to $\mathbb{A}$.

[18] Or we can compare (4) and (6) to see they become the same inductive definition in this instance.

In light of this remark, we mention two very similar results: given a set A, μ^Σ_A is a homomorphism between $T_\Sigma T_\Sigma A$ and $T_\Sigma A$, and given function $f : A \to B$, $T_\Sigma f$ is a homomorphism between $T_\Sigma A$ and $T_\Sigma B$.

Lemma 17. *For any function $f : A \to B$, the following squares commute.*[19]

$$
\begin{array}{ccc}
T_\Sigma T_\Sigma T_\Sigma A & \xrightarrow{T_\Sigma \mu_A^\Sigma} & T_\Sigma T_\Sigma A \\
{\scriptstyle \mu^\Sigma_{T_\Sigma A}} \downarrow & & \downarrow {\scriptstyle \mu_A^\Sigma} \\
T_\Sigma T_\Sigma A & \xrightarrow[\mu_A^\Sigma]{} & T_\Sigma
\end{array}
\qquad (12)
\qquad\qquad
\begin{array}{ccc}
T_\Sigma T_\Sigma A & \xrightarrow{T_\Sigma T_\Sigma B} & T_\Sigma T_\Sigma B \\
{\scriptstyle \mu^\Sigma_{A}} \downarrow & & \downarrow {\scriptstyle \mu_B^\Sigma} \\
T_\Sigma & \xrightarrow[T_\Sigma f]{} & T_\Sigma B
\end{array}
\qquad (13)
$$

[19] *Proof.* We have already shown both these squares commute. Indeed, (12) is an instance of (11) where we identify μ_A^Σ with the interpretation $[\![-]\!]_{T_\Sigma A}$ as explained in Remark 16, and (13) is the naturality square (5).

Another consequence of (12) is that if you have a term in $T_\Sigma^n A$ for any $n \in \mathbb{N}$, there are $(n-1)!$ ways to flatten it[20] by successively applying an instance of $T_\Sigma^i \mu^\Sigma_{T_\Sigma^j A}$ with different i and j (i.e. flattening at different levels inside the term), but all these ways lead to the same end result in $T_\Sigma A$. It is like when you have an expression built out of additions with possibly lots of nested bracketing, you can compute the sums in any order you want and it will give the same result. That property of addition is called associativity, and we will also say μ^Σ is associative.

[20] There is 1 way to flatten a term in $T_\Sigma^2 A$ to one in $T_\Sigma A$, and there are $n-1$ ways to flatten from $T_\Sigma^n A$ to $T_\Sigma^{(n-1)} A$. By induction, we find $(n-1)!$ possible combinations of flattening $T_\Sigma^n A \to T_\Sigma A$.

Given a set E of equations, we say $\mathbb{A}$ satisfies E and write $\mathbb{A} \vDash E$ if $\mathbb{A} \vDash \phi$ for all $\phi \in E$.[21] A (Σ, E)**-algebra** is a Σ-algebra that satisfies E. We define $\mathbf{Alg}(\Sigma, E)$, the category of (Σ, E)-algebras, to be the full subcategory of $\mathbf{Alg}(\Sigma)$ containing only those algebras that satisfy E. There is an evident forgetful functor $U : \mathbf{Alg}(\Sigma, E) \to \mathbf{Set}$ which is the composition of the inclusion functor $\mathbf{Alg}(\Sigma, E) \to \mathbf{Alg}(\Sigma)$ and $U : \mathbf{Alg}(\Sigma) \to \mathbf{Set}$.[22]

[21] Similarly for satisfaction under a particular assignment ι:
$$\mathbb{A} \vDash^\iota E \iff \forall \phi \in E, \mathbb{A} \vDash^\iota \phi.$$

[22] We will denote all the forgetful functors with the symbol U unless we need to emphasize the distinction. However, thanks to the `knowledge` package, you can click on (or hover) that symbol to check exactly which forgetful functor it is referring to.

Examples 18. 1. With $\Sigma = \{\mathsf{p} : 0\}$, there are morally only three different equations:[23]

$$\vdash \mathsf{p} = \mathsf{p}, \quad x \vdash \mathsf{p} = x, \text{ and } \quad x, y \vdash x = y.$$

Any algebra $\mathbb{A}$ satisfies the first because $[\![\mathsf{p}]\!]_A^\iota = [\![\mathsf{p}]\!]_A^\iota$, where $\iota : \varnothing \to A$ is the only possible assignment.[24] If $\mathbb{A}$ satisfies the second, it means that A is a singleton because for any $a, b \in A$, the assignments $\iota_a = x \mapsto a$ and $\iota_b = x \mapsto b$ give us[25]

$$a = \iota_a(x) = [\![x]\!]_A^{\iota_a} = [\![\mathsf{p}]\!]_A^{\iota_a} = [\![\mathsf{p}]\!]_A^{\iota_b} = [\![x]\!]_A^{\iota_b} = \iota_b(x) = b.$$

If $\mathbb{A}$ satisfies the third equation, it is also a singleton because for any $a, b \in A$, the assignment ι sending x to a and y to b gives us

$$a = \iota(x) = [\![x]\!]_A^\iota = [\![y]\!]_A^\iota = \iota(y) = b.$$

Therefore,[26] there are only two things $\mathbf{Alg}(\Sigma, E)$ can be for any E, either it is all of $\mathbf{Alg}(\Sigma)$, or it contains only the singletons.

[23] Let us not formally argue that here, but your intuition about equality and the fact that terms in $T_\Sigma X$ are either $x \in X$ or p should be enough to convince you at this point.

[24] We write nothing before the turnstile ($\vdash$) instead of the empty set $\varnothing$.

[25] We find $a = b$ for any $a, b \in A$ and A contains at least one element, the interpretation of the constant p, so A is a singleton.

[26] Modulo the argument about these being all the possible equations over Σ.

2. With $\Sigma = \{+ : 2, \mathsf{e} : 0\}$, there are many more possible equations, but the following three are quite famous:

$$x, y, z \vdash x + (y + z) = (x + y) + z, \quad x, y \vdash x + y = y + x, \text{ and } \quad x \vdash x + \mathsf{e} = x. \quad (14)$$

We already saw in Example 12 that the first asserts associativity of the interpretation of $+$. With a similar argument, one shows that the second asserts $[\![+]\!]$ is commutative, and the third asserts $[\![\mathsf{e}]\!]$ is a neutral element (on the right) for $[\![+]\!]$.[27] Moreover, note that a homomorphism of Σ-algebras from $\mathbb{A}$ to $\mathbb{B}$ is any

[27] i.e. if $\mathbb{A}$ satisfies $x \vdash x + \mathsf{e} = x$, then for all $a \in A$,
$$[\![+]\!]_A(a, [\![\mathsf{e}]\!]_A) = a.$$

function $h : A \to B$ that satisfies

$$\forall a, a' \in A, \quad h(\llbracket + \rrbracket_A(a, a')) = \llbracket + \rrbracket_B(h(a), h(a')) \text{ and } h(\llbracket \mathsf{e} \rrbracket_A) = \llbracket \mathsf{e} \rrbracket_B.$$

Namely, a homomorphism preserves the "addition" and its neutral element. Thus, letting E be the set containing the three equations in (14), we find that $\mathbf{Alg}(\Sigma, E)$ is the category **CMon** of commutative monoids and monoid homomorphisms.

3. We can add a unary operation symbol $-$ to get $\Sigma = \{+ : 2, \mathsf{e} : 0, - : 1\}$, and add the equation $x \vdash x + (-x) = \mathsf{e}$ to those in (14),[28] and we can show that $\mathbf{Alg}(\Sigma, E)$ is the category **Ab** of abelian groups and group homomorphisms.

[28] While the signature has changed between the two examples, the equations of (14) can be understood over both signatures because they concern terms constructed using the symbols common to both signatures.

Definition 19 (Algebraic theory)**.** Given a set E of equations over Σ, the **algebraic theory** generated by E, denoted by $\mathfrak{Th}(E)$, is the class of equations (over Σ) that are satisfied in all (Σ, E)-algebras:[29]

[29] Note that there is no guarantee that $\mathfrak{Th}(E)$ is a set (in fact it never is) because there is no set of all equations (because the context can be any set).

$$\mathfrak{Th}(E) = \{X \vdash s = t \mid \forall \mathbb{A} \in \mathbf{Alg}(\Sigma, E), \mathbb{A} \vDash X \vdash s = t\}.$$

Formulated differently, $\mathfrak{Th}(E)$ contains the equations that are semantically entailed by E, namely $\phi \in \mathfrak{Th}(E)$ if and only if

$$\forall \mathbb{A} \in \mathbf{Alg}(\Sigma), \quad \mathbb{A} \vDash E \implies \mathbb{A} \vDash \phi.$$

Of course, $\mathfrak{Th}(E)$ contains all of E,[30] but also many more equations like $x \vdash x = x$ which is satisfied by any Σ-algebra. We will see in §1.3 how to find which equations are entailed by others.

[30] Because a (Σ, E)-algebra satisfies E by definition.

We call a class of equations an algebraic theory if it equals $\mathfrak{Th}(E)$ for some set E of generating equations.

Example 20. If E contains the equations in (14), then $\mathfrak{Th}(E)$ will contain all the equations that every commutative monoid satisfies. Here is a non-exhaustive list:

- $x \vdash \mathsf{e} + x = x$ says that $\llbracket \mathsf{e} \rrbracket$ is a neutral element on the left for $\llbracket + \rrbracket$ which is true because, by equations in (14), it $\llbracket \mathsf{e} \rrbracket$ is neutral on the right and $\llbracket + \rrbracket$ is commutative.
- $z, w \vdash z + w = w + z$ also states commutativity of $\llbracket + \rrbracket$ but with different variable names.
- $x, y, z, w \vdash (x + w) + (x + z) + (x + y) = ((x + x) + x) + (y + (z + (\mathsf{e} + w)))$ is just a random equation that can be shown using the properties of commutative monoids.

1.2 Free Algebras

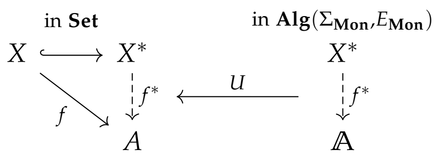
Up to now we have not given a single concrete example of an algebra, we give here a very special example.

Example 21 (Words). Let $\Sigma_{\mathbf{Mon}} = \{\cdot : 2, \mathtt{e} : 0\}$, $X = \{\mathtt{a}, \mathtt{b}, \cdots, \mathtt{z}\}$ be the set of (lowercase) letters in the latin alphabet, and X^* be the set of finite words using only these letters.[31] There is a natural $\Sigma_{\mathbf{Mon}}$-algebra structure on X^* where $+$ is interpreted as concatenation, i.e. $[\![\cdot]\!]_{X^*}(u,v) = uv$, and $\mathtt{e}$ as the empty word ε. This algebra satisfies the equations defining a monoid given in (15).

[31] We are talking about words in a mathematical sense, so X^* contains weird stuff like `aczlp` and the empty word ε.

$$E_{\mathbf{Mon}} = \{x,y,z \vdash x \cdot (y \cdot z) = (x \cdot y) \cdot z, \quad x \vdash x \cdot \mathtt{e} = x, \quad x \vdash \mathtt{e} \cdot x = x\}. \tag{15}$$

In fact, X^* is the *free* monoid over X. This means that for any other $(\Sigma_{\mathbf{Mon}}, E_{\mathbf{Mon}})$-algebra $\mathbb{A}$ and any function $f : X \to A$, there exists a unique homomorphism $f^* : X^* \to \mathbb{A}$ such that $f^*(x) = f(x)$ for all $x \in X \subseteq X^*$. This can be summarized in the following diagram.



(16)

The free $(\Sigma_{\mathbf{Mon}}, E_{\mathbf{Mon}})$-algebra over any set is always[32] the set of finite words over that set with $\cdot$ and $\mathtt{e}$ interpreted as concatenation and the empty word respectively.

[32] We have to say up to isomorphism here if we want to be fully rigorous. Let us avoid this bulkiness here and later in most places where it can be inferred.

At a first look, X^* does not seem correlated to the operation symbols in $\Sigma_{\mathbf{Mon}}$ and the equations in $E_{\mathbf{Mon}}$, so it may seem hopeless to generalize this construction of free algebra for an arbitrary Σ and E. It is possible however to describe the algebra X^* starting from $\Sigma_{\mathbf{Mon}}$ and $E_{\mathbf{Mon}}$.

Recall that $T_{\Sigma_{\mathbf{Mon}}}X$ is the set of all terms constructed with the symbols in $\Sigma_{\mathbf{Mon}}$ and the elements of X.[33] Since we want the interpretation of $\mathtt{e}$ to be a neutral element for the interpretation of $\cdot$, we could identify many terms together like $\mathtt{e}$ and $\mathtt{e} \cdot \mathtt{e}$, in fact whenever a term has an occuence of $\mathtt{e}$, we can remove it with no effect on its interpretation in a $(\Sigma_{\mathbf{Mon}}, E_{\mathbf{Mon}})$-algebra. Similarly, since we want $\cdot$ to be interpreted as an associative operation, we could identify $\mathtt{r} \cdot (\mathtt{s} \cdot \mathtt{m})$ and $(\mathtt{r} \cdot \mathtt{s}) \cdot \mathtt{m}$, and more generally, we can rearrange the parentheses in a term with no effect on its interpretation in a $(\Sigma_{\mathbf{Mon}}, E_{\mathbf{Mon}})$-algebra.

[33] For instance, it contains $\mathtt{e}$, $\mathtt{e} \cdot \mathtt{e}$, $\mathtt{a} \cdot \mathtt{a}$, $\mathtt{a} \cdot (\mathtt{r} \cdot (\mathtt{e} \cdot \mathtt{u}))$, and so on.

Squinting a bit, you can convince yourself that a $\Sigma_{\mathbf{Mon}}$-term over X considered modulo occurrences of $\mathtt{e}$ and parentheses is the same thing as a finite word in X^*.[34] Under this correspondence, we find that the interpretation of $\cdot$ on X^* (which was concatenation) can be realized syntactically by the symbol $\cdot$. For example, the concatenation of the words corresponding to $\mathtt{r} \cdot \mathtt{r}$ and $\mathtt{u} \cdot \mathtt{p}$ is the word corresponding to $(\mathtt{r} \cdot \mathtt{r}) \cdot (\mathtt{u} \cdot \mathtt{p})$. The interpretation of $\mathtt{e}$ in X^* is the empty word which corresponds to $\mathtt{e}$. We conclude that the algebra X^* could have been described entirely using the syntax of $\Sigma_{\mathbf{Mon}}$ and equations in $E_{\mathbf{Mon}}$.

[34] For instance, both $\mathtt{r} \cdot (\mathtt{s} \cdot \mathtt{m})$ and $(\mathtt{r} \cdot \mathtt{s}) \cdot \mathtt{m}$ become the word `rsm` and $\mathtt{e}$, $\mathtt{e} \cdot \mathtt{e}$ and $\mathtt{e} \cdot (\mathtt{e} \cdot \mathtt{e})$ all become the empty word.

We promptly generalize this to other signatures and sets of equations. Fix a signature Σ and a set E of equations over Σ. For any set X, we can define a binary relation $\equiv_E$ on Σ-terms[35] that contains the pair (s,t) whenever the interpretation of s and t coincide in any (Σ, E)-algebra. Formally, we have for any $s, t \in T_\Sigma X$,

[35] We omit the set X from the notation as it would be more bulky than illuminative.

$$s \equiv_E t \iff X \vdash s = t \in \mathfrak{Th}(E). \tag{17}$$

We now show $\equiv_E$ is a congruence relation.

Lemma 22. *For any set X, the relation $\equiv_E$ is reflexive, symmetric, transitive, and satisfies for any $\mathsf{op}: n \in \Sigma$ and $s_1, \ldots, s_n, t_1, \ldots, t_n \in T_\Sigma X$,*

$$\forall 1 \leq i \leq n, s_i \equiv_E t_i \implies \mathsf{op}(s_1, \ldots, s_n) \equiv_E \mathsf{op}(t_1, \ldots, t_n). \tag{18}$$

Proof. Briefly, reflexivity, symmetry and transitivity all follow from the fact that equality satisfies these properties, and (18) follows from the fact that operation symbols are interpreted as *deterministic* functions, so they preserve equality. We detail this below.

(*Reflexivity*) For any $t \in T_\Sigma X$, and any Σ-algebra $\mathbb{A}$, $\mathbb{A} \vDash X \vdash t = t$ because it holds that $[\![t]\!]^\iota_A = [\![t]\!]^\iota_A$ for all $\iota : X \to A$.

(*Symmetry*) For any $s, t \in T_\Sigma X$ and $\mathbb{A} \in \mathbf{Alg}(\Sigma)$, if $\mathbb{A} \vDash X \vdash s = t$, then $\mathbb{A} \vDash X \vdash t = s$. Indeed, is $[\![s]\!]^\iota_A = [\![t]\!]^\iota_A$ holds for all ι, then $[\![t]\!]^\iota_A = [\![s]\!]^\iota_A$ holds too. Symmetry follows because if all (Σ, E)-algebras satisfy $X \vdash s = t$, then they also satisfy $X \vdash t = s$.

(*Transitivity*) For any $s, t, u \in T_\Sigma X$, if all (Σ, E)-algebras satisfy $X \vdash s = t$ and $X \vdash t = u$, then they also satisfy $X \vdash s = u$.[36] Transitivity follows.

[36] Just like for symmetry, it is because for any $\mathbb{A} \in \mathbf{Alg}(\Sigma)$ and $\iota : X \to A$, $[\![s]\!]^\iota_A = [\![t]\!]^\iota_A$ with $[\![t]\!]^\iota_A = [\![u]\!]^\iota_A$ imply $[\![s]\!]^\iota_A = [\![u]\!]^\iota_A$.

(18) For any $\mathsf{op}: n \in \Sigma$, $s_1, \ldots, s_n, t_1, \ldots, t_n \in T_\Sigma X$, and $\mathbb{A} \in \mathbf{Alg}(\Sigma)$, if $\mathbb{A}$ satisfies $X \vdash s_i = t_i$ for all i, then for any assignment $\iota : X \to A$, we have $[\![s_i]\!]^\iota_A = [\![t_i]\!]^\iota_A$ for all i. Hence,

$$\begin{aligned} [\![\mathsf{op}(s_1, \ldots, s_n)]\!]^\iota_A &= [\![\mathsf{op}]\!]_A([\![s_1]\!]^\iota_A, \ldots, [\![s_n]\!]^\iota_A) && \text{by (6)} \\ &= [\![\mathsf{op}]\!]_A([\![t_1]\!]^\iota_A, \ldots, [\![t_n]\!]^\iota_A) && \forall i, [\![s_i]\!]^\iota_A = [\![t_i]\!]^\iota_A \\ &= [\![\mathsf{op}(s_1, \ldots, s_n)]\!]^\iota_A && \text{by (6),} \end{aligned}$$

which means $\mathbb{A} \vDash X \vdash \mathsf{op}(s_1, \ldots, s_n) = \mathsf{op}(t_1, \ldots, t_n)$. This was true for all Σ-algebras, so we can use the same arguments as above to conclude (18).

□

This lemma show $\equiv_E$ is an equivalence relation, so we can define terms modulo E. Given Σ, E and X, let $T_{\Sigma,E}X = T_\Sigma X / {\equiv_E}$ denote the set of **Σ-terms modulo** E. We will write $[-]_E : T_\Sigma X \to T_{\Sigma,E}X$ for the canonical quotient map, so $[t]_E$ is the equivalence class of t in $T_{\Sigma,E}X$.

This yields a functor $T_{\Sigma,E} : \mathbf{Set} \to \mathbf{Set}$ which sends a function $f : X \to Y$ to the unique function $T_{\Sigma,E}f$ making (19) commute, i.e. satisfying $T_{\Sigma,E}f([t]_E) = [T_\Sigma f(t)]_E$. By definition, $[-]_E$ is also a natural transformation from T_Σ to $T_{\Sigma,E}$.

$$\begin{array}{ccc} T_\Sigma X & \xrightarrow{[-]_E} & T_{\Sigma,E}X \\ {\scriptstyle T_\Sigma f}\downarrow & & \downarrow{\scriptstyle T_{\Sigma,E}f} \\ T_\Sigma Y & \xrightarrow[{[-]_E}]{} & T_{\Sigma,E}Y \end{array} \tag{19}$$

Definition 23 (Term algebra, semantically)**.** The **term algebra** for (Σ, E) on X is the Σ-algebra whose carrier is $T_{\Sigma,E}X$ and whose interpretation of $\mathsf{op}: n \in \Sigma$ is defined by[37]

$$[\![\mathsf{op}]\!]_{\mathbb{T}X}([t_1]_E, \ldots, [t_n]_E) = [\mathsf{op}(t_1, \ldots, t_n)]_E. \tag{20}$$

[37] This is well-defined (i.e. invariant under change of representative) by (18).

We denote this algebra by $\mathbb{T}_{\Sigma,E}X$ or simply $\mathbb{T}X$.

A main motivation behind this definition is that it makes $[-]_E : T_\Sigma X \to T_{\Sigma,E}X$ a homomorphism,[38] namely, (21) commutes.

[38] Indeed, (20) looks exactly like (1) with $h = [-]_E$, $A = T_\Sigma X$ and $B = \mathbb{T}X$.

$$\begin{array}{ccc} T_\Sigma T_\Sigma X & \xrightarrow{T_\Sigma[-]_E} & T_\Sigma T_{\Sigma,E} X \\ {\scriptstyle \mu^\Sigma_X}\downarrow & & \downarrow{\scriptstyle [\![-]\!]_{\mathbb{T}X}} \\ T_\Sigma X & \xrightarrow[{[-]_E}]{} & T_{\Sigma,E} X \end{array} \tag{21}$$

It is very easy to "compute" in the term algebra because all operations are realized syntactically, i.e. only by manipulating symbols. Let us first look at the interpretation of Σ-terms in $\mathbb{T}X$, i.e. the function $[\![-]\!]_{\mathbb{T}X} : T_\Sigma T_{\Sigma,E} X \to T_{\Sigma,E} X$. It was defined inductively to yield[39]

$$[\![\eta^\Sigma_{T_{\Sigma,E}X}([t]_E)]\!]_{\mathbb{T}X} = [t]_E \text{ and } [\![\mathsf{op}(t_1,\dots,t_n)]\!]_{\mathbb{T}X} = [\![\mathsf{op}]\!]_{\mathbb{T}X}([\![t_1]\!]_{\mathbb{T}X},\dots,[\![t_n]\!]_{\mathbb{T}X}). \tag{22}$$

[39] where $t \in T_\Sigma X$, $\mathsf{op}: n \in \Sigma$, and $t_1,\dots,t_n \in T_\Sigma T_{\Sigma,E} X$.

