Lifting Algebraic Reasoning to Generalized Metric Spaces

THÈSE

En vue de l'obtention du grade de Docteur, délivré par
L'École Normale Supérieure de Lyon

ED512
École Doctorale en Informatique et Mathématique de Lyon
Discipline: Informatique

Sera soutenue publiquement le 3 septembre 2024, par

RALPH SARKIS

Devant le jury composé de:

Gordon PLOTKIN	Professeur, University of Edinburgh	Rapporteur
Jiří ADÁMEK	Professeur, Czech Technical University in Prague	Rapporteur
Christine TASSON	Professeure, ISAE Supareo	Examinatrice
Rory B. B. LUCYSHYN-WRIGHT	Professeur associé, Brandon University	Examinateur
Matteo MIO	Chargé de recherche, CNRS & ENS de Lyon	Directeur
Valeria VIGNUDELLI	Chargée de recherche, CNRS & ENS de Lyon	Co-encadrante

Abstract

Algebraic reasoning is ubiquitous in mathematics and computer science, and it has been generalized to many different settings. In 2016, Mardare, Panangaden, and Plotkin introduced quantitative algebras, that is, metric spaces equipped with operations that are nonexpansive relative to the metric. They proved counterparts to important results in universal algebra, and in particular they provided a sound and complete deduction system generalizing Birkhoff's equational logic by replacing equality with equality up to ε. This allowed them to give algebraic axiomatizations for several important metrics like the Hausdorff and Kantorovich distances.

In this thesis, we make two modifications to Mardare et al.'s framework. First, we replace metrics with a more general notion that captures pseudometrics, partial orders, probabilistic metrics, and more. Second, we do not require the operations in a quantitative algebra to be nonexpansive. We provide a sound and complete deduction system, we construct free quantitative algebras, and we demonstrate the value of our generalization by proving that any monad on generalized metric spaces that lifts a monad on sets can be presented with a quantitative algebraic theory. We apply this last result to obtain an axiomatization for the Łukaszyk–Karmowski distance.

Résumé

On retrouve le raisonnement algébrique partout en mathématique et en informatique, et il a déjà été généralisé à pleins de contextes différents. En 2016, Mardare, Panangaden et Plotkin ont introduit les algèbres quantitatives, c'est-à-dire, des éspaces métriques équippés d'opérations 1-lipschitzienne relativement à la métrique. Ils ont prouvées des homologues à des résultats importants en algèbre universelle, et en particulier ils ont donné un système de deduction correct et complet qui génralise la logique équationelle de Birkhoff en remplçant l'égalité par l'égalité à ε près. Ça leur a permis de donné une axiomatisation algébrique pour quelques métriques importantes comme la distance de Hausdorff et celle de Kantorovich.

Dans cette thèse, on modifie deux aspects du cadre de Mardare et al. Premièrement, on remplace les métriques par une notion plus générale qui englobe les pseudométriques, les ordres partiels, les métriques probabilistes, entre autres. Deuxièmement, on n'exige pas que les operations de nos algèbres quantitatives soient lipshitzienne. On donne un système de deduction correct et complet, on construit les algèbres quantitatives libres, et on démontre la valeur de notre généralisation en prouvant que toute monade sur les éspaces métriques généralisés qui est le relèvement d'une monade finitaire sur les ensembles peut être présentées par une théorique algébrique quantitative. On applique ce dernier résultat pour obtenir une axiomatization de la distance de Łukaszyk–Karmowski.

Contents

Preface

Tamacun

Rodrigo y Gabriela

In place of the traditional citations as epigraphs at the start of every chapter, I put (links to) music I enjoyed listening to while writing this manuscript.

This document was not optimized for printing. The two main reasons are:

1. I use a slightly customized version of the `tufte-book` document class [KWG15]. This puts the main body of text closer to the left margin and all footnotes[0] in the right margin. This allows me to use a lot of footnotes throughout the text. I use them as if they were big parentheses, to add details, to digress, to add references, or to display diagrams. Printing with these margins can be complicated, and the text in the margins is a bit smaller.

[0] Like this one.

2. I use the **`knowldege`** package [Col24]. This allows me to easily add hyperlinks towards the definition of a symbol or a term every time I use that symbol or term. In particular, if you want to start reading at say Chapter 3, you do not even have to go over the notation introduced earlier, you can simply click on a symbol or word you don't recognize to see how it was defined. What is more, in the appendix, I put a draft of a book on category theory that I am writing, so there is no background section on categories, but every time I use a notion from that book (e.g. $\mathrm{Hom}_{\mathbf{C}}(A, B)$, functor, natural transformation), the `knowldege` link will go there.[1] Combined with the links to results, equations, and references (like Theorem 3.66, (3.12), and [MPP16]) there are more than twenty thousand links in this document.

[1] Test the links right now! Some pdf viewers are better than others to navigate a document with lots of links. Most have a navigation history so you can follow a sequence of links and get back to your original position by e.g. pressing `Alt+P` or the back button on a mouse. Some viewers also display a preview of the target of a hyperlink when you hover it, so there is no need to click.

With that said, if you would rather read on paper, I do not think there will be major difficulties since there is adequate numbering throughout the main text. However, I suggest you do not print the appendix (which is longer than the main text) because the links to the appendix are rarely numbered, and I did not make an index.

Notations and Conventions

Here are several standard and non-standard notations and conventions that I use throughout the text.

- Starting now, we will use the pronoun "we" when referring to us, author and readers. Occasionally, "I" will be used to refer to me, and "we" will be used to refer to me and my supervisors Matteo and Valeria.

- We use the following abbreviations:[2]
  - **I.H.** indicates a step of a proof that relies on the induction hypothesis (which is often left implicit).
  - **resp.** stands for "respectively".
  - **`L.H.S.`** stands for "left hand side" (of an equation usually).
  - **`R.H.S.`** stands for "right hand side".
- When defining a function $f : A/\sim \to B$ whose domain is a quotient by giving a value for $f(a)$ for each $a \in A$, we say it is **well-defined** if $f(a) = f(a')$ whenever $a \sim a'$.
- We sometimes have to deal with **proper classes**, i.e. collections of things that cannot be sets. We use classes to mean a collection that is either a set or a proper class.[3]
- We use the term **classical** to refer to universal algebra (the subject of Chapter 1),[4] usually in opposition to universal quantitative algebra (the subject of Chapter 3).
- The two principal references [MPP16, FMS21] are given special colors to help recognizing them when reading.

[2] Paired with a `knowldege` link going to this list.

[3] Really the only reason we need classes is for the collection of all sets, so nothing very fancy.

[4] In opposition to universal quantitative algebra (the subject of Chapter 3).

Acknowledgements

0 Introduction