Remark 24. In particular, when E is empty, the set $T_{\Sigma,\emptyset} X$ is $T_\Sigma X$ quotiented by $\equiv_\emptyset$, but one can show that equivalence relation $\equiv_\emptyset$ is equal to equality ($=$), i.e. $\mathfrak{Th}(\emptyset)$ only contains equation of the form $X \vdash t = t$.[40] Therefore, $T_{\Sigma,\emptyset} X = T_\Sigma X$. Moreover, since $[-]_\emptyset$ is the identity map, we find that (20) becomes the definition of the interpretations given in Remark 16, so $\mathbb{T}_{\Sigma,\emptyset} X$ is the algebra on $T_\Sigma X$ we had defined. Also, we find the interpretation of terms $[\![-]\!]_{\mathbb{T}_{\Sigma,\emptyset}X}$ is the flattening.[41]

[40] For any other equation $X \vdash s = t$ where s and t are not the same term, the Σ-algebra $T_\Sigma X$ does not satisfy it. Indeed, take the assignment $\eta^\Sigma_X : X \to T_\Sigma X$.

[41] By Remark 16 or by comparing (22) when $E = \emptyset$ and μ^Σ_X.

Example 25. Let $\Sigma = \Sigma_{\mathbf{Mon}}$ and $E = E_{\mathbf{Mon}}$ be the signature and equations defning monoids as explained in Example 21. We saw informally that $T_{\Sigma,E} X$ is in correspondence with the set X^* of finite words over X, and we already have a monoid structure on X^*.[42] Thus, we may wonder whether the term algebra $\mathbb{T}X$ describes the same monoid. Let us compute the interpretation of $u \cdot (v \cdot w)$ where $u = \mathtt{uu}$, $v = \mathtt{vv}$ and $w = \mathtt{www}$ are words in $X^* \cong T_{\Sigma,E} X$. First we use the inductive definition:

[42] The interpretation of $\cdot$ and e is concatenation and the empty word.

$$[\![u \cdot (v \cdot w)]\!]_{\mathbb{T}X} = [\![\cdot]\!]_{\mathbb{T}X}([\![u]\!]_{\mathbb{T}X}, [\![v \cdot w]\!]_{\mathbb{T}X}) = [\![\cdot]\!]_{\mathbb{T}X}([\![u]\!]_{\mathbb{T}X}, [\![\cdot]\!]_{\mathbb{T}X}([\![v]\!]_{\mathbb{T}X}, [\![w]\!]_{\mathbb{T}X})).$$

Next, we choose a representative for $u, v, w \in T_{\Sigma,E} X$ and apply the base step of the inductive definition:

$$[\![u \cdot (v \cdot w)]\!]_{\mathbb{T}X} = [\![\cdot]\!]_{\mathbb{T}X}([\mathtt{u} \cdot \mathtt{u}]_E, [\![\cdot]\!]_{\mathbb{T}X}([\mathtt{v} \cdot \mathtt{v}]_E, [\mathtt{w} \cdot (\mathtt{w} \cdot \mathtt{w})]_E)).$$

Finally, we can apply (20) a couple times to find

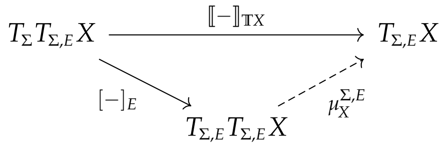
$$[\![u \cdot (v \cdot w)]\!]_{\mathbb{T}X} = [\![\cdot]\!]_{\mathbb{T}X}([\mathtt{u} \cdot \mathtt{u}]_E, [(\mathtt{v} \cdot \mathtt{v}) \cdot (\mathtt{w} \cdot (\mathtt{w} \cdot \mathtt{w}))]_E) = [(\mathtt{u} \cdot \mathtt{u}) \cdot ((\mathtt{v} \cdot \mathtt{v}) \cdot (\mathtt{w} \cdot (\mathtt{w} \cdot \mathtt{w})))]_E,$$

which means that the word corresponding to $[\![u \cdot (v \cdot w)]\!]_{\mathbb{T}X}$ is $\mathtt{uuvvwww}$, i.e. the concatenation of u, v and w.

In general (for other signatures), what happens when applying $[\![-]\!]_{\mathbb{T}X}$ to some big term in $T_\Sigma T_{\Sigma,E} X$ can be decomposed in three steps.

1. Apply the inductive definition until you have an expression built out of many $[\![\mathsf{op}]\!]_{\mathbb{T}X}$ and $[\![c]\!]_{\mathbb{T}X}$ where $\mathsf{op} \in \Sigma$ and c is an equivalence class of Σ-terms.
2. Choose a representative for each such classes (i.e. $c = [t]_E$).
3. Use (20) repeatedly until the result is just an equivalence class in $T_{\Sigma,E} X$.

More intuitively, if you have written the term with a choice of representative for all equivalence classes it uses, you can remove all brackets inside and put one pair around the whole term. In this sense, $[\![-]\!]_{\mathbb{T}X}$ looks a lot like the flattening μ_X^Σ except it deals with equivalence classes of terms. This motivates the definition of $\mu_X^{\Sigma,E}$ to be the unique function making (23) commute.[43]



(23)

[43] This guarantees $\mu_X^{\Sigma,E}$ satisfies the following equations that look like the inductive definition of μ_X^Σ in (4): for any $t \in T_\Sigma X$, $\mu_X^{\Sigma,E}([[t]_E]_E) = [t]_E$ and for any $\mathsf{op} : n \in \Sigma$ and $t_1, \dots, t_n \in T_\Sigma X$,

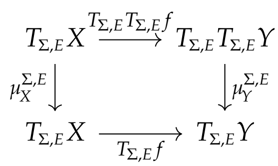
$$\mu_X^{\Sigma,E}([\mathsf{op}([t_1]_E, \dots, [t_n]_E)]_E) = [\mathsf{op}(t_1, \dots, t_n)]_E.$$

Thanks to Remark 24, we can immediately see that $\mu_X^{\Sigma,\emptyset} = \mu_X^\Sigma$ because $[-]_\emptyset$ is the identity, and $[\![-]\!]_{\mathbb{T}_{\Sigma,\emptyset}X} = \mu_X^\Sigma$.

The first thing we showed when defining μ_X^Σ was that it yielded a natural transformation $\mu^\Sigma : T_\Sigma T_\Sigma \Rightarrow T_\Sigma$. We can also do this for $\mu^{\Sigma,E}$.

Proposition 26. *The family of maps $\mu_X^{\Sigma,E} : T_{\Sigma,E}T_{\Sigma,E}X \to T_{\Sigma,E}X$ is natural in X.*

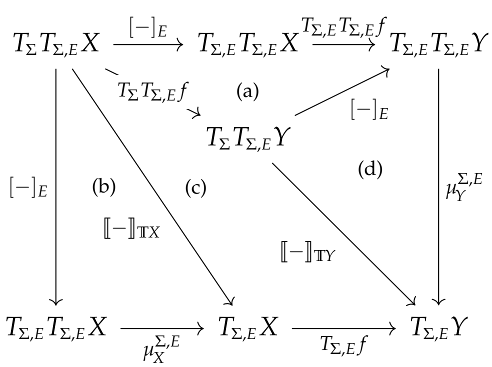
Proof. We need to prove that for any function $f : X \to Y$, the following square commutes.



(24)

We can pave the following diagram.[44]

[44] By paving a diagram, we mean to build a large diagram out of smaller ones, showing all the smaller one commute, and then concluding the bigger must commute. We often refer parts of the diagram with them letters written inside them, and explain how each of them commutes one at a time.

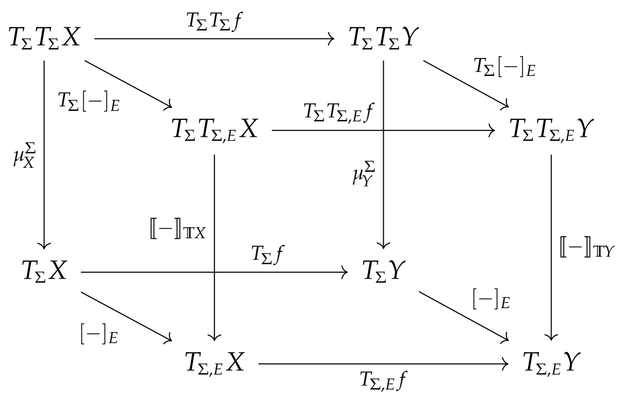


All of (a), (b) and (d) commute by definition. In more details, (a) is an instance of (19) with X replaced by $T_{\Sigma,E}X$, Y by $T_{\Sigma,E}Y$ and f by $T_{\Sigma,E}f$, and both (b) and (d) are instances of (23). To show (c) commutes, we draw another diagram that looks like a cube and where (c) is the front face. We can show all the other faces commute, and then use the fact that $T_\Sigma[-]_E$ is surjective (i.e. epic) to conclude that the front face must also commute.[45]

[45] In more details, the left and right faces commute by (21), the bottom and top faces commute by (19), and the back face commutes by (5).

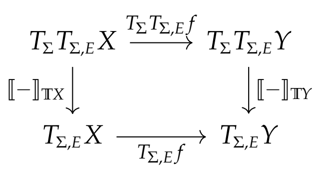
The function $T_\Sigma[-]_E$ is surjective (i.e. epic) because $[-]_E$ is (it is a canonical quotient map) and functors on **Set** preserve epimorphisms (if we assume the axiom of choice). Thus, it suffices to show that $T_\Sigma[-]_E$ pre-composed with the bottom path or the top path of the front face gives the same result.

Now it is just a matter of going around the cube using the commutativity of the other faces. Here is



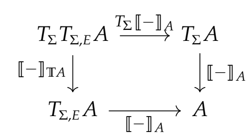
□

The front face of the cube is interesting on its own, it says that for any function $f : X \to Y$, $T_{\Sigma,E} f$ is a homomorphism from $\mathbb{T}_{\Sigma,E} X$ to $\mathbb{T}_{\Sigma,E} Y$. We redraw it below for future reference.



(25)

Stating it like this may remind you of Lemma 15 and Remark 16. We will need a variant of Lemma 15 for $T_{\Sigma,E}$, but there is a slight obstacle due to types. Indeed, given a Σ-algebra $\mathbb{A}$ we would like to prove a square like in (26) commutes.

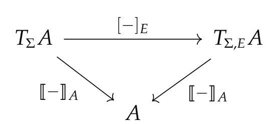


(26)

However, the arrows on top and bottom do not really exist, the interpretation $[\![-]\!]_A$ takes terms over A as input, not equivalence classes of terms. The quick fix is to assume that $\mathbb{A}$ satisfies the equations in E. This means that $[\![-]\!]_A$ is well-defined on equivalence class of terms becuase if $[s]_E = [t]_E$, then $A \vdash s = t \in \mathfrak{Th}(E)$, so $\mathbb{A}$ satisfies that equation, i.e. taking the assignment $\mathrm{id}_A : A \to A$,

$$[\![s]\!]_A = [\![s]\!]_A^{\mathrm{id}_A} = [\![t]\!]_A^{\mathrm{id}_A} = [\![t]\!]_A.$$

When $\mathbb{A}$ is a (Σ, E)-algebra, we abusively write $[\![-]\!]_A$ for the interpretation of terms and equivalence classes of terms as in (27).



(27)

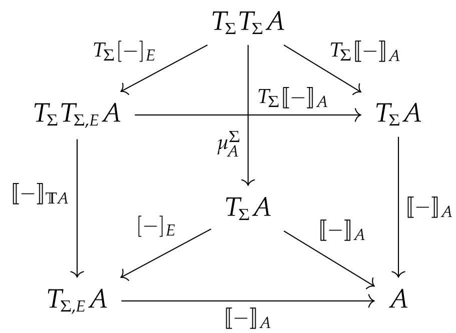
Lemma 27. *For any (Σ, E)-algebra $\mathbb{A}$, the square (26) commutes.*

Proof. Consider the following diagram that we can view as a triangular prism and whose front face is (26). Both triangles commute by (27), the square face at the back and on the left commutes by (21), and the square face at the back and on the right commutes by (11). With the same trick as in the proof of Proposition 26 using the surjectivity of $T_\Sigma[-]_E$, we conclude that the front face commutes.[46]

[46] Here is the complete derivation.

$[\![-]\!]_A \circ [\![-]\!]_{\mathbb{T}A} \circ T_\Sigma[-]_E$	
$= [\![-]\!]_A \circ [-]_E \circ \mu^\Sigma_A$	left
$= [\![-]\!]_A \circ \mu^\Sigma_A$	bottom
$= [\![-]\!]_A \circ T_\Sigma [\![-]\!]_A$	right
$= [\![-]\!]_A \circ T_\Sigma [\![-]\!]_A \circ T_\Sigma[-]_E$	top

Then, since $T_\Sigma[-]_E$ is epic, we conclude that $[\![-]\!]_A \circ [\![-]\!]_{\mathbb{T}A} = [\![-]\!]_A \circ T_\Sigma [\![-]\!]_A$.



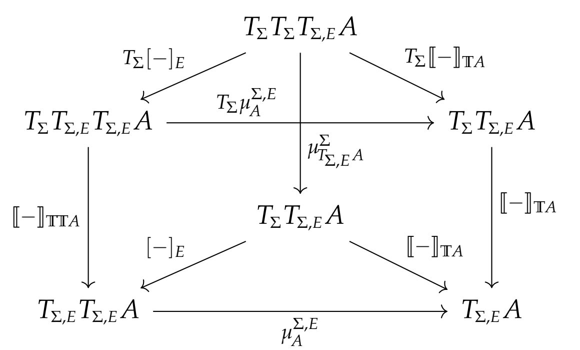
□

An important consequence of Lemma 15 was (12) saying that flattening is a homomorphism from $\mathbb{T}_{\Sigma,\varnothing}\mathbb{T}_{\Sigma,\varnothing}A$ to $\mathbb{T}_{\Sigma,\varnothing}A$. This is also true when E is not empty, i.e. $\mu_A^{\Sigma,E}$ is a homomorphism frmo $\mathbb{T}\mathbb{T}A$ to $\mathbb{T}A$.

Lemma 28. *For any set A, the following square commutes.*

$$\begin{array}{ccc} T_\Sigma T_{\Sigma,E} T_{\Sigma,E} A & \xrightarrow{T_\Sigma \mu_A^{\Sigma,E}} & T_\Sigma T_{\Sigma,E} A \\ {\scriptstyle [\![-]\!]_{\mathbb{T}\mathbb{T}A}} \downarrow & & \downarrow {\scriptstyle [\![-]\!]_{\mathbb{T}A}} \\ T_{\Sigma,E} T_{\Sigma,E} A & \xrightarrow[\mu_A^{\Sigma,E}]{} & T_{\Sigma,E} A \end{array} \tag{28}$$

Proof. We prove it exactly like Lemma 27 with the following diagram.[47]



□

[47] The top and bottom faces commute by definition of $\mu_A^{\Sigma,E}$ (23), the back-left face by (21), and the back-right face by (11).
Then, $T_\Sigma[-]_E$ is epic, so the following derivation suffices.

$$\begin{aligned} & \mu_A^{\Sigma,E} \circ [\![-]\!]_{\mathbb{T}\mathbb{T}A} \circ T_\Sigma[-]_E \\ &= \mu_A^{\Sigma,E} \circ [-]_E \circ \mu^\Sigma_{T_{\Sigma,E}A} && \text{left} \\ &= [\![-]\!]_{\mathbb{T}A} \circ \mu^\Sigma_{T_{\Sigma,E}A} && \text{bottom} \\ &= [\![-]\!]_{\mathbb{T}A} \circ T_\Sigma [\![-]\!]_{\mathbb{T}A} && \text{right} \\ &= [\![-]\!]_{\mathbb{T}A} \circ T_\Sigma \mu_A^{\Sigma,E} \circ T_\Sigma[-]_E && \text{top} \end{aligned}$$

In a moment, we will show that $\mathbb{T}_{\Sigma,E}X$ is not only a Σ-algebra, but also a (Σ,E)-algebra. This requires us to talk about satisfaction of equations, hence about the interpretation of terms in some $T_\Sigma Y$ under an assignment $\sigma : Y \to T_{\Sigma,E}X$. By the definition $[\![-]\!]^\sigma_{\mathbb{T}X} = [\![-]\!]_{\mathbb{T}X} \circ T_\Sigma\sigma$, and our informal description of $[\![-]\!]_{\mathbb{T}X}$, we can infer that $[\![t]\!]^\sigma_{\mathbb{T}X}$ is the equivalence class of the term t where all occurences of the variable y have been substituted by a representative of $\sigma(y)$.

In particular, this means that under the assignment $\sigma : X \to T_{\Sigma,E}X$ that sends a variable x to its equivalence class $[x]_E$, the interpretation of a term $t \in T_\Sigma X$ is $[t]_E$.[48] We prove this formally below.

[48] The representative chosen for $\sigma(x)$ is x so the term t is not modified.

Lemma 29. *Let* $\sigma = X \xrightarrow{\eta^\Sigma_X} T_\Sigma X \xrightarrow{[-]_E} T_{\Sigma,E} X$ *be an assignment. Then,* $[\![-]\!]^\sigma_{\mathbb{T}X} = [-]_E$.

Proof. We proceed by induction. For the base case, we have

$$\begin{aligned}
[\![\eta^\Sigma_X(x)]\!]^\sigma_{\mathbb{T}X} &= [\![T_\Sigma\sigma(\eta^\Sigma_X(x))]\!]_{\mathbb{T}X} && \text{by (7)}\\
&= [\![T_\Sigma[-]_E(T_\Sigma\eta^\Sigma_X(\eta^\Sigma_X(x)))]\!]_{\mathbb{T}X} && \text{by Proposition 6}\\
&= [\![T_\Sigma[-]_E(\eta^\Sigma_{T_\Sigma X}(\eta^\Sigma_X(x)))]\!]_{\mathbb{T}X} && \text{by (3)}\\
&= [\![\eta^\Sigma_{T_{\Sigma,E}X}([\eta^\Sigma_X(x)]_E)]\!]_{\mathbb{T}X} && \text{by (3)}\\
&= [\eta^\Sigma_X(x)]_E && \text{by (22)}
\end{aligned}$$

For the inductive step, if $t = \mathsf{op}(t_1, \dots, t_n)$, we have

$$\begin{aligned}
[\![t]\!]^\sigma_{\mathbb{T}X} &= [\![T_\Sigma\sigma(t)]\!]_{\mathbb{T}X} && \text{by (7)}\\
&= [\![T_\Sigma[-]_E(T_\Sigma\eta^\Sigma_X(t))]\!]_{\mathbb{T}X} && \text{by Proposition 6}\\
&= [\![T_\Sigma[-]_E(T_\Sigma\eta^\Sigma_X(\mathsf{op}(t_1,\dots,t_n)))]\!]_{\mathbb{T}X} &&\\
&= [\![T_\Sigma[-]_E(\mathsf{op}(T_\Sigma\eta^\Sigma_X(t_1),\dots,T_\Sigma\eta^\Sigma_X(t_n)))]\!]_{\mathbb{T}X} && \text{by (2)}\\
&= [\![\mathsf{op}(T_\Sigma[-]_E(T_\Sigma\eta^\Sigma_X(t_1)),\dots,T_\Sigma[-]_E(T_\Sigma\eta^\Sigma_X(t_n)))]\!]_{\mathbb{T}X} && \text{by (2)}\\
&= [\![\mathsf{op}]\!]_{\mathbb{T}X}\left([\![T_\Sigma[-]_E(T_\Sigma\eta^\Sigma_X(t_1))]\!]_{\mathbb{T}X},\cdots,[\![T_\Sigma[-]_E(T_\Sigma\eta^\Sigma_X(t_n))]\!]_{\mathbb{T}X}\right) && \text{by (22)}\\
&= [\![\mathsf{op}]\!]_{\mathbb{T}X}\left([t_1]_E,\cdots,[t_n]_E\right) && \text{I.H.}\\
&= [\mathsf{op}(t_1,\dots,t_n)]_E && \text{by (20)} \quad \square
\end{aligned}$$

We will denote that special assignment $\eta^{\Sigma,E}_X = [-]_E \circ \eta^\Sigma_X : X \to T_{\Sigma,E}X$.[49] A quick corollary of the previous lemma is that for any equation ϕ with context X, ϕ belongs to E if and only if the algebra $\mathbb{T}_{\Sigma,E}X$ satisfies it under the assignment $\eta^{\Sigma,E}_X$.

[49] Note that $\eta^{\Sigma,E}$ becomes a natural transformation $\mathrm{id}_{\mathbf{Set}} \to T_{\Sigma,E}$ because it is the vertical composition $[-]_E \cdot \eta^\Sigma$.

Lemma 30. *Let* $s,t \in T_\Sigma X$*,* $X \vdash s = t \in E$ *if and only if* $\mathbb{T}_{\Sigma,E}X \vDash^{\eta^{\Sigma,E}_X} X \vdash s = t$.[50]

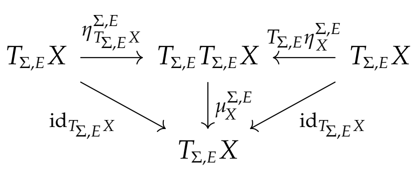
[50] *Proof.* By Lemma 29, we have

$$[\![s]\!]^{\eta^{\Sigma,E}_X}_{\mathbb{T}X} = [s]_E \text{ and } [\![t]\!]^{\eta^{\Sigma,E}_X}_{\mathbb{T}X} = [t]_E,$$

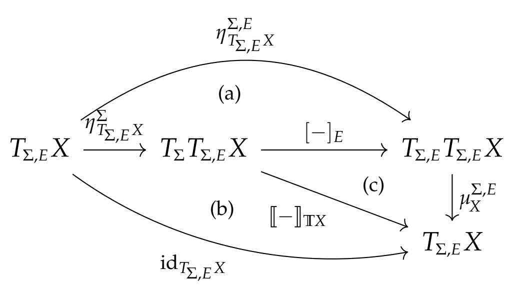
then by definition of $\equiv_E$, $X \vdash s = t \in E$ if and only if $[s]_E = [t]_E$.

The interaction between μ^Σ and η^Σ is mimicked by $\mu^{\Sigma,E}$ and $\eta^{\Sigma,E}$.

Lemma 31. *The following diagram commutes.*



Proof. For the triangle on the left, we pave the following diagram.



(29)

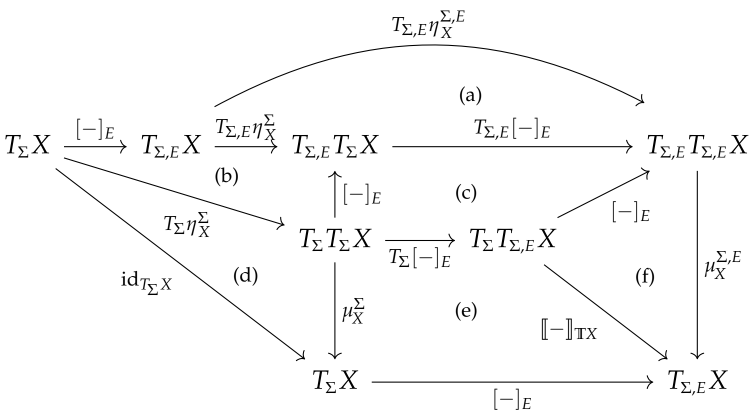
Showing (29) commutes:

(a) Definition of $\eta^{\Sigma,E}_X$.

(b) Definition of $[\![-]\!]_{\mathbb{T}X}$ (22).

(c) Definition of $\mu^{\Sigma,E}_X$ (23).

For the triangle on the right, we show that $[-]_E = \mu_X^{\Sigma,E} \circ T_{\Sigma,E}\eta_X^{\Sigma,E} \circ [-]_E$ by paving (30), and we can conclude since $[-]_E$ is surjective (or epic) that $\mathrm{id}_{T_{\Sigma,E}X} = \mu_X^{\Sigma,E} \circ T_{\Sigma,E}\eta_X^{\Sigma,E}$.



(30)

Showing (30) commutes:
(a) Definition of $\eta_X^{\Sigma,E}$ and functoriality of $T_{\Sigma,E}$.
(b) Naturality of $[-]_E$ (19).
(c) Naturality of $[-]_E$ again.
(d) Definition of μ_X^Σ (4).
(e) By (21).
(f) By (23).

□

We single out another special case of interpretation in a term algebra when E is empty (recall from Remark 24 that $\mathbb{T}_{\Sigma,\emptyset}X$ is the algebra on $T_\Sigma X$ whose interpretation of op applies op syntactically).

Definition 32 (Substitution)**.** Given a signature Σ, an empty set of equations, and an assignment $\sigma : Y \to T_\Sigma X$,[51] we call $[\![-]\!]^\sigma_{\mathbb{T}X}$ the **substitution** map, and we denote it by $\sigma^* : T_\Sigma Y \to T_\Sigma X$. We saw in Remark 24 that $[\![-]\!]_{\mathbb{T}X} = \mu_X^\Sigma$, thus substitution is

$$\sigma^* = T_\Sigma Y \xrightarrow{T_\Sigma \sigma} T_\Sigma T_\Sigma X \xrightarrow{\mu_X^\Sigma} T_\Sigma X. \tag{31}$$

In words, σ^* replaces the occurences of a variable y by $\sigma(y)$.[52]

[51] We can identify $T_\Sigma X$ with $T_{\Sigma,\emptyset}X$ because $\equiv_\emptyset$ is the equality relation.

[52] You may be more familiar with the notation $t[\sigma(y)/y]$ (e.g. from substitution in the λ-calculus). An inductive definition can also be given: for any $y \in Y$, $\sigma^*(\eta_Y^\Sigma(y)) = \sigma(y)$, and

$$\sigma^*(\mathsf{op}(t_1,\dots,t_n)) = \mathsf{op}(\sigma^*(t_1),\dots,\sigma^*(t_n)).$$

That simple description makes substitution a little special, and the following result has even deeper implications. It morally says that substitution preserves the satisfaction of equations.

Lemma 33. *Let $Y \vdash s = t$ be an equation, $\sigma : Y \to T_\Sigma X$ an assignment, and $\mathbb{A}$ a Σ-algebra. If $\mathbb{A}$ satisfies $Y \vdash s = t$, then it also satisfies $X \vdash \sigma^*(s) = \sigma^*(t)$.*

Proof. Let $\iota : X \to A$ be an assignment, we need to show $[\![\sigma^*(s)]\!]^\iota_A = [\![\sigma^*(t)]\!]^\iota_A$. Define the assignment $\iota_\sigma : Y \to A$ that sends $y \in Y$ to $[\![\sigma(y)]\!]^\iota_A$, we claim that $[\![-]\!]^{\iota_\sigma}_A = [\![\sigma^*(-)]\!]^\iota_A$. The lemma then follows because by hypothesis, $[\![s]\!]^{\iota_\sigma}_A = [\![t]\!]^{\iota_\sigma}_A$. The following derivation proves our claim.

$$\begin{aligned}
[\![-]\!]^{\iota_\sigma}_A &= [\![-]\!]_A \circ T_\Sigma(\iota_\sigma) && \text{by (7)}\\
&= [\![-]\!]_A \circ T_\Sigma([\![\sigma(-)]\!]^\iota_A) && \text{definition of } \iota_\sigma\\
&= [\![-]\!]_A \circ T_\Sigma\left([\![-]\!]_A \circ T_\Sigma\iota \circ \sigma\right) && \text{by (7)}\\
&= [\![-]\!]_A \circ T_\Sigma[\![-]\!]_A \circ T_\Sigma T_\Sigma\iota \circ T_\Sigma\sigma && \text{by Proposition 6}\\
&= [\![-]\!]_A \circ \mu^\Sigma_A \circ T_\Sigma T_\Sigma\iota \circ T_\Sigma\sigma && \text{by (11)}
\end{aligned}$$

$$
\begin{aligned}
&= [\![-]\!]_A \circ T_\Sigma \iota \circ \mu_Y^\Sigma \circ T_\Sigma \sigma && \text{by (5)}\\
&= [\![-]\!]_A \circ T_\Sigma \iota \circ \sigma^* && \text{by (31)}\\
&= [\![\sigma^*(-)]\!]_A^\iota . && \text{by (7)} \qquad \square
\end{aligned}
$$

We are finally ready to show that $\mathbb{T}_{\Sigma,E}A$ is a (Σ, E)-algebra.[53]

[53] All the work we have been doing finally pays off.

Proposition 34. *For any set A, the term algebra $\mathbb{T}_{\Sigma,E}A$ satisfies all the equations in E.*

Proof. Let $X \vdash s = t$ belong to E and $\iota : X \to T_{\Sigma,E}A$ be an assignment. We need to show that $[\![s]\!]^\iota_{\mathbb{T}A} = [\![t]\!]^\iota_{\mathbb{T}A}$. We factor ι into[54]

$$\iota = X \xrightarrow{\eta_X^{\Sigma,E}} T_{\Sigma,E}X \xrightarrow{T_{\Sigma,E}\iota} T_{\Sigma,E}T_{\Sigma,E}A \xrightarrow{\mu_A^{\Sigma,E}} T_{\Sigma,E}A.$$

[54] This factoring is correct because

$$
\begin{aligned}
\iota &= \mathrm{id}_{T_{\Sigma,E}A} \circ \iota\\
&= \mu_A^{\Sigma,E} \circ \eta^{\Sigma,E}_{T_{\Sigma,E}A} \circ \iota && \text{Lemma 31}\\
&= \mu_A^{\Sigma,E} \circ T_{\Sigma,E}\iota \circ \eta_X^{\Sigma,E}. && \text{naturality of } \eta^{\Sigma,E}
\end{aligned}
$$

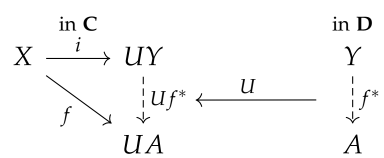
Now, Lemma 30 says that the equation is satisfied in $\mathbb{T}X$ under the assignment $\eta_X^{\Sigma,E}$, i.e. that $[\![s]\!]^{\eta_X^{\Sigma,E}}_{\mathbb{T}X} = [\![t]\!]^{\eta_X^{\Sigma,E}}_{\mathbb{T}X}$. We also know by Lemma 14 that homomorphisms preserve satisfaction, so we can apply it twice using the facts that $T_{\Sigma,E}\iota$ and $\mu_A^{\Sigma,E}$ are homomorphisms (by (25) and (28) respectively) to conclude that

$$[\![s]\!]^\iota_{\mathbb{T}A} = [\![s]\!]^{\mu_A^{\Sigma,E}\circ T_{\Sigma,E}\iota\circ\eta_X^{\Sigma,E}}_{\mathbb{T}A} = [\![t]\!]^{\mu_A^{\Sigma,E}\circ T_{\Sigma,E}\iota\circ\eta_X^{\Sigma,E}}_{\mathbb{T}A} = [\![t]\!]^\iota_{\mathbb{T}A}. \qquad \square$$

We now know that $\mathbb{T}_{\Sigma,E}X$ belongs to $\mathbf{Alg}(\Sigma, E)$, in order to tie up the parallel with Example 21, we will show that $\mathbb{T}_{\Sigma,E}X$ is the free (Σ, E)-algebra over X.

Definition 35 (Free object)**.** Let $\mathbf{C}$ and $\mathbf{D}$ be categories, $U : \mathbf{D} \to \mathbf{C}$ be a functor between them, and $X \in \mathbf{C}_0$. A **free object** on X (with respect to U) is an object $Y \in \mathbf{D}_0$ along with a morphism $i \in \mathrm{Hom}_{\mathbf{C}}(X, UY)$ such that for any object $A \in \mathbf{D}_0$ and morphism $f \in \mathrm{Hom}_{\mathbf{C}}(X, UA)$, there exists a unique morphism $f^* \in \mathrm{Hom}_{\mathbf{D}}(Y, A)$ such that $Uf^* \circ i = f$. This is summarized in the following diagram.[55]

[55] This is almost a copy of (16).



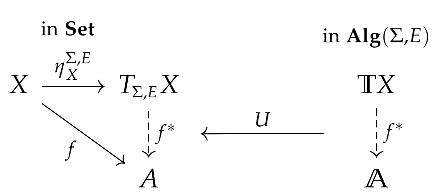
(32)

Proposition 36. *Free objects are unique up to isomorphism, namely, if Y and Y' are free objects on X, then $Y \cong Y'$.*[56]

[56] Very abstractly: a free object on X is the same thing as an initial object in the comma category $\Delta(X) \downarrow U$, and initial objects are unique up to isomorphism.

Proposition 37. *For any set X, the term algebra $\mathbb{T}X$ is the free (Σ, E)-algebra on X.*

Proof. Let $\mathbb{A}$ be another (Σ, E)-algebra and $f : X \to A$ a function. We claim that $f^* = [\![-]\!]_A \circ T_{\Sigma,E}f$ is the unique homomorphism making the following commute.



First, f^* is a homomorphism because it is the composite of two homomorphisms $T_{\Sigma,E}f$ (by (25)) and $[\![-]\!]_A$ (by Lemma 27 since $\mathbb{A}$ satisfies E). Next, the triangle commutes by the following derivation.

$$\begin{aligned} [\![-]\!]_A \circ T_{\Sigma,E}f \circ \eta_X^{\Sigma,E} &= [\![-]\!]_A \circ \eta_A^{\Sigma,E} \circ f && \text{naturality of } \eta^{\Sigma,E} \\ &= [\![-]\!]_A \circ [-]_E \circ \eta_A^{\Sigma} \circ f && \text{definition of } \eta^{\Sigma,E} \\ &= [\![-]\!]_A \circ \eta_A^{\Sigma} \circ f && \text{by (27)} \\ &= f && \text{definition of } [\![-]\!]_A \text{ (6)} \end{aligned}$$

Finally, uniqueness follows from the inductive definition of $\mathbb{T}X$ and the homomorphism property. Briefly, if we know the action of a homomorphism on equivalence classes of terms of depth 0, we can infer all of its action because all other classes of terms can be obtained by applying operation symbols.[57] □

[57] Formally, let $f, g : \mathbb{T}X \to \mathbb{A}$ be two homomorphisms such that for any $x \in X$, $f[x]_E = g[x]_E$, then, we can show that $f = g$. For any $t \in T_\Sigma X$, we showed in Lemma 29 that $[t]_E = [\![t]\!]_{\mathbb{T}X}^{\eta_X^{\Sigma,E}}$. Then using (9), we have

$$f[t]_E = [\![t]\!]_A^{f\circ\eta_X^{\Sigma,E}} = [\![t]\!]_A^{g\circ\eta_X^{\Sigma,E}} = g[t]_E,$$

where the second inequality follows by hypothesis that f and g agree on equivalence classes of terms of depth 0.

Once we have free objects, we have an adjunction, and once we have an adjunction, we have a monad, so we need to talk about monads. Unfortunately, our univeral algebra spiel is not finished yet, we will get back to monads shortly.

1.3 Equational Logic

We were happy that interpretations in the term algebra are computed syntactically, but there is a big caveat. Evertything is done modulo $\equiv_E$ which was defined in (17) to morally contain all equations in $\mathfrak{Th}(E)$, that is, all equations semantically implied by E. Equational logic is a deductive system that allows to derive syntactically all of $\mathfrak{Th}(E)$ starting from E.

In Lemma 22, we proved that $\equiv_E$ is a congruence (i.e. reflexive, symmetric, transitive, and invariant under operations), and in Lemma 33 we showed $\equiv_E$ is also preserved by substitutions. This can help us syntactically derive $\mathfrak{Th}(E)$ because, for instance, if we know $X \vdash s = t \in E$, we can conclude $X \vdash t = s \in \mathfrak{Th}(E)$ by symmetry. Then, by transitivity, we can conclude that $X \vdash s = s \in \mathfrak{Th}(E)$, which we already knew by reflexivity. This can be summarized with the inference rules of **equational logic** in Figure 1.2.

$$\frac{}{X \vdash t = t}\,\textsc{Refl} \qquad \frac{X \vdash s = t}{X \vdash t = s}\,\textsc{Symm} \qquad \frac{X \vdash s = t \qquad X \vdash t = u}{X \vdash s = u}\,\textsc{Trans}$$

$$\frac{\mathsf{op} : n \in \Sigma \qquad \forall 1 \leq i \leq n, X \vdash s_i = t_i}{X \vdash \mathsf{op}(s_1, \dots, s_n) = \mathsf{op}(t_1, \dots, t_n)}\,\textsc{Cong}$$

$$\frac{\sigma : X \to T_\Sigma Y \qquad X \vdash s = t}{Y \vdash \sigma^*(s) = \sigma^*(t)}\,\textsc{Sub}$$

Figure 1.2: Rules of equational logic over the signature Σ, where X and Y can be any set, and s, t, u, s_i and t_i can be any terms in $T_\Sigma X$. As indicated in the premises of the rules Cong and Sub, they can be instantiated for any n-ary operation symbol and for any function σ respectively.

Definition 38 (Derivation)**.** A **derivation**[58] of $X \vdash s = t$ in equational logic with axioms E (a set of equations) is a finite rooted tree such that:

- all nodes are labelled by equations,
- the root is labelled by $X \vdash s = t$,
- when an internal node (not a leaf) is labelled by ϕ and its children are labelled by $\phi_1, \dots, \phi_n$, there is a inference rule in Figure 1.2 which concludes ϕ from $\phi_1, \dots, \phi_n$, and
- all the leaves are either in E or instances of REFL, i.e. an equation $Y \vdash u = u$ for some set Y and $u \in T_\Sigma Y$.

[58] Many other definitions of derivation exist, and our treatment of them will not be 100% rigorous.

Example 39. We write a derivation with the same notation used to specify the inference rules in Figure 1.2. Consider the signature $\Sigma = \{+:2, \mathsf{e}:0\}$ with E containing the equations defining commutative monoids in (14). Here is a derivation of $x, y, z \vdash x + (y + z) = z + (x + y)$ in equational logic with axioms E.

$$\dfrac{\dfrac{}{x,y,z \vdash x+(y+z) = (x+y)+z}\in E \qquad \dfrac{\sigma = \begin{matrix} x \mapsto x+y \\ y \mapsto z \end{matrix} \qquad \dfrac{}{x,y \vdash x+y = y+x}\in E}{x,y,z \vdash (x+y)+z = z+(x+y)}\,\textsc{Sub}}{x,y,z \vdash x+(y+z) = z+(x+y)}\,\textsc{Trans}$$

Given any set of equations E, we denote by $\mathfrak{Th}'(E)$ the class of equations that can be proven from E in equational logic, i.e. $\phi \in \mathfrak{Th}'(E)$ if and only if there is a derivation of ϕ in equational logic with axioms E.

Our goal for the rest of this section is to prove that $\mathfrak{Th}'(E) = \mathfrak{Th}(E)$. We say that equational logic is sound and complete for (Σ, E)-algebras. Less concisely, soundness means that whenever equational logic proves an equation ϕ with axioms E, then ϕ is satisfied by all (Σ, E)-algebras, and completeness says that whenever an equation ϕ is satisfied by all (Σ, E)-algebras, then there is a derivation of ϕ in equational logic with axioms E.

Soundness is a straightforward consequence of earlier results.[59]

[59] In the story we are telling here, the rules of equational logic were designed to be sound because we knew some properties of $\equiv_E$ already. In general when defining rules of a logic, we may use intuitions and later prove soundness to confirm them, or realize that soundness does not hold and infirm them.

Theorem 40 (Soundness)**.** *If* $\phi \in \mathfrak{Th}'(E)$*, then* $\phi \in \mathfrak{Th}(E)$.

Proof. In the proof of Lemma 22, we proved that each of REFL, SYMM, TRANS, and CONG are sound rules for a fixed arbitrary algebra. Namely, if $\mathbb{A} \in \mathbf{Alg}(\Sigma)$ satisfies the equations on top, then it satisfies the one on the bottom. Lemma 33 states the same soundness property for SUB. This implies a weaker property: if all (Σ, E)-algebras satisfy the equations on top, then they satisfy the one on the bottom.[60]

[60] This is a classical theorem of first order logic:

$$(\forall A.(PA \Rightarrow QA)) \Rightarrow (\forall A.PA \Rightarrow \forall A.QA)$$

Now, if $\phi \in \mathfrak{Th}'(E)$ was proven using equational logic and the axioms in E, then since all $\mathbb{A} \in \mathbf{Alg}(\Sigma, E)$ satisfy all the axioms, by repeatedly applying the weaker property above for each rule in the derivation, we find that all $\mathbb{A} \in \mathbf{Alg}(\Sigma, E)$ satisfy ϕ, i.e. $\phi \in \mathfrak{Th}(E)$. □

Completeness is a wilder beast we need to tame. The more classical proofs rely on a theory of congruences. Our method is based on uniqueness free algebras

(Proposition 36). We will define an algebra exactly like $\mathbb{T}A$ but using the equality relation induced by $\mathfrak{Th}'(E)$ instead $\equiv_E$ which is induced by $\mathfrak{Th}(E)$. We then show that algebra is the free (Σ, E)-algebra and conclude that $\mathfrak{Th}(E)$ and $\mathfrak{Th}'(E)$ must coincide (this proves soundness again).

Fix a signature Σ and a set E of equations over Σ. For any set X, we can define a binary relation $\equiv'_E$ on Σ-terms[61] that contains the pair (s,t) whenever $X \vdash s = t$ can be proven in equational logic. Formally, we have for any $s, t \in T_\Sigma X$,

[61] Again, we omit the set X from the notation.

$$s \equiv'_E t \iff X \vdash s = t \in \mathfrak{Th}'(E). \tag{33}$$

We can show $\equiv'_E$ is a congruence relation.

Lemma 41. *For any set X, the relation $\equiv'_E$ is reflexive, symmetric, transitive, and for any $\mathsf{op} : n \in \Sigma$ and $s_1, \dots, s_n, t_1, \dots, t_n \in T_\Sigma X$,*

$$\forall 1 \leq i \leq n, s_i \equiv'_E t_i \implies \mathsf{op}(s_1, \dots, s_n) \equiv'_E \mathsf{op}(t_1, \dots, t_n). \tag{34}$$

Proof. This is immediate from the presence of Refl, Symm, Trans, and Cong in the rules of equational logic. $\square$

We write $\wr - \int_E : T_\Sigma X \to T_\Sigma X/{\equiv'_E}$ for the canonical quotient map, so $\wr t \int_E$ is the equivalence class of t modulo the congruence $\equiv'_E$ induced by equational logic.

Definition 42 (Term algebra, syntactically)**.** The new term algebra for (Σ, E) on X is the Σ-algebra whose carrier is $T_\Sigma X/{\equiv'_E}$ and whose interpretation of $\mathsf{op} : n \in \Sigma$ is defined by[62]

$$[\![\mathsf{op}]\!]_{\mathbb{T}'X}(\wr t_1 \int_E, \dots, \wr t_n \int_E) = \wr \mathsf{op}(t_1, \dots, t_n) \int_E. \tag{35}$$

[62] This is well-defined (i.e. invariant under change of representative) by (34).

We denote this algebra by $\mathbb{T}'_{\Sigma,E}X$ or simply $\mathbb{T}'X$.

We will prove that this alternative definition of the term algebra coincides with $\mathbb{T}X$ because both are the free (Σ, E)-algebra on X. First, we have to show that $\mathbb{T}'X$ belongs to $\mathbf{Alg}(\Sigma, E)$ like we did for $\mathbb{T}X$ in Proposition 34, and we prove a technical lemma before that.

Lemma 43. *Let $\iota : Y \to T_\Sigma X/{\equiv'_E}$ be an assignment. For any function $\sigma : Y \to T_\Sigma X$ satisfying $\wr \sigma(y) \int_E = \iota(y)$ for all $y \in Y$, we have $[\![-]\!]^\iota_{\mathbb{T}'X} = \wr \sigma^*(-) \int_E$.*

Proof. We proceed by induction. For the base case, we have by definition of the interpretation of terms (6), definition of σ, and definition of σ^* (31),

$$[\![\eta^\Sigma_Y(y)]\!]^\iota_{\mathbb{T}'X} \stackrel{(6)}{=} \iota(y) = \wr \sigma(y) \int_E \stackrel{(31)}{=} \wr \sigma^*(\eta^\Sigma_Y(y)) \int_E.$$

For the inductive step, we have

$$\begin{aligned}
[\![\mathsf{op}(t_1, \dots, t_n)]\!]^\iota_{\mathbb{T}'X} &= [\![\mathsf{op}]\!]_{\mathbb{T}'X}([\![t_1]\!]^\iota_{\mathbb{T}'X}, \dots, [\![t_n]\!]^\iota_{\mathbb{T}'X}) && \text{by (6)} \\
&= [\![\mathsf{op}]\!]_{\mathbb{T}'X}(\wr \sigma^*(t_1) \int_E, \dots, \wr \sigma^*(t_n) \int_E) && \text{I.H.} \\
&= \wr \mathsf{op}(\sigma^*(t_1), \dots, \sigma^*(t_n)) \int_E && \text{by (35)} \\
&= \wr \sigma^*(\mathsf{op}(t_1, \dots, t_n)) \int_E. && \text{definition of } \sigma^*
\end{aligned}$$

$\square$

Proposition 44. *For any set X, $\mathbb{T}'X$ satisfies all the equations in E.*

Proof. Let $Y \vdash s = t$ belong to E and $\iota : Y \to T_\Sigma X/{\equiv'_E}$ be an assignment. By the axiom of choice,[63] there is a function $\sigma : Y \to T_\Sigma X$ satisfying $\wr \sigma(y) \int_E = \iota(y)$ for all $y \in Y$. Thanks to Lemma 43, it is enough to show $\wr \sigma^*(s) \int_E = \wr \sigma^*(t) \int_E$.[64] Equivalently, by definition of $\wr - \int_E$ and $\mathfrak{Th}'(E)$, we can exhibit a derivation of $X \vdash \sigma^*(s) = \sigma^*(t)$ in equational logic with axioms E. This is rather simple because that equation can be proven with the Sub rule instantiated with $\sigma^* : Y \to T_\Sigma X$ and the equation $Y \vdash s = t$ which is an axiom. □

[63] Choice implies the quotient map $\wr - \int_E$ has a left inverse $r : T_\Sigma X/{\equiv'_E} \to T_\Sigma X$, and we can then set $\sigma = r \circ \iota$.

[64] By Lemma 43, it implies

$$[\![s]\!]^\iota_{\mathbb{T}'X} = \wr \sigma^*(s) \int_E = \wr \sigma^*(t) \int_E = [\![t]\!]^\iota_{\mathbb{T}'X},$$

and since ι was an arbitrary assignment, we conclude that $\mathbb{T}'X \vDash Y \vdash s = t$.

Completeness of equational logic readily follows.

Theorem 45 (Completeness)**.** *If $\phi \in \mathfrak{Th}(E)$, then $\phi \in \mathfrak{Th}'(E)$.*

Proof. Write $\phi = X \vdash s = t \in \mathfrak{Th}(E)$. By Proposition 44 and definition of $\mathfrak{Th}(E)$, we know that $\mathbb{T}'X \vDash \phi$. In particular, $\mathbb{T}'X$ satisfies ϕ under the assignment

$$\iota = X \xrightarrow{\eta^\Sigma_X} T_\Sigma X \xrightarrow{\wr - \int_E} T_\Sigma X/{\equiv'_E},$$

namely, $[\![s]\!]^\iota_{\mathbb{T}'X} = [\![t]\!]^\iota_{\mathbb{T}'X}$. Moreover with $\sigma = \eta^\Sigma_X$, we can show σ satisfies the hypothesis of Lemma 43 and $\sigma^* = \mathrm{id}_{T_\Sigma X}$,[65] thus we conclude

$$\wr s \int_E = [\![s]\!]^\iota_{\mathbb{T}'X} = [\![t]\!]^\iota_{\mathbb{T}'X} = \wr t \int_E.$$

By definition of $\wr - \int_E$, this implies $s \equiv'_E t$ which in turn means $X \vdash s = t$ belongs to $\mathfrak{Th}'(E)$. □

[65] We defined ι precisely to have $\wr \sigma(x) \int_E = \iota(x)$. To show $\sigma^* = {\eta^\Sigma_X}^*$ is the identity, use (31) and the fact that $\mu^\Sigma \cdot \eta^\Sigma T_\Sigma = \mathbb{1}_{T_\Sigma}$ (it holds by definition (4)).

Note that because $\mathbb{T}X$ and $\mathbb{T}'X$ were defined in the same way in terms of $\mathfrak{Th}(E)$ and $\mathfrak{Th}'(E)$ respectively, and since we have proven the latter to be equal, we obtain that $\mathbb{T}X$ and $\mathbb{T}'X$ are the same algebra. In the sequel, we will work with $\mathbb{T}X$ mostly but we may use the fact that $s \equiv_E t$ if and only if there is a derivation of $X \vdash s = t$ in equational logic.

Remark 46. We have used the axiom of choice twice in proving completeness of equational logic. That is only an artifact of our presentation that deals with arbitrary contexts. Since terms are finite and operation symbols have finite arities, it is possible make do with only finite contexts (which removes the need for choice). Formally, one can prove by induction on the derivation that a proof of $X \vdash s = t$ can be transformed into a proof of $\mathrm{FV}\{s,t\} \vdash s = t$ which uses only equations with finite contexts.[66]

[66] We denoted by $\mathrm{FV}\{s,t\}$ the set of **free variables** used in s and t. This can be defined inductively as follows:

$$\mathrm{FV}\{\eta^\Sigma_X(x)\} = \{x\}$$
$$\mathrm{FV}\{\mathsf{op}(t_1,\dots,t_n)\} = \mathrm{FV}\{t_1\} \cup \cdots \cup \mathrm{FV}\{t_n\}$$
$$\mathrm{FV}\{t_1,\dots,t_n\} = \mathrm{FV}\{t_1\} \cup \cdots \cup \mathrm{FV}\{t_n\}.$$

Note that $\mathrm{FV}\{-\}$ applied to a finite set of terms is always finite.

We mention now two related results for the sake of comparison when we introduce quantitative equational logic. For any set X and variable y, the following rules are derivable in equational logic.

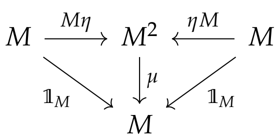
$$\frac{X \vdash s = t}{X \cup \{y\} \vdash s = t}\ \textsc{Add} \qquad \frac{X \vdash s = t \qquad y \notin \mathrm{FV}\{s,t\}}{X \setminus \{y\} \vdash s = t}\ \textsc{Del}$$

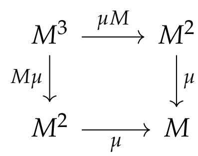
In words, Add says that you can always add a variable to the context, and Del says you can remove a variable from the context when it is not used in the terms of the equations. Both these rules are instances of SUB. For the first, take σ to be the inclusion of X in $X \cup \{y\}$ (it may be the identity if $y \in X$). For the second, let σ send y to whatever element of $X \setminus \{y\}$ and all the other elements of X to themselves[67], then since y is not in the free variables of s and t, $\sigma^*(s) = s$ and $\sigma^*(t) = t$.

[67] When X is empty, the equations on the top and bottom of DEL coincide, so the rule is clearly derivable.

1.4 Monads

Definition 47 (Monad)**.** A **monad** on a category $\mathbf{C}$ is a triple (M, η, μ) comprised of an endofunctor $M : \mathbf{C} \to \mathbf{C}$ and two natural transformations $\eta : \mathrm{id}_{\mathbf{C}} \Rightarrow M$ and $\mu : M^2 \Rightarrow M$ called the **unit** and **multiplication** respectively that make (36) and (37) commute in $[\mathbf{C}, \mathbf{C}]$.[68]


(36)


(37)

[68] In equations, ie means for any object $A \in \mathbf{C}_0$, $\mu_A \circ M\eta_A = \mathrm{id}_A$, $\mu_A \circ \eta_{MA} = \mathrm{id}_A$, and $\mu_A \circ \mu_{MA} = \mu_A \circ M\mu_A$.

In this chapter we will mostly talk about monads on **Set**, but it is good to keep some arguments general for later. Here are some very important examples (for the literature and especially for this manuscript).

Example 48 (Maybe)**.** Suppose $\mathbf{C}$ has (binary) coproducts and a terminal object $\mathbf{1}$, then $(- + \mathbf{1}) : \mathbf{C} \to \mathbf{C}$ is a monad. It is called the **maybe monad**. We write inl^{X+Y} (resp. inr^{X+Y}) for the coprojection of X (resp. Y) into $X + Y$.[69] First, note that for a morphism $f : X \to Y$,

$$f + \mathbf{1} = [\mathsf{inl}^{Y+\mathbf{1}} \circ f, \mathsf{inr}^{Y+\mathbf{1}}] : X + \mathbf{1} \to Y + \mathbf{1}.$$

[69] These notations are very common in the community of programming language research, they stand for *injection left* (resp. *right*). We may omit the superscript in case it is too cumbersome.

The components of the unit are given by the coprojections, i.e. $\eta_X = \mathsf{inl}^{X+\mathbf{1}} : X \to X + \mathbf{1}$, and the components of the multiplication are

$$\mu_X = [\mathsf{inl}^{X+\mathbf{1}}, \mathsf{inr}^{X+\mathbf{1}}, \mathsf{inr}^{X+\mathbf{1}}] : X + \mathbf{1} + \mathbf{1} \to X + \mathbf{1}.$$

Checking that (36) and (37) commute is an exercise in reasoning with coproducts. It is much more interesting to give the intuition in **Set** where $+$ is the disjoint union and $\mathbf{1}$ is the singleton $\{*\}$:[70]

[70] This intuition should carry over well to many categories where the coproduct and terminal objects have similar behaviors.

- $X + \mathbf{1}$ is the set X with an additional (fresh) element $*$,
- the function $f + \mathbf{1}$ acts like f on X and sends the new element $* \in X$ to the new element $* \in Y$,
- the unit $\eta_X : X \to X + \mathbf{1}$ is the injection (sending $x \in X$ to itself),
- the multiplication μ_X acts like the identity on X and sends the two new elements of $X + \mathbf{1} + \mathbf{1}$ to the single new element $X + \mathbf{1}$,
- one can check (36) and (37) commute by hand because (briefly) $x \in X$ is always sent to $x \in X$ and $*$ is always sent to $*$.

Example 49 (Powerset)**.** The covariant **non-empty finite powerset** functor $\mathcal{P}_{\text{ne}}$: **Set** $\to$ **Set** sends a set X to the set of non-empty finite subsets of X which we denote by $\mathcal{P}_{\text{ne}}X$. It acts on functions just like the usual powerset functor, i.e. given a function $f : X \to Y$, $\mathcal{P}_{\text{ne}}f$ is the direct image function, it sends $S \subseteq X$ to $f(S) = \{f(x) \mid x \in S\}$.[71]

One can show $\mathcal{P}_{\text{ne}}$ is a monad with the following unit and multiplication:

$$\eta_X : X \to \mathcal{P}_{\text{ne}}(X) = x \mapsto \{x\} \text{ and } \mu_X : \mathcal{P}_{\text{ne}}(\mathcal{P}_{\text{ne}}(X)) \to \mathcal{P}_{\text{ne}}(X) = F \mapsto \bigcup_{s \in F} s.$$

[71] It is clear that $f(S)$ is non-empty and finite when S is non-empty and finite.

Example 50 (Distributions)**.** The functor $\mathcal{D}$: **Set** $\to$ **Set** sends a set X to the set of **finitely supported distributions** on X:[72]

$$\mathcal{D}(X) := \{\varphi : X \to [0,1] \mid \sum_{x \in X} \varphi(x) = 1 \text{ and } \varphi(x) \neq 0 \text{ for finitely many } x\text{'s}\}.$$

[72] We will simply call them distributions.

We call $\varphi(x)$ the **weight** of φ at x, and let $\text{supp}(\varphi)$ denote the **support** of φ, that is, $\text{supp}(\varphi)$ contains all the elements $x \in X$ such that $\varphi(x) \neq 0$. On morphisms, $\mathcal{D}$ sends a function $f : X \to Y$ to the function between sets of distributions defined by

$$\mathcal{D}f : \mathcal{D}X \to \mathcal{D}Y = \varphi \mapsto \left(y \mapsto \sum_{x \in X, f(x) = y} \varphi(x) \right).$$

In words, the weight of $\mathcal{D}f(\varphi)$ at y is equal to the total weight of φ on the preimage of y under f.

One can show that $\mathcal{D}$ is a monad with unit $\eta_X = x \mapsto \delta_x$, where δ_x is the **Dirac** distribution at x (the weight of δ_x is 1 at x and 0 everywhere else), and multiplication

$$\mu_X = \Phi \mapsto \left(x \mapsto \sum_{\varphi \in \text{supp}(\Phi)} \Phi(\varphi)\varphi(x) \right).$$

In words, the weight $\mu_X(\Phi)$ at x is the average weight at x of distributions in the support of Φ.

Monads have historically been the prevailing categorical approach to universal algebra.[73] This is due to a result of Linton [Lin66] stating that any algebraic theory gives rise to a monad. Given a signature Σ and a set E of equations, the monad Linton constructed is $T_{\Sigma,E}$.

[73] Although this has been changing, in part due to [HP07] (and the articles leading to that paper, e.g. [PP01, HPP06]) where the authors argue for using Lawvere theories instead.

Proposition 51. *The functor $T_{\Sigma,E}$:* **Set** $\to$ **Set** *defines a monad on* **Set** *with unit $\eta^{\Sigma,E}$ and multiplication $\mu^{\Sigma,E}$. We call it the **term monad** for (Σ, E).*

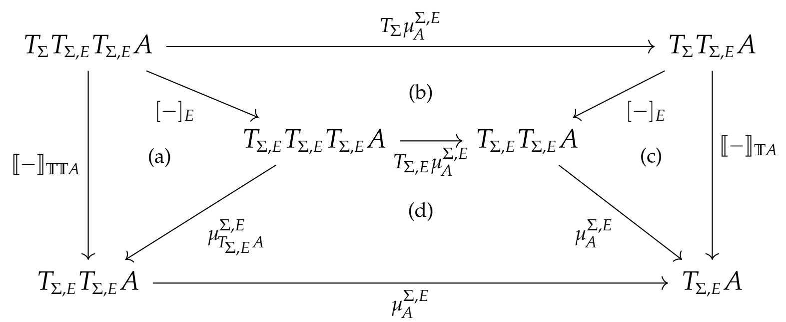
Proof. We have done most of the work already.[74] We showed that $\eta^{\Sigma,E}$ and $\mu^{\Sigma,E}$ are natural transformations of the right type in Footnote 49 and Proposition 26 respectively, and we showed the appropriate instance of (36) commutes in Lemma 31. It remains to prove (37) commutes which, instantiated here, means proving the fol-

[74] In fact, we have done it twice because we showed that $\mathbb{T}_{\Sigma,E}A$ is the free (Σ, E)-algebra on A for every set A, and that automatically yields (through abstract categorical arguments) a monad sending A to the carrier of $\mathbb{T}_{\Sigma,E}A$, i.e. $T_{\Sigma,E}A$.

lowing diagram commutes for every set A.

$$\begin{array}{ccc}
T_{\Sigma,E}T_{\Sigma,E}T_{\Sigma,E}A & \xrightarrow{T_{\Sigma,E}\mu^{\Sigma,E}_A} & T_{\Sigma,E}T_{\Sigma,E}A \\
\mu^{\Sigma,E}_{T_{\Sigma,E}A}\downarrow & & \downarrow\mu^{\Sigma,E}_A \\
T_{\Sigma,E}T_{\Sigma,E}A & \xrightarrow[\mu^{\Sigma,E}_A]{} & T_{\Sigma,E}A
\end{array}$$

It follows from the following paved diagram.[75]



[75] We know that (a), (b) and (c) commute by (23), (19), and (23) respectively. This means that (d) precomposed by the epimorphism $[-]_E$ yields the outer square. Moreover, we know the outer square commutes by (28), therefore, (d) must also commute.

Note that when E is empty, we get a monad $(T_\Sigma, \eta^\Sigma, \mu^\Sigma)$.[76] □

[76] Here is an alternative proof that T_Σ is a monad. We showed η^Σ and μ^Σ are natural in (3) and (5) respectively. The right triangle of (36) commutes by definition of μ^Σ (4), the left triangle commutes by Lemma 9, and the square (37) commutes by (12).

It makes sense now to ask to go in the other direction, namely, given a monad, how do we obtain a signature and a set of equations? First, just like (Σ, E)-algebras are models of the theory (Σ, E), we can define models for a monad, which we also call algebras.

Definition 52 (M–algebra)**.** Let (M, η, μ) be a monad on $\mathbf{C}$, an M**-algebra** is a pair (A, α) comprising an object $A \in \mathbf{C}_0$ and a morphism $\alpha : MA \to A$ such that (38) and (39) commute.

$$\begin{array}{ccc}
A & \xrightarrow{\eta_A} & MA \\
& \searrow^{\mathrm{id}_A} & \downarrow\alpha \\
& & A
\end{array} \qquad (38) \qquad\qquad
\begin{array}{ccc}
M^2A & \xrightarrow{\mu_A} & MA \\
M\alpha\downarrow & & \downarrow\alpha \\
MA & \xrightarrow[\alpha]{} & A
\end{array} \qquad (39)$$

We call A the carrier and we may write only α to refer to an M-algebra.

Definition 53 (Homomorphism)**.** Let (M, η, μ) be a monad and (A, α) and (B, β) be two M-algebras. An M-algebra **homomorphism** or simply M-homomorphism from α to β is a morphism $h : A \to B$ in $\mathbf{C}$ making (40) commute.

$$\begin{array}{ccc}
MA & \xrightarrow{Mh} & MB \\
\alpha\downarrow & & \downarrow\beta \\
A & \xrightarrow[h]{} & B
\end{array} \qquad (40)$$

The composition of two M-homomorphisms is an M-homomorphism and id_A is an M-homomorphism from (A, α) to itself whenever α is an M-algebra, thus we get a category of M-algebras and M-homomorphisms called the **Eilenberg–Moore category** of M and denoted by $\mathbf{EM}(M)$.[77]

[77] Named after the authors of the article introducing that category [EM65].

Since $\mathbf{EM}(M)$ was built from objects and morphisms in $\mathbf{C}$, there is an obvious forgetful functor $U^M : \mathbf{EM}(M) \to \mathbf{C}$ sending an M-algebra (A, α) to its carrier A and an M-homomorphism to its underlying morphism.

The terminology suggests that (Σ, E)-algebras and $T_{\Sigma,E}$-algebras are the same thing. Let us check this.

Proposition 54. *There is an isomorphism* $\mathbf{Alg}(\Sigma, E) \cong T_{\Sigma,E}$.

Proof. ... □

What about algebras for other monads?

Example 55 (Maybe)**.** In **Set**, an $(- + \mathbf{1})$-algebra is a function $\alpha : A + \mathbf{1} \to A$ making the following diagrams commute.

$$\begin{array}{ccc} A & \xrightarrow{\eta_A} & A + \mathbf{1} \\ & \searrow^{\mathrm{id}_A} & \downarrow{\scriptstyle \alpha} \\ & & A \end{array} \qquad\qquad \begin{array}{ccc} A + \mathbf{1} + \mathbf{1} & \xrightarrow{\mu_A} & A + \mathbf{1} \\ {\scriptstyle \alpha+\mathbf{1}}\downarrow & & \downarrow{\scriptstyle \alpha} \\ A + \mathbf{1} & \xrightarrow[\alpha]{} & A \end{array}$$

Reminding ourselves that η_A is the inclusion in the left component, the triangle commuting enforces α to act like the identity function on all of A. We can also write $\alpha = [\mathrm{id}_A, \alpha(*)]$.[78] The square commuting ads no additional constraint. Thus, an algebra for the maybe monad on **Set** is just a set with a distinguished point. Let $h : A \to B$ be a function, commutativity of (41) is equivalent to $h(\alpha(*)) = \beta(*)$. Hence, a $(- + \mathbf{1})$-homomorphism is a function that preserves the distinguished point.

[78] We identify the element $\alpha(*) \in A$ with the function $\alpha(*) : \mathbf{1} \to A$ picking out that element.

$$\begin{array}{ccc} A + \mathbf{1} & \xrightarrow{h+\mathbf{1}} & B + \mathbf{1} \\ {\scriptstyle [\mathrm{id}_A, \alpha(*)]}\downarrow & & \downarrow{\scriptstyle [\mathrm{id}_B, \beta(*)]} \\ A & \xrightarrow[h]{} & B \end{array} \tag{41}$$

Seeing the distinguished point of a $(- + \mathbf{1})$-algebra as the interpretation of a constant, we recognize that the category $\mathbf{EM}(- + \mathbf{1})$ is isomorphic to the category $\mathbf{Alg}(\Sigma)$ where $\Sigma = \{\mathsf{p} : 0\}$ contains a single constant.

An other option to recognize $\mathbf{EM}(- + \mathbf{1})$ as a category of algebras is via monad isomorphisms.

Definition 56 (Monad morphism)**.** Let (M, η^M, μ^M) and (N, η^N, μ^N) be two monads on $\mathbf{C}$. A **monad morphism** from M to N is a natural transformation $\rho : M \Rightarrow N$ making (42) and (43) commute.[79]

$$\begin{array}{ccc} \mathrm{id}_{\mathbf{C}} & & \\ {\scriptstyle \eta^M}\downarrow & \searrow^{\eta^N} & \\ M & \xrightarrow[\rho]{} & N \end{array} \tag{42}$$

$$\begin{array}{ccc} MM & \xrightarrow{\rho \diamond \rho} & NN \\ {\scriptstyle \mu^M}\downarrow & & \downarrow{\scriptstyle \mu^N} \\ M & \xrightarrow[\rho]{} & N \end{array} \tag{43}$$

[79] Recall that $\rho \diamond \rho$ denotes the horizontal composition of ρ with itself, i.e.

$$\rho \diamond \rho = \rho N \cdot M\rho = N\rho \cdot \rho M.$$

As expected ρ is called a monad isomorphism when there is a monad morphism $\rho^{-1} : N \Rightarrow M$ satisfying $\rho \cdot \rho^{-1} = \mathbb{1}_N$ and $\rho^{-1} \cdot \rho = \mathbb{1}_M$. In fact, it is enough that all the components of ρ are isomorphisms in $\mathbf{C}$ to guarantee ρ is a monad isomorphism.[80]

[80] One checks that natural isomorphisms are precisely the natural transformations whose components are all isomorphisms, and that the inverse of a monad morphism is automatically a monad morphism.

Example 57. For the signature $\Sigma = \{\mathsf{p} : 0\}$, the term monad T_Σ is isomorphic to $- + \mathbf{1}$. Indeed, recall that a Σ-term over A is either an element of A or p, this yields a bijection $\rho_A : A + \mathbf{1} \to T_\Sigma A$ that sends any $a \in A$ to itself and $*$ to the constant $\mathsf{p} \in T_\Sigma A$. To verify that ρ is a monad morphism, we check these diagrams commute.[81]

[81] All of them commute essentially because ρ_A acts like the identity on A.

$$\begin{array}{ccc} A+\mathbf{1} & \xrightarrow{\rho_A} & T_\Sigma A \\ {\scriptstyle f+\mathbf{1}}\downarrow & & \downarrow{\scriptstyle T_\Sigma f} \\ B+\mathbf{1} & \xrightarrow[\rho_B]{} & T_\Sigma B \end{array} \quad (44) \qquad \begin{array}{ccc} A & & \\ {\scriptstyle \eta_A}\downarrow & \searrow^{\eta^\Sigma_A} & \\ A+\mathbf{1} & \xrightarrow[\rho_A]{} & T_\Sigma A \end{array} \quad (45) \qquad \begin{array}{ccc} A+\mathbf{1}+\mathbf{1} & \xrightarrow{\rho_{T_\Sigma A}\circ(\rho_A+\mathbf{1})} & T_\Sigma T_\Sigma A \\ {\scriptstyle \mu_A}\downarrow & & \downarrow{\scriptstyle \mu^\Sigma_A} \\ A+\mathbf{1} & \xrightarrow[\rho_A]{} & T_\Sigma A \end{array} \quad (46)$$

We obtained a monad isomorphism between the maybe monad and the term monad for the signature Σ with only a constant. We can recover the isomorphism between the categories of algebras $\mathbf{EM}(-+\mathbf{1})$ and $\mathbf{Alg}(\Sigma)$ from Example 55 with the following result.

Proposition 58. *If $\rho : M \Rightarrow N$ is a monad morphism, then there is a functor $-\rho : \mathbf{EM}(N) \to \mathbf{EM}(M)$. If ρ is a monad isomorphism, then $-\rho$ is also an isomorphism.*

Proof. Given an N-algebra $\alpha : NA \to A$, we show that $\alpha \circ \rho_A : MA \to A$ is an M-algebra by paving the following diagrams.

$$\begin{array}{ccc} A & \xrightarrow{\eta^M_A} & MA \\ & \searrow^{\eta^N_A}\ \text{(a)} & \downarrow{\scriptstyle \rho_A} \\ & \text{(b)} & NA \\ & \searrow_{\mathrm{id}_A} & \downarrow{\scriptstyle \alpha} \\ & & A \end{array} \qquad \begin{array}{ccccc} MMA & & \xrightarrow{\mu^M_A} & & MA \\ {\scriptstyle M\rho_A}\downarrow & & \text{(c)} & & \downarrow{\scriptstyle \rho_A} \\ MNA & \xrightarrow{\rho_{NA}} & NNA & \xrightarrow{\mu^N_A} & NA \\ {\scriptstyle M\alpha}\downarrow & \text{(d)} & {\scriptstyle N\alpha}\downarrow & \text{(e)} & \downarrow{\scriptstyle \alpha} \\ MA & \xrightarrow[\rho_A]{} & NA & \xrightarrow[\alpha]{} & A \end{array} \quad (47)$$

Showing (47) commutes:

(a) By (42).

(b) By (38) for $\alpha : NA \to A$.

(c) By (43), noting that $(\rho \diamond \rho)_A = \rho_{NA} \circ M\rho_A$.

(d) Naturality of ρ.

(e) By (39) for $\alpha : NA \to A$.

Moreover, if $h : A \to B$ is an N-homomorphism from α to β, then it is also a M-homomorphism from $\alpha \circ \rho_A$ to $\beta \circ \rho_B$ by the paving below.[82]

$$\begin{array}{ccc} MA & \xrightarrow{Mh} & MB \\ {\scriptstyle \rho_A}\downarrow & & \downarrow{\scriptstyle \rho_B} \\ NA & \xrightarrow[Nh]{} & NB \\ {\scriptstyle \alpha}\downarrow & & \downarrow{\scriptstyle \beta} \\ A & \xrightarrow[h]{} & B \end{array}$$

We obtain a functor $-\rho : \mathbf{EM}(N) \to \mathbf{EM}(M)$ taking an algebra (A, α) to $(A, \alpha \circ \rho_A)$ and a homomorphism $h : (A, \alpha) \to (B, \beta)$ to $h : (A, \alpha \circ \rho_A) \to (B, \beta \circ \rho_B)$.

Furthermore, it is easy to see that $-\rho = \mathrm{id}_{\mathbf{EM}(M)}$ when $\rho = \mathbb{1}_M$ is the identity monad morphism, and that for any other monad morphism $\rho' : N \Rightarrow L$, $-(\rho' \cdot \rho) = (-\rho) \circ (-\rho')$.[83] Thus, when ρ is a monad isomorphism with inverse ρ^{-1}, $-\rho^{-1}$ is the inverse of $-\rho$, so $-\rho$ is an isomorphism. □

[82] The top square commutes by naturality of ρ and the bottom square commutes because h is an N-homomorphism (40).

[83] In other words, the assignments $M \mapsto \mathbf{EM}(M)$ and $\rho \mapsto -\rho$ becomes a functor from the category of monads on $\mathbf{C}$ and monad morphisms to the category of categories.

With the monad isomorphism $-+\mathbf{1} \cong T_\Sigma$ of Example 57, we obtain an isomorphism $\mathbf{EM}(-+\mathbf{1}) \cong \mathbf{EM}(T_\Sigma)$, and composing it with the isomorphism of Proposition 54 (instantiating $E = \emptyset$), we get back the result from Example 55 that algebras for the maybe monad are the same thing as algebras for the signature with only a constant.

This motivates the following definition.

Definition 59 (**Set** presentation)**.** Let M be a monad on **Set**, an **algebraic presentation** of M is signature Σ and a set of equations E along with a monad isomorphism $M \cong T_{\Sigma,E}$. We also say M is presented by (Σ, E).

We have proven in Example 57 that $\Sigma = \{\mathsf{p} : 0\}$ and $E = \emptyset$ is an algebraic presentation for the maybe monad on **Set**. Here is a couple of additional examples.

Example 60 (Powerset)**.** The powerset monad $\mathcal{P}_{\mathrm{ne}}$ is presented by the theory of **semilattices** $(\Sigma_{\mathbf{SLat}}, E_{\mathbf{SLat}})$ where $\Sigma_{\mathbf{SLat}} = \{\oplus : 2\}$ and $E_{\mathbf{SLat}}$ contains the following equations stating that $\oplus$ is idempotent, commutative and associative resepctively.

$$x \vdash x = x \oplus x \qquad x, y \vdash x \oplus y = y \oplus x \qquad x, y, z \vdash x \oplus (y \oplus z) = (x \oplus y) \oplus z$$

...

Example 61 (Distributions)**.** The distribution monad $\mathcal{D}$ is presented by the theory of **convex algebras** $(\Sigma_{\mathbf{CA}}, E_{\mathbf{CA}})$ where $\Sigma_{\mathbf{CA}} = \{+_p : 2 \mid p \in (0,1)\}$ and $E_{\mathbf{CA}}$ contains the following equations for all $p, q \in (0,1)$.

$$x \vdash x = x +_p x \qquad x, y \vdash x +_p y = y +_{1-p} x$$
$$x, y, z \vdash (x +_p y) +_q z = x +_{pq} + (y +_{\frac{p(1-q)}{1-pq}} z)$$

...

Remark 62*.* Not all monads on **Set** have an algebraic presentation. Linton also gave in [Lin66] a characterization of which monads can be presented by a signature with finitary operation symbols, such monads are aptly called **finitary monads**.

2 Generalized Metric Spaces

The Homeless Wanderer

Emahoy Tsegué-Maryam Guèbrou

2.1 L-Spaces

Definition 63 (Complete lattice)**.** A **complete lattice** is a partially ordered set $(\mathsf{L}, \leq)$[84] where all subsets $S \subseteq \mathsf{L}$ have a infimum and a supremum denoted by $\inf S$ and $\sup S$ respectively. In particular, L has a **bottom element** $\bot = \sup \emptyset$ and a **top element** $\top = \inf \emptyset$ that satisfy $\bot \leq \varepsilon \leq \top$ for all $\varepsilon \in \mathsf{L}$. We use L to refer to the lattice and its underlying set.

[84] i.e. L is a set and $\leq \subseteq \mathsf{L} \times \mathsf{L}$ is a binary relation on L that is reflexive, transitive and antisymmetric.

Let us describe two central (for this thesis) examples of complete lattices.

Example 64 (Unit interval)**.** The **unit interval** $[0, 1]$ is the set of real numbers between 0 and 1. It is a poset with the usual order $\leq$ ("less than or equal") on numbers. It is usually an axiom in the definition of $\mathbb{R}$[85] that all non-empty bounded subsets of real numbers have an infimum and a supremum. Since all subsets of $[0, 1]$ are bounded (by 0 and 1), we conclude that $([0, 1], \leq)$ is a complete lattice with $\bot = 0$ and $\top = 1$.

[85] Or possibly a theorem proven after constructing $\mathbb{R}$.

Later in this section, we will see elements of $[0, 1]$ as distances between points of some space. It would make sense, then, to extend the interval to contain values bigger than 1. Still because a complete lattice must have a top element there must be a number above all others. We could either stop at some arbitrary $0 \leq B \in \mathbb{R}$ and consider $[0, B]$, or we can consider ∞ to be a number as done below.[86]

[86] If one needs negative distances, it is also possible to work with any interval $[A, B]$ with $A \leq B \in \mathbb{R}$, or even $[-\infty, \infty]$. We will stick to $[0, 1]$ and $[0, \infty]$.

Example 65 (Extended interval)**.** Similarly to the unit interval, the **extended interval** is the set $[0, \infty]$ of positive real numbers extended with ∞, and it is a poset after asserting $\varepsilon \leq \infty$ for all $\varepsilon \in [0, \infty]$. It is also a complete lattice because non-empty bounded subsets of $[0, \infty)$ still have an infimum and supremum, and if a subset is not bounded above or contains ∞, then its supremum is ∞. We find that 0 is bottom and ∞ is top.

It is the prevailing custom to consider distances valued in the extended interval.[87] However, in our research, we preferred to use the unit interval for a very subtle and

[87] In fact, $[0, \infty]$ is also famous under the name *Lawvere quantale* because of Lawvere's seminal paper [Law02]. In that work, he used some structure on $[0, \infty]$ (now called a quantale) to give a categorical definition very close to that of a metric, the most accepted abstract notion of distance.

inconsequential reason (explained in **??**), and that is why most examples will have distances valued in $[0,1]$.

There are many other interesting complete lattices, although (unfortunately) they are rarely viewed as possible places to value distances.

Example 66 (Booleans)**.** The **Boolean lattice** B is the complete lattice containing only two elements, bottom and top. Its name comes from the interpretation of $\bot$ as a `false` value and $\top$ as a `true` value which makes the infimum act like an `AND` and the supremum like an `OR`.

Example 67 (Extended natural numbers)**.** The set $\mathbb{N}_\infty$ of natural numbers extended with ∞ is a sublattice of $[0,\infty]$.[88] Indeed, it is a poset with the usual order and the infimum and supremum of a subset of natural numbers is either itself a natural number of ∞ (when the subset is unbounded).

[88] As expected, a **sublattice** of $(\mathsf{L},\leq)$ is a set $\mathsf{S}\subseteq\mathsf{L}$ closed under taking infimums and supremums. Note that the top and bottom of S need not coincide with those of L. For instance $[0,1]$ is a sublattice of $[0,\infty]$, but $\top=1$ in the former and $\top=\infty$ in the latter.

Example 68 (Powerset lattice)**.** For any set X, we denote the powerset of X by $\mathcal{P}(X)$. The inclusion relation $\subseteq$ between subsets of X makes $\mathcal{P}(X)$ a poset. The infimum of a family of subsets $S_i\subseteq X$ is the intersection $\cap_{i\in I}S_i$, and its supremum is the union $\cup_{i\in I}S_i$. Hence, $\mathcal{P}(X)$ is a complete lattice. The bottom element is $\emptyset$ and the top element is X.

It is well-known that subsets of X correspond to functions $X\to\{\bot,\top\}$.[89] Endowing the two-element set with the complete lattice structure of B is what yields the complete lattice structure on $\mathcal{P}(X)$. The following example generalizes this construction.

[89] A subset $S\subseteq X$ is sent to the characteristic function χ_S, and a function $f : X\to\mathsf{B}$ is sent to $f^{-1}(\top)$. We say that $\{\bot,\top\}$ is the subobject classifier of **Set**.

Example 69 (Function space)**.** Given a complete lattice $(\mathsf{L},\leq)$, for any set X, we denote the set of functions from X to L by L^X. The pointwise order on functions defined by

$$f\leq_* g \iff \forall x\in X, f(x)\leq g(x)$$

is a partial order on L^X. The infimums and supremums of families of functions are also computed pointwise.[90] Namely, given $\{f_i : X\to\mathsf{L}\}_{i\in I}$, for all $x\in X$:

$$(\inf_{i\in I} f_i)(x)=\inf_{i\in I} f_i(x)\quad\text{and}\quad(\sup_{i\in I} f_i)(x)=\sup_{i\in I} f_i(x).$$

This makes L^X a complete lattice. The bottom element is the function that is constant at $\bot$ and the top element is the function that is constant at $\top$.

[90] Taking $\mathsf{L}=\mathsf{B}$, we find that $\mathcal{P}(X)$ and B^X are isomorphic as complete lattices under the usual correspondence. Namely, pointwise infimums and supremums become intersections and unions respectively. For example, if $\chi_S,\chi_T : X\to\mathsf{B}$ are the characteristic functions of $S,T\subseteq X$, then

$$\begin{aligned}\inf\{\chi_S,\chi_T\}(x)=\top &\Leftrightarrow \chi_S(x)=\chi_T(x)=\top\\ &\Leftrightarrow x\in S \text{ and } x\in T\\ &\Leftrightarrow x\in S\cap T.\end{aligned}$$

As a special case of function spaces, it is easy to show that when X is a set with two elements, L^X is isomorphic (as complete lattices) to the product $\mathsf{L}\times\mathsf{L}$ as defined below.

Example 70 (Product)**.** Let $(\mathsf{L},\leq_\mathsf{L})$ and $(\mathsf{K},\leq_\mathsf{K})$ be two complete lattices. Their **product** is the poset $(\mathsf{L}\times\mathsf{K},\leq_{\mathsf{L}\times\mathsf{K}})$ on the Cartesian product of L and K with the order defined by

$$(\varepsilon,\delta)\leq_{\mathsf{L}\times\mathsf{K}}(\varepsilon',\delta')\iff\varepsilon\leq_\mathsf{L}\varepsilon' \text{ and } \delta\leq_\mathsf{K}\delta'. \tag{48}$$

It is a complete lattice where the infimums and supremums are computed coordinatewise, namely, for any $S \subseteq \mathsf{L} \times \mathsf{K}$,[91]

$$\inf S = (\inf\{\pi_{\mathsf{L}}(c) \mid c \in S\}, \inf\{\pi_{\mathsf{K}}(c) \mid c \in S\}) \text{ and}$$
$$\sup S = (\sup\{\pi_{\mathsf{L}}(c) \mid c \in S\}, \sup\{\pi_{\mathsf{K}}(c) \mid c \in S\}).$$

The bottom (resp. top) element of $\mathsf{L} \times \mathsf{K}$ is the pairing of the bottom (resp. top) elements of L and K. i.e. $\bot_{\mathsf{L}\times\mathsf{K}} = (\bot_{\mathsf{L}}, \bot_{\mathsf{K}})$ and $\top_{\mathsf{L}\times\mathsf{K}} = (\top_{\mathsf{L}}, \top_{\mathsf{K}})$.

[91] Where π_{L} and π_{K} are the projections from $\mathsf{L} \times \mathsf{K}$ to L and K respectively.

The following example is also based on functions and it appears in many works on generalized notions of distances, e.g. [Fla97, HR13].

Example 71 (CDF)**.** A **cumulative distribution function**[92] (or CDF for short) is a function $f : [0,\infty] \to [0,1]$ that is monotone (i.e. $\varepsilon \leq \delta \implies f(\varepsilon) \leq f(\delta)$) and satisfies

$$f(\delta) = \sup\{f(\varepsilon) \mid \varepsilon < \delta\}. \tag{49}$$

[92] Although cumulative *sub*distribution function might be preferred.

Intuitively, (49) says that f cannot abruptly change value at some $x \in [0,\infty]$, but it can do that "after" some x.[93] For instance, out of the two functions below, only $f_{>1}$ is a CDF.

$$f_{\geq 1} = x \mapsto \begin{cases} 0 & x < 1 \\ 1 & x \geq 1 \end{cases} \qquad f_{>1} = x \mapsto \begin{cases} 0 & x \leq 1 \\ 1 & x > 1 \end{cases}$$

[93] This property is often called *right-continuity*.

We denote by $\mathsf{CDF}([0,\infty])$ the subset of $[0,1]^{[0,\infty]}$ containing all CDFs, it inherits a poset structure (pointwise ordering), and we can show it is a complete lattice.[94]

[94] Note however that $\mathsf{CDF}([0,\infty])$ is not a sublattice of $[0,1]^{[0,\infty]}$ because the infimums are not always taken pointwise. For instance, given $0 < n \in \mathbb{N}$, define f_n by (see them on Desmos)

$$f_n(x) = \begin{cases} 0 & x \leq 1 - \frac{1}{n} \\ nx & 1 - \frac{1}{n} < x < 1 \\ 1 & 1 \leq x \end{cases}.$$

The pointwise infimum of $\{f_n\}_{n\in\mathbb{N}}$ clearly sends everything below 1 to 0 and everything above and including 1 to 1, so it does not satisfy $f(1) = \sup_{\varepsilon<1} f(\varepsilon)$. We can find the infimum with the general formula that defines infimums in terms of supremums:

$$\inf_{n>0} f_n = \sup\{f \in \mathsf{CDF}([0,\infty]) \mid \forall n > 0, f \leq_* f_n\}.$$

We find that $\inf_{n>0} f_n = f_{>1}$.

Let $\{f_i : [0,\infty] \to [0,1]\}_{i\in I}$ be a family of CDFs. We will show the pointwise supremum $\sup_{i\in I} f_i$ is a CDF, and that is enough since having all supremums implies having all infimums.

- If $\varepsilon \leq \delta$, since all f_is are monotone, we have $f_i(\varepsilon) \leq f_i(\delta)$ for all $i \in I$ which implies
$$(\sup_{i\in I} f_i)(\varepsilon) = \sup_{i\in I} f_i(\varepsilon) \leq \sup_{i\in I} f_i(\delta) = (\sup_{i\in I} f_i)(\delta).$$
- For any $\delta \in [0,\infty]$, we have
$$(\sup_{i\in I} f_i)(\delta) = \sup_{i\in I} f_i(\delta) = \sup_{i\in I}\sup_{\varepsilon<\delta} f_i(\varepsilon) = \sup_{\varepsilon<\delta}\sup_{i\in I} f_i(\varepsilon) = \sup_{\varepsilon<\delta}(\sup_{i\in I} f_i)(\varepsilon).$$

Nothing prevents us from defining CDFs on other domains, and we will write $\mathsf{CDF}(\mathsf{L})$ for the complete lattice of functions $\mathsf{L} \to [0,1]$ that are monotone and satisfy (49). We could also change the codomain, but we will stick to $[0,1]$.

Definition 72 (L-space)**.** Given a complete lattice L and a set X, an L**-relation** on X is a function $d : X \times X \to \mathsf{L}$. We refer to the pair (X, d) as an L**-space**, and we will also use a single bold-face symbol $\mathbf{X}$ to refer to an L-space with underlying set X and L-relation $d_{\mathbf{X}}$.[95] The set X is called the **carrier** or the **underlying** set.

A **nonexpansive** map from $\mathbf{X}$ to $\mathbf{Y}$ is a function $f : X \to Y$ between the underlying sets of $\mathbf{X}$ and $\mathbf{Y}$ that satisfies

$$\forall x, x' \in X, \quad d_{\mathbf{Y}}(f(x), f(x')) \leq d_{\mathbf{X}}(x, x'). \tag{50}$$

[95] We will often switch between referring to spaces with $\mathbf{X}$ or $(X, d_{\mathbf{X}})$, and we will try to match the symbol for the space and the one for its underlying set only modifying the former with `mathbf`.

The identity maps $\mathrm{id}_X : X \to X$ and the composition of two nonexpansive maps are always nonexpansive[96], therefore we have a category whose objects are L-spaces and morphisms are nonexpansive maps. We denote it by L**Spa**.

This category is concrete over **Set** with the forgetful functor $U : \mathsf{L}\mathbf{Spa} \to \mathbf{Set}$ which sends an L-space $\mathbf{X}$ to its carrier and a morphism to the underlying function between carriers.

Remark 73. In the sequel, we will not distinguish between the morphism $f : \mathbf{X} \to \mathbf{Y}$ and the underlying function $f : X \to Y$. Although, we may write Uf for the latter, when disambiguation is necessary.

Instantiating L for different complete lattices, we can get a feel for what the categories L**Spa** look like. We also give concrete examples of L-spaces.

Examples 74 (Binary relations)**.** When $\mathsf{L} = \mathsf{B}$, a function $d : X \times X \to \mathsf{B}$ is the same thing as a subset of $X \times X$, which is the same thing as a binary relation on X.[97] Then, a B-space is a set equipped with a binary relation, and we choose to have, as a convention, $d(x,y) = \bot$ when x and y are related and $d(x,y) = \top$ when they are not.[98] A nonexpansive map from $\mathbf{X}$ to $\mathbf{Y}$ is a function $f : X \to Y$ such that for any $x, x' \in X$, $f(x)$ and $f(x')$ are related when x and x' are. When x and x' are not related, $f(x)$ and $f(x')$ might still be related.[99] The category B**Spa** is well-known under different names, **EndoRel** in [Vig23], **Rel** in [AHS06] (although that name is more commonly used for the category where relations are morphisms) and 2**Rel** in my book. Here are a couple of fun examples of B-spaces:

1. **Chess.** Let P be the set of positions on a chess board (`a2`, `d6`, `f3`, etc.) and $d_B : P \times P \to \mathsf{B}$ send a pair (p,q) to $\bot$ if and only if q is accessible from p in one bishop's move. The pair (P, d_B) is an object of B**Spa**. Let d_Q be the B-relation sending (p,q) to $\bot$ if and only if q is accessible from p in one queen's move. The pair (P, d_Q) is another object of B**Spa**. The identity function $\mathrm{id}_P : P \to P$ is nonexpansive from (P, d_B) to (P, d_Q) because whenever a bishop can go from p to q, a queen can too. However, it is not nonexpansive from (P, d_Q) to (P, d_B) because e.g. a queen can go from `a1` to `a2` but a bishop cannot.[100]
2. **Siblings.** Let H be the set of all humans (me, Paul Erdős, my brother Paul, etc.) and $d_S : H \times H \to \mathsf{B}$ send (h,k) to $\bot$ if and only if h and k are full siblings.[101] The pair (H, d_S) is an object of B**Spa**. Let $d_=$ be the B-relation sending (h,k) to $\bot$ if and only if h and k are the same person. The pair $(H, d_=)$ is another object of B**Spa**. The function $f : H \to H$ sending h to their biological mother is nonexpansive from (H, d_S) to $(H, d_=)$ because whenever h and k are full siblings, they have the same biologcial mother.

Examples 75 (Distances)**.** The main examples of L-spaces in this thesis are $[0,1]$-spaces or $[0,\infty]$-spaces. These are sets X equipped with a function $d : X \times X \to [0,1]$ or $d : X \times X \to [0,\infty]$, and we can usually understand $d(x,y)$ as the distance between two points $x, y \in X$. With this interpretation, a function is nonexpansive when applying it never increases the distances between points.[102] Let us give several examples of $[0,1]$- and $[0,\infty]$-spaces:

[96] Fix three L-spaces $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{Z}$ with two nonexpansive maps $f : X \to Y$ and $g : Y \to Z$, we have by nonexpansiveness of g then f:

$$d_{\mathbf{Z}}(gf(x), gf(x')) \leq d_{\mathbf{Y}}(f(x), f(x')) \leq d_{\mathbf{X}}(x, x').$$

[97] Hence, the choice of terminology L-relation.

[98] This conventions might look backwards, but it makes sense with the morphisms.

[99] Note that this interpretation of nonexpansiveness depends on our just chosen convention. Swapping the meaning of $d(x,y) = \top$ and $d(x,y) = \bot$ is the same thing as taking the opposite order on B (i.e $\top \leq \bot$), namely, morphisms become functions $f : X \to Y$ such that for any $x, x' \in X$, $f(x)$ and $f(x')$ are *not* related when neither are x and x'.

[100] In other words, the set of valid moves for a bishop is included in the set of valid moves for a queen, but not vice versa.

[101] Full siblings share the same biological parents.

[102] This is a justification for the term nonexpansive. In the setting of distances being real-valued, another popular term is 1-Lipschitz.

1. **Euclidean.** Probably the most famous notion of distance in mathematics is the Euclidean distance on real numbers $d : \mathbb{R} \times \mathbb{R} \to [0, \infty] = (x, y) \mapsto |x - y|$. The distance between any two points is unbounded, but it is never ∞. The pair $(\mathbb{R}, d)$ is an object of $[0, \infty]\mathbf{Spa}$. Multiplication by $r \in \mathbb{R}$ is a nonexpansive function $r \cdot - : (\mathbb{R}, d) \to (\mathbb{R}, d)$ if and only if r is between -1 and 1. Intuitively, a function $f : (\mathbb{R}, d) \to (\mathbb{R}, d)$ is nonexpansive when its derivative at any point is between -1 and 1.[103]

2. **Collaboration.** Let H be the set of humans again. The collaboration distance d between two humans h and k is the length of the shortest chain of humans between h and k in which two humans can be linked only if they co-authored a scientific paper together.[104] For instance $d(\text{me}, \text{Paul Erdős}) = 4$ as computed by `csauthors.net` on December 8th 2023:

$$\text{me} \xrightarrow{\text{[PS21]}} \text{D. Petrişan} \xrightarrow{\text{[GPR16]}} \text{M. Gehrke} \xrightarrow{\text{[EGP07]}} \text{M. Erné} \xrightarrow{\text{[EE86]}} \text{P. Erdős}$$

   The pair (H, d) is a $[0, \infty]$-space, but it could also be seen as a $\mathbb{N}_\infty$-space (because the length of a chain is always an integer).

3. **Hamming.** Let W be the set of words of the English language. If two words u and v have the same number of letters, the Hamming distance $d(u, v)$ between u and v is the number of positions in u and v where the letters do not match.[105] When u and v are of different lengths, we let $d(u, v) = \infty$, and we obtain a $[0, \infty]$-space (W, d). (It is also a $\mathbb{N}_\infty$-space.)

Remark 76. As Examples 75 come with many important intuitions, we will often call an L-relation $d : X \times X \to \mathsf{L}$ a **distance function** and $d(x, y)$ the **distance** from x to y,[106] even when L is neither $[0, 1]$ nor $[0, \infty]$.

Examples 77. We give more examples of L-spaces to showcase the potential of our abstract framework.

1. **Diversion.**[107] Let J be the set of products available to consumers inside a vending machine (including a "no purchase" option), the second-choice diversion $d(p, q)$ from product p to product q is the fraction of consumers that switch from buying p to buying q when p is removed (or out of stock) from the machine. That fraction is always contained between 0 and 1, so we have a function $d : J \times J \to [0, 1]$ which makes (J, d) an object of $[0, 1]\mathbf{Spa}$.[108]

2. **Rank.** Let P be the set of web pages available on the internet. In [BP98], the authors introduce an algorithm to measure the importance of a page $p \in P$ giving it a rank $R(p) \in [0, 1]$. This data can be organized in a function $d_R : P \times P \to [0, 1]$ which assigns $R(p)$ to a pair (p, p) and 0 (or 1) to a pair (p, q) with $p \neq q$.[109] This yields a $[0, 1]$-space (P, d_R).

   The rank of a page varies over time (it is computed from the links between all web pages which change quite frequently), so if we let T be the set of instants of

[103] The derivatives might not exist, so this is just an intuitive explaination.

[104] As conventions, the length of a chain is number of links, not humans. Also, $d(h, k) = \infty$ when no such chain exists between h and k, except when $h = k$, then $d(h, h) = 0$ (or we could say it is the length of the empty chain from h to h).

[105] For instance $d(\texttt{carrot}, \texttt{carpet}) = 2$ because these words differ only in two positions, the second and third to last ($\texttt{r} \neq \texttt{p}$ and $\texttt{o} \neq \texttt{e}$).

[106] The asymmetry in the terminology "distance from x to y" is justified because, in general, nothing guarantees $d(x, y) = d(y, x)$.

[107] This example takes inspiration from the diversion matrices in [CMS23], where they consider the automobile market in the U.S. instead of a vending machine.

[108] Eventhough d is valued in $[0, 1]$, calling it a distance function does not fit our intuition because when $d(p, q)$ is big, it means the products p and q are probably very similar.

[109] The values $d_R(p, q)$ when $p \neq q$ are considered irrelevant, so they are filled with an arbitrary value, e.g. 0 or 1.

time, we can define $d'_R(p,p)$ to be the function of type $T \to [0,1]$ which sends t to the rank $R(p)$ computed at time t.[110] This makes (P, d'_R) into a $[0,1]^T$-space.

[110] Again, $d_R(p,q)$ can be set to some unimportant constant value.

In order to create a search engine, we also need to consider the input of the user looking for some web page.[111] If U is the set of possible user inputs, we can define $d''_R(p,p)$ to depend on U and T, so that (P, d''_R) is a $[0,1]^{U\times T}$-space.

[111] The rank of a Wikipedia page about ramen will be lower when the user inputs "`Genre Humaine`" than when they input "`Ramen_Lord`".

While the categories $\mathsf{B}\mathbf{Spa}$, $[0,1]\mathbf{Spa}$ and $[0,\infty]\mathbf{Spa}$ are interesting on their own, they contain subcategories which are more widely studied. For instance, the category **Poset** of posets and monotone maps is a full subcategory of $\mathsf{B}\mathbf{Spa}$ where we only keep B-spaces (X,d) where the binary relation corresponding to d is reflexive, transitive and antisymmetric. Similarly, a $[0,\infty]$-space (X,d) where the distance function satisfies the triangle inequality $d(x,z) \leq d(x,y) + d(y,z)$ and reflexivity $d(x,x) \leq 0$ is known as a Lawvere metric space [Law02].

The next section lays out the language we will use to state conditions as those above on L-spaces. It implicitly relies on the following equivalent definition of L-spaces.

Definition 78 (L-structure)**.** Given a complete lattice L, an **L-structure**[112] is a set X equipped with a family of binary relations $R_\varepsilon \subseteq X \times X$ indexed by $\varepsilon \in \mathsf{L}$ satisfying

[112] We borrow the name "structure" from the very abstract notion of relational structure used in [FMS21, **?**, **?**].

- **monotonicity** in the sense that if $\varepsilon \leq \varepsilon'$, then $R_\varepsilon \subseteq R_{\varepsilon'}$, and
- **continuity** in the sense that for any I-indexed family of elements $\varepsilon_i \in \mathsf{L}$,[113]

$$\bigcap_{i\in I} R_{\varepsilon_i} = R_\delta, \text{ where } \delta = \inf_{i\in I} \varepsilon_i.$$

[113] By monotonicity, $R_\delta \subseteq R_{\varepsilon_i}$ so the inclusion $R_\delta \subseteq \cap_{i\in I} R_{\varepsilon_i}$ always holds.
A consequence of continuity (take $I = \emptyset$) is that $R_\top$ is the full binary relation $X \times X$.

Intuitively[114] $(x,y) \in R_\varepsilon$ should be interpreted as bounding the distance from x to y above by ε. Then, monotonicity means the points that are at a distance below ε are also at a distance below ε' when $\varepsilon \leq \varepsilon'$. Continuity means the points that are at a distance below a bunch of bounds ε_i are also at a distance below the infimum of those bounds $\inf_{i\in I} \varepsilon_i$.

[114] The proof of Proposition 79 will shed more light on these objects by equating them with L-spaces.

The names for these conditions come from yet another equivalent definition.[115] Organising the data of an L-structure into a function $R : \mathsf{L} \to \mathcal{P}(X \times X)$ sending ε to R_ε, we can recover monotonicity and continuity by seeing $\mathcal{P}(X \times X)$ as a complete lattice like in Example 68. Indeed, monotonicity is equivalent to R being a monotone function between the posets $(\mathsf{L}, \leq)$ and $(\mathcal{P}(X \times X), \subseteq)$, and continuity is equivalent to R preserving infimums. Seeing L and $\mathcal{P}(X \times X)$ as posetal categories, we can simply say that R is a continuous functor.[116]

[115] This time more directly equivalent.

[116] Limits in a posetal category are always computed by taking the infimum of all the points in the diagram, so preserving limits and preserving infimums is the same thing.

A morphism between two L-structures $(X, \{R_\varepsilon\})$ and $(Y, \{S_\varepsilon\})$ is a function $f : X \to Y$ satisfying

$$\forall \varepsilon \in \mathsf{L}, \forall x, x' \in X, (x,x') \in R_\varepsilon \implies (f(x), f(x')) \in S_\varepsilon. \tag{51}$$

This should feel similar to nonexpansive maps.[117] Let us call $\mathsf{L}\mathbf{Str}$ the category of L-structures.

[117] In words, (51) reads as: if x and x' are at a distance below ε' then so are $f(x)$ and $f(x')$.

Proposition 79. *For any complete lattice* L*, the categories* $\mathsf{L}\mathbf{Spa}$ *and* $\mathsf{L}\mathbf{Str}$ *are isomorphic.*[118]

[118] This result is a stripped down version of [MPP17, Theorem 4.3]

Proof. Given an L-relation (X, d), we define the binary relations $R^d_\varepsilon \subseteq X \times X$ by

$$(x, x') \in R^d_\varepsilon \iff d(x, x') \leq \varepsilon. \tag{52}$$

This family satisfies monotonicity because for any $\varepsilon \leq \varepsilon'$ we have

$$(x, x') \in R^d_\varepsilon \overset{(52)}{\iff} d(x, x') \leq \varepsilon \implies d(x, x') \leq \varepsilon' \overset{(52)}{\iff} (x, x') \in R^d_{\varepsilon'}.$$

It also satisfies continuity because if $(x, x') \in R_{\varepsilon_i}$ for all $i \in I$, then $d(x, x') \leq \varepsilon_i$ for all $i \in I$. By defintion of infimum, we must have $d(x, x') \leq \inf_{i \in I} \varepsilon_i$, hence $(x, x') \in R_{\inf_{i \in I} \varepsilon_i}$. We conclude the forward inclusion ($\subseteq$) of continuity holds, the converse ($\supseteq$) follows from monotonicity.

Taking $\mathsf{L} = \mathsf{B}$, Proposition 79 gives back our interpretation of B**Spa** as the category **2Rel** from Examples 74. Indeed, a B-structure is just a set X equipped with a binary relation $R_\perp \subseteq X \times X$ (because $R_\top$ is required to equal $X \times X$), and morphisms of B-structures are functions that preserve that binary relation. This also justifies our weird choice of $d(x, y) = \perp$ meaning x and y are related.

Any nonexpansive map $f : (X, d) \to (Y, \Delta)$ in L**Spa** is also a morphism between the L-structures $(X, \{R^d_\varepsilon\})$ and $(Y, \{R^\Delta_\varepsilon\})$ because for all $\varepsilon \in \mathsf{L}$ and $x, x' \in X$, we have

$$(x, x') \in R^d_\varepsilon \overset{(52)}{\iff} d(x, x') \leq \varepsilon \overset{(50)}{\implies} \Delta(f(x), f(x')) \leq \varepsilon \overset{(52)}{\iff} (f(x), f(x')) \in R^\Delta_\varepsilon.$$

It follows that the assignment $(X, d) \mapsto (X, \{R^d_\varepsilon\})$ is a functor $F : \mathsf{L}\mathbf{Spa} \to \mathsf{L}\mathbf{Str}$ acting trivially on morphisms.

Given an L-structure $(X, \{R_\varepsilon\})$, we define the function $d_R : X \times X \to \mathsf{L}$ by

$$d_R(x, x') = \inf \left\{ \varepsilon \in \mathsf{L} \mid (x, x') \in R_\varepsilon \right\}.$$

Note that monotonicity and continuity of the family $\{R_\varepsilon\}$ imply[119]

$$d_R(x, x') \leq \varepsilon \iff (x, x') \in R_\varepsilon. \tag{53}$$

[119] The converse implication ($\Leftarrow$) is by definition of infimum. For ($\Rightarrow$), continuity says that

$$R_{d_R(x,x')} = \bigcap_{\varepsilon \in \mathsf{L}, (x,x') \in R_\varepsilon} R_\varepsilon,$$

so $R_{d_R(x,x')}$ contains (x, x'), then by monotonicity, $d_R(x, x') \leq \varepsilon$ implies R_ε also contains (x, x').

This allows us to prove that a morphism $f : (X, \{R_\varepsilon\}) \to (Y, \{S_\varepsilon\})$ is nonexpansive from (X, d_R) to (Y, d_S) because for all $\varepsilon \in \mathsf{L}$ and $x, x' \in X$, we have

$$d_R(x, x') \leq \varepsilon \overset{(53)}{\iff} (x, x') \in R_\varepsilon \overset{(51)}{\implies} (f(x), f(x')) \in S_\varepsilon \overset{(53)}{\iff} d_S(f(x), f(x')) \leq \varepsilon,$$

hence putting $\varepsilon = d(x, x')$, we obtain $d_S(f(x), f(x')) \leq d_R(x, x')$. It follows that the assignment $(X, \{R_\varepsilon\}) \mapsto (X, d_R)$ is a functor $G : \mathsf{L}\mathbf{Str} \to \mathsf{L}\mathbf{Spa}$ acting trivially on morphisms.

Observe that (52) and (53) together say that $R^{d_R}_\varepsilon = R_\varepsilon$ and $d_{R^d} = d$, so F and G are inverses to each other on objects. Since both functors do nothing to morphisms, we conclude that F and G are inverses to each other, and that $\mathsf{L}\mathbf{Spa} \cong \mathsf{L}\mathbf{Str}$. □

2.2 Equational Constraints

It is often the case one wants to impose conditions on the L-spaces they consider. For instance, recall that when L is $[0, 1]$ or $[0, \infty]$, L-spaces are sets with a notion of distance between points. Starting from our intuition on the distance between points of the space we live in, people have come up with several abstract conditions they enforce on distance functions. For example, we can restate (with a slight modification[120]) the axioms defining metric spaces.

[120] The separation axiom is now divided in two, (55) and (56).

First, symmetry says that the distance from x to y is the same as the distance from y to x:

$$\forall x, y \in X, \quad d(x,y) = d(y,x). \tag{54}$$

Reflexivity, also called indiscernibility of identicals, says that the distance between x and itself is 0 (i.e. the smallest distance possible):

$$\forall x \in X, \quad d(x,x) = 0. \tag{55}$$

Identity of indiscernibles, also called Leibniz's law, says that if two points x and y are at distance 0, then x and y must be the same:

$$\forall x, y \in X, \quad d(x,y) = 0 \implies x = y. \tag{56}$$

Finally, the triangle inequality says that the distance from x to z is always smaller than the sum of the distances from x to y and from y to z:

$$\forall x, y, z \in X, \quad d(x,z) \leq d(x,y) + d(y,z). \tag{57}$$

There are also very famous axioms on B-spaces (X, d) that arise from viewing the binary relation corresponding to d as some kind of order on elements of X.

First, reflexivity says that any element x is related to itself.[121] Translating back to the B-relation, this is equivalent to:

[121] We abstract orders that look like the "smaller or equal" order $\leq$ on say real numbers rather than the strict order $<$.

$$\forall x \in X, \quad d(x,x) = \bot. \tag{58}$$

Antisymmetry says that if both (x,y) and (y,x) are in the order relation, then they must be equal:

$$\forall x, y \in X, \quad d(x,y) = \bot = d(y,x) \implies x = y. \tag{59}$$

Finally, transitivity says that if (x,y) and (y,z) belong to the order relation, then so does (x,z):

$$\forall x, y, z \in X, \quad d(x,y) = \bot = d(y,z) \implies d(x,z) = \bot. \tag{60}$$

We can immediately notice that all the axioms (54)–(60) start with a universal quantification of variables. A harder thing to see is that we never actually need to talk about equality between distances. For instance, the equation $d(x,y) = d(y,x)$ in the axiom of symmetry (54) can be replaced by two inequations $d(x,y) \leq d(y,x)$ and $d(y,x) \leq d(x,y)$, and moreover since x and y are universally quantified, only one of these inequations is necessary:

$$\forall x, y \in X, \quad d(x,y) \leq d(y,x). \tag{61}$$

If we rely on the equivalence between L-spaces and L-structures (Proposition 79), we can transform (61) into a family of implications indexed by all $\varepsilon \in \mathsf{L}$:[122]

$$\forall x, y \in X, \quad (y,x) \in R_\varepsilon^d \implies (x,y) \in R_\varepsilon^d. \tag{62}$$

[122] Recall that $(x,y) \in R_\varepsilon^d$ is the same thing as $d(x,y) \leq \varepsilon$. Hence, (61) and (62) are equivalent because requiring $d(x,y)$ to be smaller than $d(y,x)$ is equivalent to requiring all upper bounds of $d(y,x)$ (in particular $d(y,x)$ itself) to also be upper bounds of $d(x,y)$.

Starting from the triangle inequality (57) and applying the same transformations that got us from (54) to (62), we obtain a family of implications indexed by two values $\varepsilon, \delta \in \mathsf{L}$:[123]

$$\forall x, y, z \in X, \quad (x, y) \in R^d_\varepsilon \text{ and } (y, z) \in R^d_\delta \implies (x, z) \in R^d_{\varepsilon+\delta}. \tag{63}$$

[123] You can try to prove how (57) and (63) are equivalent if the process of going from the former to the latter was not clear to you.

The last conceptual step is to make the L.H.S. of the implication part of the universal quantification. That is, instead of saying "for all x and y, if P then Q", we say "for all x and y such that P, Q". We do this by introducing a syntax very similar to the equations of universal algebra. We fix a complete lattice $(\mathsf{L}, \leq)$, but as mentionned before, you can keep in mind the examples $\mathsf{L} = [0, 1]$ and $\mathsf{L} = [0, \infty]$.

Definition 80 (Quantitative equation)**.** A **quantitative equation** (over L) is a tuple comprising an L-space $\mathbf{X}$ called the **context**, two elements $x, y \in X$ and optionally an element $\varepsilon \in \mathsf{L}$. We write these as $\mathbf{X} \vdash x = y$ when no ε is given or $\mathbf{X} \vdash x =_\varepsilon y$ when it is given.

An L-space $\mathbf{A}$ **satisfies** a quantitative equation

- $\mathbf{X} \vdash x = y$ if for any nonexpansive assignment $\iota : \mathbf{X} \to \mathbf{A}$, $\iota(x) = \iota(y)$.
- $\mathbf{X} \vdash x =_\varepsilon y$ if for any nonexpansive assignment $\iota : \mathbf{X} \to \mathbf{A}$, $d_\mathbf{A}(\iota(x), \iota(y)) \leq \varepsilon$.

We use ϕ and ψ to refer to a quantitative equation, and we write $\mathbf{A} \vDash \phi$ when $\mathbf{A}$ satisfies ϕ.[124] We will also write $\mathbf{A} \vDash^\iota \phi$ when the equality $\iota(x) = \iota(y)$ or the bound $d_\mathbf{A}(\iota(x), \iota(y)) \leq \varepsilon$ holds for a particular assignment $\iota : \mathbf{X} \to \mathbf{A}$.[125]

[124] Of course, satisfaction generalizes straightforwardly to sets of quantitative equations, i.e. if E is a set of quantitative equations, $\mathbf{A} \vDash E$ means $\mathbf{A} \vDash \phi$ for all $\phi \in E$.

[125] and not necessarily for all assignments.

Example 81 (Symmetry)**.** With $\mathsf{L} = [0, 1]$ or $\mathsf{L} = [0, \infty]$, we want to translate (62) into a quantitative equation. A first approximation would be replacing the relation R^d_ε with our new syntax $=_\varepsilon$ to obtain something like

$$x, y \vdash y =_\varepsilon x \implies x =_\varepsilon y.$$

We are not allowed to use implications like this, so we have implement the last step mentionned above by putting the premise $y =_\varepsilon x$ into the context. This means we need to quantify over variables x and y with a bound ε on the distance from y to x.

Note that when defining satisfaction of a quantitative equation, the quantification happens at the level of assignments $\iota : \mathbf{X} \to \mathbf{A}$. Hence, we have to find a context $\mathbf{X}$ such that nonexpansive assignments $\mathbf{X} \to \mathbf{A}$ correspond to choices of two elements in $\mathbf{A}$ with the same bound ε on their distance.

Let the context $\mathbf{X}_\varepsilon$ be the L-space with two elements x and y such that $d_{\mathbf{X}_\varepsilon}(y, x) = \varepsilon$ and all other distances are $\top$ ($\top$ is either 1 or ∞). A nonexpansive assignment $\iota : \mathbf{X}_\varepsilon \to \mathbf{A}$ is just a choice of two elements $\iota(x), \iota(y) \in A$ satisfying $d_\mathbf{A}(\iota(y), \iota(x)) \leq \varepsilon$.[126] For all of these, we have to impose the condition $d_\mathbf{A}(\iota(x), \iota(y)) \leq \varepsilon$. Therefore, our quantitative equation is

$$\mathbf{X}_\varepsilon \vdash x =_\varepsilon y. \tag{64}$$

[126] Indeed, since $\top$ is the top element of L, the other values of $d_\mathbf{X}$ being $\top$ means that they impose no further condition on $d_\mathbf{A}$.

For a fixed $\varepsilon \in \mathsf{L}$, an L-space $\mathbf{A}$ satisfies (64) if and only if it satisfies (62). Hence,[127] if $\mathbf{A}$ satisfies that quantitative equation for all $\varepsilon \in \mathsf{L}$, then it satisfies (54), i.e. the distance $d_\mathbf{A}$ is symmetric.

[127] Recall our argument in Footnote 122.

In practice, defining the context like this is more cumbersome than need be, so we will define some syntactic sugar to remedy this. Before that, we take the time to do another example.

Example 82 (Triangle inequality)**.** Again with $\mathsf{L} = [0,1]$ or $\mathsf{L} = [0,\infty]$, let the context $\mathbf{X}_{\varepsilon,\delta}$ be the L-space with three elements x, y and z such that $d_{\mathbf{X}_{\varepsilon,\delta}}(x,y) = \varepsilon$ and $d_{\mathbf{X}_{\varepsilon,\delta}}(y,z) = \delta$, and all other distances are $\top$. A nonexpansive assignment $\iota : \mathbf{X}_{\varepsilon,\delta} \to \mathbf{A}$ is just a choice of three elements $a = \iota(x), b = \iota(y), c = \iota(z) \in A$ such that $d_{\mathbf{A}}(a,b) \leq \varepsilon$ and $d_{\mathbf{A}}(b,c) \leq \delta$. Hence, if $\mathbf{A}$ satisfies

$$\mathbf{X}_{\varepsilon,\delta} \vdash x =_{\varepsilon+\delta} z, \tag{65}$$

it means that for any such assignment, $d_{\mathbf{A}}(a,c) \leq \varepsilon + \delta$ also holds. We conclude that $\mathbf{A}$ satisifes (63). If $\mathbf{A}$ satisfies $\mathbf{X}_{\varepsilon,\delta} \vdash x =_{\varepsilon+\delta} z$ for all $\varepsilon, \delta \in \mathsf{L}$, then $\mathbf{A}$ satisfies the triangle inequality (57).

Notice that in the contexts $\mathbf{X}_\varepsilon$ and $\mathbf{X}_{\varepsilon,\delta}$, we only needed to set one or two distances and all the others where the maximum they could be $\top$. In our syntactic sugar for quantitative equations, we will only write the distances that are important (using the syntax $=_\varepsilon$), and we understand the underspecified distances to be as high as they can be. For instance, (64) will be written[128]

$$y =_\varepsilon x \vdash x =_\varepsilon y, \tag{66}$$

and (65) will be written

$$x =_\varepsilon y, y =_\delta z \vdash x =_{\varepsilon+\delta} z. \tag{67}$$

In this syntax, we call **premises** everything on the left of the turnstile $\vdash$ and **conclusion** what is on the right.

More generally, when we write $\{x_i =_{\varepsilon_i} y_i\}_{i\in I} \vdash x =_\varepsilon y$ (resp. $\{x_i =_{\varepsilon_i} y_i\}_{i\in I} \vdash x = y$), it corresponds to the quantitative equation $\mathbf{X} \vdash x =_\varepsilon y$ (resp. $\mathbf{X} \vdash x = y$), where the context $\mathbf{X}$ contains the variables in[129]

$$X = \{x, y\} \cup \{x_i \mid i \in I\} \cup \{y_i \mid i \in I\},$$

and the L-relation is defined for $u, v \in X$ by[130]

$$d_{\mathbf{X}}(u,v) = \inf\{\varepsilon \mid u =_\varepsilon v \in \{x_i =_\varepsilon y_i\}_{i\in I}\}.$$

Here are some more translations:

- (55) becomes $\vdash x =_0 x$,[131]
- (56) becomes $x =_0 y \vdash x = y$,
- (58) becomes $\vdash x =_\perp x$,
- (59) becomes $x =_\perp y, y =_\perp x \vdash x = y$, and
- (60) becomes $x =_\perp y, y =_\perp z \vdash x =_\perp z$.

[128] We can understand this syntax as putting back the information in the context into an implication. For instance, you can read (66) as "if the distance from y to x is bounded above by ε, then so is the distance from x to y".

[129] Note that the x_is, y_is, x and y need not be distinct. In fact, x and y almost always appear in the x_is and y_is.

[130] In words, the distance from u to v is the smallest value ε such that $u =_\varepsilon v$ was a premise. It is rare that u and v appear several times, but our definition allows it.

[131] We write nothing to the left of the turnstile $\vdash$ instead of writing $\emptyset$.

Remark 83. The translations of (55) and (58) look very close. In fact, noting that 0 is the bottom element of $[0,1]$ and $[0,\infty]$, the quantitative equation $\vdash x =_\bot x$ can state the reflexivity of a distance in $[0,1]$ or $[0,\infty]$ or the reflexivity of a binary relation.

Similarly, in the translation of the triangle inequality (67), if we let ε and δ range over B and interpret $+$ as an OR, we get three vacuous quantitative equations[132] and the translation of (60) above. So transitivity and triangle inequality are the same under this abstract point of view.[133]

[132] When either ε or δ equals $\top$, $\varepsilon + \delta = \top$, but when the conclusion of a quantitative equation is $x =_\top z$, it must be satisfied because $\top$ is an upper bound on all distances by definition.

[133] These observations were probably folkloric since at least the original publication of [Law02] in 1973.

Let us continue this list of examples for a while, just in case it helps a reader that is looking to translate an axiom into a quantitative equation. We will also give some results later which could imply that reader's axiom cannot be translated in this language.

Examples 84. Let $\mathsf{L} = [0,1]$ or $\mathsf{L} = [0,\infty]$.

1. The **strong triangle inequality** states that $d(x,z) \leq \max\{d(x,y), d(y,z)\}$, it is equivalent to the satisfaction of the following family of quantitative equations

$$\forall \varepsilon, \delta \in \mathsf{L}, \quad x =_\varepsilon y, y =_\delta z \vdash x =_{\max\{\varepsilon,\delta\}} z. \tag{68}$$

Let $\mathsf{L} = \mathsf{B}$.

1. A binary relation R on $X \times X$ is said to be **functional** if there are no two distinct $y, y' \in X$ such that $(x,y) \in R$ and $(x,y') \in R$ for a single $x \in X$. This is equivalent to satisfying

$$x =_\bot y, x =_\bot y' \vdash y = y'. \tag{69}$$

2. We say $R \subseteq X \times X$ is **injective** if there are no two distinct $x, x' \in X$ such that $(x,y) \in R$ and $(x',y) \in R$ for a single $y \in X$.[134] This is equivalent to satisfying

$$x =_\bot y, x' =_\bot y \vdash x = x'. \tag{70}$$

[134] Equivalently, the opposite (or converse) of R is functional.

You may try to formulate totality or surjectivity of a binary relation with quantitative equations, but you will find that difficult. We show in Examples 95 that it is not possible.

3. We say $R \subseteq X \times X$ is **circular** if whenever (x,y) and (y,z) belong to R, then so does (z,x) (compare with transitivity (60)). This is equivalent to satisfying

$$x =_\bot y, y =_\bot z \vdash z =_\bot x. \tag{71}$$

That is enough concrete examples. We now turn to the study of subcategories of L**Spa** that are defined via (sets of) quantitative equations. The most notable examples are the categories **Poset** of posets and **Met** of (extended) metric spaces:

- **Poset** is the full subcategory of B**Spa** with all B-spaces satisfying reflexivity, antisymmetry and transitivity stated as quantitative equations:

$$E_{\mathbf{Poset}} = \{\vdash x =_\bot x, x =_\bot y, y =_\bot x \vdash x = y, x =_\bot y, y =_\bot z \vdash x =_\bot z\}.$$

- **Met** is the full subcategory of $[0,1]$**Spa** (taking $[0,\infty]$ works just as well) with all $[0,1]$-spaces satisfying symmetry, reflexivity, identity of indiscernibles and triangle inequality stated as quantitative equations: $E_{\mathbf{Met}}$ contains all of the following

$$\begin{aligned}\forall \varepsilon \in [0,1], \quad & y =_\varepsilon x \vdash x =_\varepsilon y \\ & \vdash x =_0 x \\ & x =_0 y \vdash x = y \\ \forall \varepsilon, \delta \in [0,1], \quad & x =_\varepsilon y, y =_\delta z \vdash x =_{\varepsilon+\delta} z.\end{aligned}$$

Given a set E of quantitative equations, we can define a full subcategory of L**Spa** that contains only those L-spaces that satisfy E, this is the category **GMet**(L, E) whose objects we call **generalized metric spaces** or **spaces** for short. We also write **GMet**(E) or **GMet** when the complete lattices L or the set E are fixed or irrelevant. There is an evident forgetful functor U : **GMet** $\to$ **Set** which is the composition of the inclusion functor **GMet** $\to$ L**Spa** and U : L**Spa** $\to$ **Set**.[135]

[135] Recall that while we use the same symbol for both forgetful functors, you can disambiguate them with the hyperlinks.

2.3 The Categories GMet

In this section, we study various properties of the categories of generalized metric spaces. We fix a complete lattice L and a set of quantitative equations E throughout, and denote by **GMet** the category of L-spaces that satisfy E.

The goal here is mainly to become familiar with L-spaces and quantitative equations, so not all results will be useful later. This also means we will often avoid the use of some abstract results (many will be proved later) that can (sometimes drastically) simplify some proofs.[136] In order to keep all the information about **GMet** in the same place, we will quickly mention at the end some natural things that can be derived via the big theorems of **??**.

[136] For instance, we will see that U : **GMet** $\to$ **Set** is a right adjoint, so it has many nice properties which we could use in this section.

We also take some time to identify some (well-known) conditions on L-spaces that cannot be expressed via quantitative equations.[137] These proofs are always in the same vein, we know **GMet** has some property, we show the class of L-spaces with a condition does not have that property, hence that condition is not expressible as a set of quantitative equations.

[137] Unfortunately, we cannot make an exhaustive list since the literature on different notions of metric spaces is too vast.

Products

The category **GMet** has all products. We prove this in three steps. First, we find the terminal object, second we show L**Spa** has all products, and third we show the products of L-spaces which all satisfy some quantitative equation also satisfies that quantitative equation.

Proposition 85. *The category* **GMet** *has a terminal object.*

Proof. The terminal object $\mathbf{1}$ in L**Spa** is relatively easy to find,[138] it is a singleton $\{*\}$ with the L-relation $d_{\mathbf{1}}$ sending $(*,*)$ to $\bot$. Indeed, for any L-space $\mathbf{X}$, we have a function $! : X \to *$ that sends any x to $*$, and because $d_{\mathbf{1}}(*,*) = \bot \leq d_{\mathbf{X}}(x,x')$ for

[138] Again, many abstract results could help guide our search, but it is enough to have a bit of intuition about L-spaces.

any $x, x' \in X$, ! is nonexpansive. We obtain a morphism $! : \mathbf{X} \to \mathbf{1}$, and since any other morphism $\mathbf{X} \to \mathbf{1}$ must have the same underlying function[139], ! is the unique morphism of this type.

[139] Because $\{*\}$ is terminal in **Set**.

Since **GMet** is a full subcategory of L**Spa**, it is enough to show **1** is in **GMet** to conclude it is the terminal object in this subcategory. We can do this by showing **1** satisfies absolutely all quantitative equations, and in particular those of E.[140] Let $\mathbf{X}$ be any L-space, $x, y \in X$ and $\varepsilon \in \mathsf{L}$. As we have seen above, there is only one assignment $\iota : \mathbf{X} \to \mathbf{1}$, and it sends x and y to $*$. This means

[140] Which defined **GMet** at the start of this section.

$$\iota(x) = * = \iota(y) \quad \text{and} \quad d_{\mathbf{1}}(\iota(x), \iota(y)) = d_{\mathbf{1}}(*, *) = \bot \leq \varepsilon.$$

Therefore **1** satisfies both $\mathbf{X} \vdash x = y$ and $\mathbf{X} \vdash x =_\varepsilon y$. □

Proposition 86. *The category* L**Spa** *has all products.*

Proof. Let $\{\mathbf{A}_i = (A_i, d_i) \mid i \in I\}$ be a family of L-spaces indexed by I. We define the L-space $\mathbf{A} = (A, d)$ with carrier $A = \prod_{i \in I} A_i$ (the Cartesian product of the carriers) and L-relation $d : A \times A \to \mathsf{L}$ defined by the following supremum:[141]

[141] For $a \in A$, we write a_i the ith coordinate of a.

$$\forall a, b \in A, \quad d(a, b) = \sup_{i \in I} d_i(a_i, b_i). \tag{72}$$

For each $i \in I$, we have the evident projection $\pi_i : \mathbf{A} \to \mathbf{A}_i$ sending $a \in A$ to $a_i \in A_i$, and it is nonexpansive because, by definition, for any $a, b \in A$,

$$d_i(a_i, b_i) \leq \sup_{i \in I} d_i(a_i, b_i) = d(a, b).$$

We will show that $\mathbf{A}$ with these projections is the product $\prod_{i \in I} \mathbf{A}_i$.

Let $\mathbf{X}$ be some L-space and $f_i : \mathbf{X} \to \mathbf{A}_i$ be a family of nonexpansive maps. By the universal property of the product in **Set**, there is a unique function $\langle f_i \rangle : X \to A$ satisfying $\pi_i \circ \langle f_i \rangle = f_i$ for all $i \in I$. It remains to show $\langle f_i \rangle$ is nonexpansive from $\mathbf{X}$ to $\mathbf{A}$. For any $x, x' \in X$, we have[142]

[142] The equation holds because the ith coordinate of $\langle f_i \rangle(x)$ is $f_i(x)$ by definition of $\langle f_i \rangle$, and the inequation holds because for all $i \in I$, $d_i(f_i(x), f_i(x')) \leq d_{\mathbf{X}}(x, x')$ by nonexpansiveness of f_i.

$$d(\langle f_i \rangle(x), \langle f_i \rangle(x')) = \sup_{i \in I} d_i(f_i(x), f_i(x')) \leq d_{\mathbf{X}}(x, x').$$

Note that a particular case of this construction for I being empty is the terminal object **1** from Proposition 85. Indeed, the empty Cartesian product is the singleton, and the empty supremum is the bottom element $\bot$. □

In order to show that satisfaction of a quantitative equation is preserved by the product of L-spaces, we first prove a simple lemma.

Lemma 87. *Let ϕ be a quantitative equation with context $\mathbf{X}$. If $f : \mathbf{A} \to \mathbf{B}$ is a nonexpansive map and $\mathbf{A} \vDash^\iota \phi$ for an assignment $\iota : \mathbf{X} \to \mathbf{A}$, then $\mathbf{B} \vDash^{f \circ \iota} \phi$.*

Proof. There are two very similar cases. If ϕ is of the form $\mathbf{X} \vdash x = y$, we have[143]

[143] The equivalences hold by definition of $\vDash$.

$$\mathbf{A} \vDash^\iota \phi \iff \iota(x) = \iota(y) \implies f\iota(x) = f\iota(y) \iff \mathbf{B} \vDash^{f \circ \iota} \phi.$$

If ϕ is of the form $\mathbf{X} \vdash x =_\varepsilon y$, we have[144]

[144] The equivalences hold by definition of $\vDash$, and the implication holds by nonexpansiveness of f.

$$\mathbf{A} \vDash^\iota \phi \iff d_{\mathbf{A}}(\iota(x), \iota(y)) \leq \varepsilon \implies d_{\mathbf{B}}(f\iota(x), f\iota(y)) \leq \varepsilon \iff \mathbf{B} \vDash^{f \circ \iota} \phi. \qquad \square$$

Proposition 88. *If all* L*-spaces* $\mathbf{A}_i$ *satisfy a quantitative equation* ϕ*, then* $\prod_{i\in I}\mathbf{A}_i \vDash \phi$*.*

Proof. Let $\mathbf{A} = \prod_{i\in I}\mathbf{A}_i$ and $\mathbf{X}$ be the context of ϕ. It is enough to show that for any assignment $\iota : \mathbf{X} \to \mathbf{A}$, the following equivalence holds:[145]

$$(\forall i \in I, \mathbf{A}_i \vDash^{\pi_i \circ \iota} \phi) \Longleftrightarrow \mathbf{A} \vDash^{\iota} \phi. \tag{73}$$

The proposition follows because if $\mathbf{A}_i \vDash \phi$ for all $i \in I$, then the L.H.S. holds for any ι, hence the R.H.S. does too, and we conclude $\mathbf{A} \vDash \phi$. Let us prove (73).

($\Rightarrow$) Consider the case $\phi = \mathbf{X} \vdash x = y$. The satisfaction $\mathbf{A}_i \vDash \phi$ means $\pi_i\iota(x) = \pi_i\iota(y)$. If it is true for all $i \in I$, then we must have $\iota(x) = \iota(y)$ by universality of the product, thus we get $\mathbf{A} \vDash^{\iota} \phi$. In case $\phi = \mathbf{X} \vdash x =_{\varepsilon} y$, the satisfaction $\mathbf{A}_i \vDash \phi$ means $d_{\mathbf{A}_i}(\pi_i\iota(x), \pi_i\iota(y)) \leq \varepsilon$. If it is true for all $i \in I$, we get $\mathbf{A} \vDash \phi$ because

$$d_{\mathbf{A}}(\iota(x), \iota(y)) = \sup_{i\in I} d_{\mathbf{A}_i}(\pi_i\iota(x), \pi_i\iota(y)) \leq \varepsilon.$$

($\Leftarrow$) Apply Lemma 87 for all π_i. □

[145] When I is empty, the L.H.S. of (73) is vacuously true, and the R.H.S. is true since $\mathbf{A}$ is the terminal object of L-space which we showed satisfies all quantitative equations in Proposition 85.

Corollary 89. *The category* **GMet** *has all products, and they are computed like in* L**Spa**.[146]

[146] We showed that products in L**Spa** of objects in **GMet** also belong to **GMet**, it follows that this is also their products in **GMet** because the latter is a full subcategory of L**Spa**.

Unfortunately, this means that the notion of metric space originally defined in [Fré06], and incidentally what the majority of mathematicians calls metric spaces, are not instances of generalized metric spaces as we defined them. Since they only allow finite distances, some infinite products do not exist.[147] In general, if one wants to bound the distance above by some $B \in \mathsf{L}$, this can be done with the equation $\vdash x =_B y$, but the value B is still allowed as a distance. For instance $[0,1]$**Spa** is the full subcategory of $[0,\infty]$**Spa** defined by the equation $\vdash x =_1 y$.

[147] For instance let $\mathbf{A}_n$ be the metric space with two points $\{a, b\}$ at distance $n > 0 \in \mathbb{N}$ from each other. Then $\mathbf{A} = \prod_{n>0\in\mathbb{N}}\mathbf{A}_n$ exists in $[0,\infty]$**Spa** as we have just proven, but

$$d_{\mathbf{A}}(a^*, b^*) = \sup_{n>0\in\mathbb{N}} d_{\mathbf{A}_n}(a, b) = \sup_{n>0\in\mathbb{N}} n = \infty,$$

which means $\mathbf{A}$ is not a metric space in the sense of [Fré06].

Arguably, this is only a superficially negative result since it is already common in parts of the literature [**?**] to allow infinite distances because the resulting category of metric spaces has better properties (like having infinite products and coproducts). There are some other conditions that one would like to impose on $[0,\infty]$-spaces which are not even preserved under finite products. We give two examples arising under the terminology partial metric.

Definition 90. An $[0,\infty]$-space (A, d) is called a **partial metric space** if it satisfies the following condiditons [Mat94, Definition 3.1]:[148]

$$\forall a, b \in A, \quad a = b \Longleftrightarrow d(a,a) = d(a,b) = d(b,b) \tag{74}$$

$$\forall a, b \in A, \quad d(a,a) \leq d(a,b) \tag{75}$$

$$\forall a, b \in A, \quad d(a,b) = d(b,a) \tag{76}$$

$$\forall a, b, c \in A, \quad d(a,c) \leq d(a,b) + d(b,c) - d(b,b) \tag{77}$$

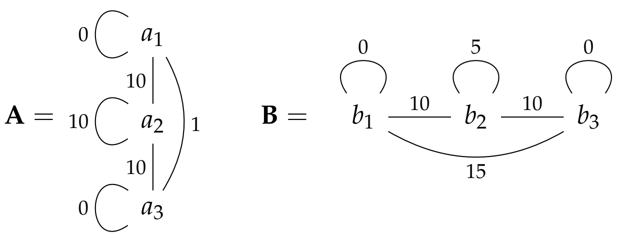
[148] There is some ambiguity in what $+$ and $-$ means when dealing with ∞ (the original paper supposes distances are finite), but it is rather unimportant to

These conditions look similar to what we were able to translate into equations before, but the first and last are problematic. We can translate (75) into $x =_{\varepsilon} y \vdash x =_{\varepsilon} x$, (76) is just symmetry which we can translate into $y =_{\varepsilon} x \vdash x =_{\varepsilon} y$.

For (74), note that the forward implication is trivial, but for the converse, we would need to compare three distances inside the context, which seems impossible because the context only bounds distances by above. For (77), the problem comes from the minus operation on distances which will not interact well with our only possibility of bounding by above. Indeed, if we tried something like $x =_{\varepsilon_1} y, y =_{\varepsilon_2} z, y =_{\varepsilon_3} y \vdash x =_{\varepsilon_1+\varepsilon_2-\varepsilon_3} z$, we could always take ε_3 really big (even ∞) and make the distance between x and z as close to 0 as we would like.

These are just informal arguments, but thanks to Corollary 89, we can prove formally that these conditions are not expressible as (sets of) quantitative equations. Let **A** and **B** be the $[0, \infty]$-spaces pictured below (the distances are symmetric).[149]

[149] The numbers on the lines indicate the distance between the ends of the line, e.g. $d_{\mathbf{A}}(a_1, a_1) = 0$, $d_{\mathbf{A}}(a_1, a_3) = 1$, and $d_{\mathbf{B}}(b_2, b_3) = 10$.

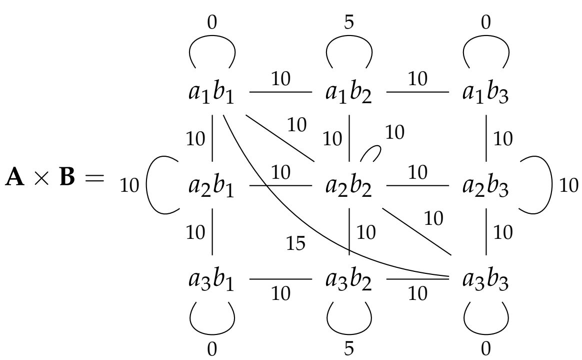


We can verify (by exhaustive checks) that **A** and **B** are partial metric spaces. If we take their product inside $[0, \infty]\mathbf{Spa}$, we find the following $[0, \infty]$-space (some distances are omitted) which does not satisfy (74) nor (77).[150]

[150] For (74), the three points in the middle row $\{a_2b_1, a_2b_2, a_2b_3\}$ are all at distance from each other and from themselves while not being equal. For (77), we have

$$d_{\mathbf{A}}(a_1b_1, a_3b_3) = 15, \text{ and}$$
$$d_{\mathbf{A}}(a_1b_1, a_2b_2) + d_{\mathbf{A}}(a_2b_2, a_3b_3) - d_{\mathbf{A}}(a_2b_2, a_2b_2) = 10,$$

but $15 > 10$.



We conclude that there is no set E of quantitative equations such that $\mathbf{GMet}([0, \infty], E)$ is the full subcategory of $[0, \infty]\mathbf{Spa}$ containing all the partial metric spaces.[151]

[151] It is still possible that the category of partial metrics and nonexpansive maps is identified with some **GMet**. That would mean (infinite) products of partial metrics exist but they are not computed with supremums.

This result is a bit more damaging to our concept of generalized metric space (especially since partial metric spaces were motivated by some considerations in programming semantics), but we had to expect something like this would happen with how much time mathematicians had to use and abuse the name metric.

Isometries

Since the forgetful functor $U : \mathsf{L}\mathbf{Spa} \to \mathbf{Set}$ preserves isomorphisms, we know that the underlying function of an isomorphism in $\mathsf{L}\mathbf{Spa}$ is a bijection between the carriers. What is more, we show in Proposition 92 it must preserve distances on the nose, i.e. it is an isometry.

Definition 91 (Isometry)**.** A morphism $f : \mathbf{X} \to \mathbf{Y}$ of L-spaces is called an **isometry** if[152]

$$\forall x, x' \in X, \quad d_{\mathbf{Y}}(f(x), f(x')) = d_{\mathbf{X}}(x, x'). \tag{78}$$

If furthermore, f is injective, we call it an **isometric embedding**.[153]

[152] The inequation in (50) was replaced by an equation.

[153] If $f : \mathbf{X} \to \mathbf{Y}$ is an isometric embedding, we can identify $\mathbf{X}$ with the subspace of $\mathbf{Y}$ containing all the elements in the image of f. Conversely, the inclusion of a subspace of $\mathbf{Y}$ in $\mathbf{Y}$ is always an isometric embedding.

Proposition 92. *In* **GMet***, isomorphisms are precisely the bijective isometries.*

Proof. We show a morphism $f : \mathbf{X} \to \mathbf{Y}$ is has an inverse $f^{-1} : \mathbf{Y} \to \mathbf{X}$ if and only if it is a bijective isometry.

($\Rightarrow$) Since the underlying functions of f and f^{-1} are inverses, they must be bijections. Moreover, using (50) twice, we find that for any $x, x' \in X$,[154]

$$d_{\mathbf{X}}(x, x') = d_{\mathbf{X}}(f^{-1}f(x), f^{-1}f(x')) \leq d_{\mathbf{Y}}(f(x), f(x')) \leq d_{\mathbf{X}}(x, x'),$$

thus $d_{\mathbf{X}}(x, x') = d_{\mathbf{Y}}(f(x), f(x'))$, so f is an isometry.

[154] This is a general argument showing that any nonexpansive function with a right inverse is an isometry, it is also an isometric embedding because a right inverse in **Set** implies injectivity.

($\Leftarrow$) Since f is bijective, it has an inverse $f^{-1} : Y \to X$ in **Set**, but we have to show f^{-1} is nonexpansive from $\mathbf{Y}$ to $\mathbf{X}$. For any $y, y' \in Y$, by surjectivity of f, there are $x, x' \in X$ such that $y = f(x)$ and $y' = f(x')$, then we have

$$d_{\mathbf{X}}(f^{-1}(y), f^{-1}(y')) = d_{\mathbf{X}}(f^{-1}f(x), f^{-1}f(x')) = d_{\mathbf{X}}(x, x') = d_{\mathbf{Y}}(f(x), f(x')) = d_{\mathbf{Y}}(y, y').$$

Hence f^{-1} is nonexpansive, it is even an isometry. $\square$

In particular, this means, as is expected, that isomorphisms preserve the satisfaction of quantitative equations. We can show a stronger statement: any isometric embedding reflects the satisfaction of quantitative equations.[155]

[155] This is stronger because we have just shown the inverse of an isomorphisms is an isometric embedding.

Proposition 93. *Let $f : \mathbf{Y} \to \mathbf{Z}$ be an isometric embedding between L-spaces and ϕ a quantitative equation, then*

$$\mathbf{Z} \vDash \phi \implies \mathbf{Y} \vDash \phi. \tag{79}$$

Proof. Let $\mathbf{X}$ be the context of ϕ. Any nonexpansive assignment $\iota : \mathbf{X} \to \mathbf{Y}$ yields an assignment $f \circ \iota : \mathbf{X} \to \mathbf{Z}$. By hypothesis, we know that $\mathbf{Z}$ satisfies ϕ for this particular assignment, namely,

$$\mathbf{Z} \vDash^{f \circ \iota} \phi. \tag{80}$$

We can use this and the fact that f is an isometric embedding to show $\mathbf{X} \vDash^{\iota} \phi$. There are two very similar cases.

If $\phi = \mathbf{X} \vdash x = y$, then we can show $\iota(x) = \iota(y)$ because we know $f\iota(x) = f\iota(x)$ by (80) and f is injective.

If $\phi = \mathbf{X} \vdash x =_{\varepsilon} y$, then we have $d_{\mathbf{Y}}(\iota(x), \iota(y)) = d_{\mathbf{Z}}(f\iota(x), f\iota(y)) \leq \varepsilon$, where the equation holds because f is an isometry and the inequation holds by (80). $\square$

Corollary 94. *Let $f : \mathbf{Y} \to \mathbf{Z}$ be an isometric embedding between L-spaces. If $\mathbf{Z}$ belongs to* **GMet***, then so does $\mathbf{Y}$. In particular, all the subspaces of a generalized metric space are also generalized metric spaces.*[156]

[156] Both parts are immediate. The first follows from applying (79) to all ϕ in E, the set of quantitative equations defining **GMet**. The second follows from Footnote 153.

Examples 95. Corollary 94 can be useful to identify some properties of L-spaces that cannot be modelled with quantitative equations. Here are a couple of examples.

1. A binary relation $R \subseteq X \times X$ is called **total** if for every $x \in X$, there exists $y \in X$ such that $(x, y) \in R$. Let **TotRel** be the full subcategory of B**Spa** containing only total relations, is **TotRel** equal to some $\mathbf{GMet}(\mathsf{B}, E)$ for some E? The existential quantification in the definition of total seems hard to simulate with a quantitative equation, but this is not a guarantee that maybe several equations cannot interact in such a counter-intuitive way.

   In order to prove that no set E defines total relations (i.e. $\mathbf{X} \vDash E$ if and only if the relation corresponding to $d_{\mathbf{X}}$ is total), we can exhibit an example of a B-space that is total with a subspace that is not total. It follows that **TotRel** is not closed under taking subspaces, so it is not a category of generalized metric spaces by Corollary 94.[157]

   Let $\mathbf{N}$ be the B-space with carrier $\mathbb{N}$ and B-relation $d_{\mathbf{N}}(n, m) = \bot \Leftrightarrow m = n + 1$ (the corresponding relation is the graph of the successor function). This space satisfies totality, but the subspace obtained by removing 1 is not total because $d_{\mathbf{N}}(0, n) = \bot$ only when $n = 1$.

   This same example works to show that surjectivity[158] cannot be defined via quantitative equations.

2. A very famous condition to impose on metric spaces is **completeness** (we do not need to define it here). Just as famous is the fact that $\mathbb{R}$ with the Euclidean distance from Examples 75 is complete but the subspace $\mathbb{Q}$ is not. Thus, completeness cannot be defined via quantitative equations.[159]

[157] Actually, we have only proven that **TotRel** cannot be defined as a subcategory of B**Spa** with quantitative equations. There may still be some convoluted way that $\mathbf{TotRel} \cong \mathbf{GMet}(\mathsf{L}, E)$ for some cleverly picked L and E (L could even be equal to B).

[158] This condition is symmetric to totality: $R \subseteq X \times X$ is **surjetive** if for every $y \in X$, there exists $x \in X$ such that $(x, y) \in R$.

[159] Still with the caveat that the category of complete metric spaces might still be isomorphic to some **GMet**.

Since isometric embeddings correspond to subspaces, one might think that they are the monomorphisms in **GMet**. Unfortunately, they are way more restrained. Any nonexpansive map that is injective is a monomorphism. To prove this, we rely on the existence of a space $\mathbb{A}$ that (informally) *can pick elements*.

Proposition 96. *There is a generalized metric space $\mathbb{A}$ on the set $\{*\}$ such that for any other space $\mathbf{X}$, any function $f : \{*\} \to X$ is a nonexpansive map $\mathbb{A} \to \mathbf{X}$.*[160]

[160] In category theory speak, $\mathbb{A}$ is a representing object of the forgetful functor $U : \mathbf{GMet} \to \mathbf{Set}$.

Proof. In L**Spa**, $\mathbb{A}$ is also easy to find, its L-relation is defined by $d_{\mathbb{A}}(*, *) = \top$. Indeed, any function $f : \{*\} \to X$ is nonexpansive because $\top$ is the maximum value $d_{\mathbf{X}}$ can assign, so

$$d_{\mathbf{X}}(f(*), f(*)) \leq \top = d_{\mathbb{A}}(*, *).$$

Unfortunately, this L-space does not satisfy some quantitative equations (e.g. reflexivity $x \vdash x =_{\bot} x$), so we cannot guarantee it belongs to **GMet**.

Recall that $\mathbf{1}$ is a generalized metric space on the same set $\{*\}$, but with $d_{\mathbf{1}}(*, *) = \bot$. However, in many cases, $\mathbf{1}$ is not the right candidate either because if every function $f : \{*\} \to X$ is nonexpansive from $\mathbf{1}$ to $\mathbf{X}$, it means $d_{\mathbf{X}}(x, x) = \bot$ for all $x \in X$, which is not always the case.[161]

[161] It is equivalent to satisfying reflexivity.

We have two L-spaces at the extremes of a range of L-spaces $\{(\{*\}, d_{\varepsilon})\}_{\varepsilon \in \mathsf{L}}$, where the L-relation d_{ε} sends $(*, *)$ to ε. At one extreme, we are guaranteed to be in **GMet**, but we are too restricted, and at the other extreme we might not belong to **GMet**.

Getting inspiration from the intermediate value theorem, we can attempt to find a middle ground, namely, a value $\varepsilon \in \mathsf{L}$ such that setting $d_{\mathbb{A}}(*,*) = \varepsilon$ yields a space that lives in **GMet** but is not too restricted.

One thing that could make sense is to take the biggest value (and hence the least restricted space that is in **GMet**). Formally, let

$$d_{\mathbb{A}}(*,*) = \sup\{\varepsilon \in \mathsf{L} \mid (\{*\}, d_\varepsilon) \vDash E\}.$$

It remains to check that any function $f : \{*\} \to X$ is nonexpansive from $\mathbb{A}$ to $\mathbf{X}$. Consider the image of f seen as a subspace of $\mathbf{X}$. By Corollary 94, it belongs to **GMet** and hence satisfies E. Moreover, it is clearly isomorphic to the L-space $(\{*\}, d_\varepsilon)$ with $\varepsilon = d_{\mathbf{X}}(f(*), f(*))$[162], which means that L-space satisfies E as well (by Corollary 94 again). We conclude that $d_{\mathbf{X}}(f(*), f(*)) \leq d_{\mathbb{A}}(*,*)$.

As a bonus, one could check that for any $\varepsilon \in \mathsf{L}$ that is smaller than $d_{\mathbb{A}}(*,*)$, $(\{*\}, d_\varepsilon)$ also belongs to **GMet**. □

[162] The isomorphism is the restriction of f to its image.

Proposition 97. *In* **GMet**, *monomorphisms are precisely the injective nonexpansive maps.*

Proof. We show a morphism $f : \mathbf{X} \to \mathbf{Y}$ is monic if and only if it is injective.

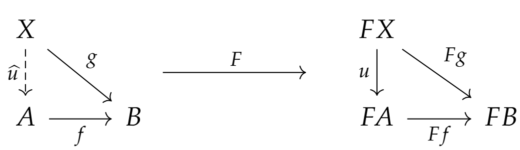
($\Rightarrow$) Let $x, x' \in X$ be such that $f(x) = f(x')$, and identify these elements with functions $x, x' : \{*\} \to X$ sending $*$ to x and x' respectively. By Proposition 96, we get two nonexpansive maps $x, x' : \mathbb{A} \to \mathbf{X}$. Post-composing by f, we find that $f \circ x = f \circ x'$ because they both send $*$ to $f(x) = f(x')$. By monicity of f, we find that $x = x'$ (as morphisms and hence as elements of X). We conclude that f is injective.

($\Leftarrow$) Suppose that $f \circ g = f \circ h$ for some nonexpansive maps $g, h : \mathbf{Z} \to \mathbf{X}$. Applying the forgetful functor $U : \mathbf{GMet} \to \mathbf{Set}$, we wind that $f \circ g = f \circ h$ also as functions. Since Uf is injective, Ug and Uh must be equal, and since U is faithful, we obtain $g = h$. □

It remains to give a categorical characterisation of isometric embeddings. This will rely on a well-known[163] abstract notion that we define here for completeness.

[163] While it is well-known, especially to those familiar with fibered category theory, it does not usually fit in a basic category theory background.

Definition 98 (Cartesian morphism)**.** Let $F : \mathbf{C} \to \mathbf{D}$ be a functor, and $f : A \to B$ be a morphism in $\mathbf{D}$. We say f is a **cartesian morphism** if for every morphism $g : X \to B$ and factorization $Fg = Ff \circ u$, there exists a unique morphism $\widehat{u} : X \to A$ with $F\widehat{u} = u$ satisfying $x = f \circ \widehat{u}$. This can be summarized (without the quantifiers) in the diagram below.



Example 99 (in **GMet**)**.** Let us unroll this in the important case for us, when F is the forgetful functor $U : \mathbf{GMet} \to \mathbf{Set}$. A nonexpansive map $f : \mathbf{A} \to \mathbf{B}$ is a cartesian morphism if for any nonexpansive map $g : \mathbf{X} \to \mathbf{B}$, all functions $u : X \to A$ satisfying $g = f \circ u$ are nonexpansive maps $u : \mathbf{X} \to \mathbf{A}$.[164]

[164] We do not bother to write $\widehat{u}$ as it is automatically unique with underlying function u because U is faithful.

We can turn this around into an equivalent definition. The morphism $f : \mathbf{A} \to \mathbf{B}$ is cartesian if for all functions $u : X \to A$, $f \circ u$ being nonexpansive from $\mathbf{X}$ to $\mathbf{B}$ implies u is nonexpansive from $\mathbf{X}$ to $\mathbf{A}$.[165] In [AHS06, Definition 8.6], f is also called an *initial morphism*.

[165] If $f \circ u$ is nonexpansive from $\mathbf{X}$ to $\mathbf{B}$, then it is equal to g for some $g : \mathbf{X} \to \mathbf{B}$ which yields $u : \mathbf{X} \to \mathbf{A}$ being nonexpansive.

Proposition 100. *A morphism $f : \mathbf{A} \to \mathbf{B}$ in* **GMet** *is an isometric embedding if and only if it is monic and cartesian.*

Proof. By Proposition 97, being an isometric embedding is equivalent to being a monomorphism (i.e. being injective) and being an isometry. Therefore, it is enough to show that when f is injective, isometry $\iff$ cartesian.

($\Rightarrow$) Suppose f is an isometry, and let $u : X \to A$ be a function such that $f \circ u$ is nonexpansive from $\mathbf{X} \to \mathbf{B}$, we need to show u is nonexpansive from $\mathbf{X} \to \mathbf{A}$.[166] This is true because

[166] We use the definition of cartesian in Example 99.

$$\forall x, x' \in \mathbf{X}, \quad d_{\mathbf{A}}(u(x), u(x')) = d_{\mathbf{B}}(fu(x), fu(x')) \leq d_{\mathbf{X}}(x, x'),$$

where the equation follows from f being an isometry and the inequation from nonexpansiveness of $f \circ u$.

($\Leftarrow$) Suppose f is cartesian. For any $a, a' \in A$, we know that $d_{\mathbf{B}}(f(a), f(a')) \leq d_A(a, a')$, but we still need to show the converse inequality. Let $\mathbf{X}$ be the subspace of B containing only the image of a and a' (its carrier is $\{f(a), f(a')\}$), and $g : X \to A$ be the function sending $f(a)$ to a and $f(a')$ to a'.[167] Notice that $f \circ g$ is the inclusion of $\mathbf{X}$ in B which is nonexpansive. Because f is cartesian, g must then be nonexpansive from $\mathbf{X}$ to $\mathbf{A}$ which implies

[167] We use the injectivity of f here.

$$d_{\mathbf{A}}(a, a') = d_{\mathbf{A}}(g(f(a)), g(f(a'))) \leq d_{\mathbf{X}}(f(a), f(a')) = d_{\mathbf{B}}(f(a), f(a')).$$

We conclude that f is an isometry. □

Corollary 101. *If the composition $\mathbf{A} \xrightarrow{f} \mathbf{B} \xrightarrow{g} \mathbf{C}$ is an isometric embedding, then f is an isometric embedding.*[168]

[168] With the characterisation of Proposition 100, this abstractly follows from [AHS06, Proposition 8.9]. We give the concrete proof anyways.

Proof. It is a standard result that if $g \circ f$ is monic then so is f. Even more standard for injectivity. Now, if $g \circ f$ is an isometry, we have for any $a, a' \in X$,[169]

[169] The equation holds by hypothesis that $g \circ f$ is an isometry and the two inequations hold by nonexpansiveness of g and f.

$$d_{\mathbf{A}}(a, a') = d_{\mathbf{C}}(gf(a), gf(a')) \leq d_{\mathbf{B}}(f(a), f(a')) \leq d_{\mathbf{A}}(a, a'),$$

and we conclude that $d_{\mathbf{A}}(a, a') = d_{\mathbf{B}}(f(a), f(a'))$, hence f is an isometry. □

The question of concretely characterizing epimorphisms is harder to settle. We can do it for L**Spa**, but not for an arbitrary **GMet**.

Proposition 102. *In L**Spa**, a morphism $f : \mathbf{X} \to \mathbf{A}$ is epic if and only if it is surjective.*

Proof. ($\Rightarrow$) Given any $a \in A$, we define the L-space $\mathbf{A}_a$ to be $\mathbf{A}$ with an additional copy of a with all the same distances. Namely, the carrier is $A + \{*_a\}$, for any $a' \in A$, $d_{\mathbf{A}_a}(*_a, a') = d_{\mathbf{A}}(a, a')$ and $d_{\mathbf{A}_a}(a', *_a) = d_{\mathbf{A}}(a', a)$, and all the other distances are as in $\mathbf{A}$.[170]

[170] This construction is already impossible to do in an arbitrary **GMet**. For instance, if $\mathbf{A}$ satisfies $x =_0 y \vdash x = y$, then $\mathbf{A}_a$ does not because $d_{\mathbf{A}_a}(a, *_a) = 0$.

If $f : \mathbf{X} \to \mathbf{A}$ is not surjective, then pick $a \in A$ that is not in the image of f, and define two functions $g_a, g_* : A \to A + \{*_a\}$ that act as identity on all A except a where $g_a(a) = a$ and $g_*(a) = *_a$. By construction, both g_a and g_* are nonexpansive from $\mathbf{A}$ to $\mathbf{A}_a$ and $g_a \circ f = g_* \circ f$. Since $g_a \neq g_*$, f cannot be epic, and we have proven the contrapositive of the forward implication.

($\Leftarrow$) Suppose that $g, g' : \mathbf{A} \to \mathbf{B}$ are morphisms in $\mathsf{L}\mathbf{Spa}$ such that $g \circ f = g' \circ f$. Apply the forgetful functor to get $Ug \circ Uf = Ug' \circ Uf$, and since U is epic in $\mathbf{Set}$, we know $Ug = Ug'$. Since U is faithful, we conclude that $g = g'$.[171] $\square$

[171] This direction works in an arbitrary **GMet**.

Proposition 103. *Let $f : \mathbf{A} \to \mathbf{B}$ be a split epimorphism between* L*-spaces and ϕ a quantitative equation, then*

$$\mathbf{A} \vDash \phi \implies \mathbf{B} \vDash \phi. \tag{81}$$

Proof. Let $g : \mathbf{B} \to \mathbf{A}$ be the right inverse of f (i.e. $f \circ g = \mathrm{id}_\mathbf{B}$) and $\mathbf{X}$ be the context of ϕ.[172] Any nonexpansive assignment $\iota : \mathbf{X} \to \mathbf{B}$ yields an assignment $g \circ \iota : \mathbf{X} \to \mathbf{A}$. By hypothesis, we know that $\mathbf{A}$ satisfies ϕ for this particular assignment, namely,

$$\mathbf{A} \vDash^{g \circ \iota} \phi. \tag{82}$$

Now, we can apply Lemma 87 with $f : \mathbf{A} \to \mathbf{B}$ to obtain $\mathbf{B} \vDash^{f \circ g \circ \iota} \phi$, and since $f \circ g = \mathrm{id}_\mathbf{B}$, we conclude $\mathbf{B} \vDash^{\iota} \phi$. $\square$

[172] Note that we already argued in Footnote 154 that the right inverse implies g is an isometric embedding. Then we could conclude by Corollary 94, and the proof given is essentially the same.

Remark 104. It is not true in general that the image $f(A)$ of a nonexpansive function $f : \mathbf{A} \to \mathbf{B}$ (seen as a subspace of $\mathbf{B}$) satisfies the same equations as $\mathbf{A}$. For instance, let $\mathbf{A}$ contain two points $\{a, b\}$ all at distance $1 \in [0, \infty]$ from each other (even from themselves). The $[0, \infty]$-relation is symmetric so it satisfies for all $\varepsilon \in [0, 1]$. $y =_\varepsilon x \vdash x =_\varepsilon y$. If we define $\mathbf{B}$ with the same points and distances except $d_\mathbf{B}(a, b) = 0.5$, then the identity function is nonexpansive from $\mathbf{A}$ to $\mathbf{B}$, but its image is $\mathbf{B}$ in which the distance is not symmetric.

Coproducts

Proposition 105. *The category* **GMet** *has an initial object.*

Proof. The initial object $\emptyset$ in $\mathsf{L}\mathbf{Spa}$ is the empty set with the only possible L-relation $\emptyset \times \emptyset \to \mathsf{L}$ (the empty function). The empty function $f : \emptyset \to X$ is always nonexpansive from $\emptyset$ to $\mathbf{X}$ because (50) is vacuously satisfied.

Just as for the terminal object, since **GMet** is a full subcategory of $\mathsf{L}\mathbf{Spa}$, it suffices to show $\emptyset$ is in **GMet** to conclude it is initial in this subcategory. We do this by showing $\emptyset$ satisfies absolutely all quantitative equations, and in particular those of E. This is easily done because when $\mathbf{X}$ is not empty,[173] there are no assignments $\mathbf{X} \to \emptyset$, so $\emptyset$ vacuously satisfies $\mathbf{X} \vdash x = y$ and $\mathbf{X} \vdash x =_\varepsilon y$. $\square$

[173] The context of a quantitative equation cannot be empty because the latter must come with some elements of the context.

Proposition 106. *The category* $\mathsf{L}\mathbf{Spa}$ *has all coproducts.*

Proof. We just showed the empty coproduct (i.e. the initial object) exists. Let $\{\mathbf{A}_i = (A_i, d_i) \mid i \in I\}$ be a family of L-spaces indexed by a non-empty set I. We define the

L-space $\mathbf{A} = (A, d)$ with carrier $A = \coprod_{i \in I} A_i$ (the disjoint union of the carriers) and L-relation $d : A \times A \to \mathsf{L}$ defined by:[174]

$$\forall a, b \in A, \quad d(a,b) = \begin{cases} d_i(a,b) & \exists i \in I, a, b \in A_i \\ \top & \text{otherwise} \end{cases}.$$

For each $i \in I$, we have the evident coprojection $\kappa_i : \mathbf{A}_i \to \mathbf{A}$ sending $a \in A_i$ to its copy in A, and it is nonexpansive because, by definition, for any $a, b \in A_i$, $d(a,b) = d_i(a,b)$.[175] We show $\mathbf{A}$ with these coprojections is the coproduct $\coprod_{i \in I} \mathbf{A}_i$.

Let $\mathbf{X}$ be some L-space and $f_i : \mathbf{A}_i \to \mathbf{X}$ be a family of nonexpansive maps. By the universal property of the coproduct in **Set**, there is a unique function $[f_i] : A \to X$ satisfying $[f_i] \circ \kappa_i = f_i$ for all $i \in I$. It remains to show $[f_i]$ is nonexpansive from $\mathbf{A}$ to $\mathbf{X}$. For any $a, b \in A$, suppose a belongs to A_i and b to A_j for some $i, j \in I$, then we have[176]

$$d_{\mathbf{X}}([f_i](a), [f_i](b)) = d_{\mathbf{X}}(f_i(a), f_j(b)) \leq \begin{cases} d_i(a,b) & i = j \\ \top & \text{otherwise} \end{cases} = d(a,b). \qquad \square$$

[174] In words, $\mathbf{A}$ is the L-space with a copy of each $\mathbf{A}_i$ where the L-relation sends two points in different copies to $\top$ (intuitively, the copies are completely unrelated inside $\mathbf{A}$).

[175] Each coprojection is even an isometric embedding.

[176] The first equation holds by definition of $[f_i]$ (it applies f_i to elements in the copy of A_i). The inequation follows by nonexpansiveness of f_i which is equal to f_j when $i = j$. The second equation is by definition of d.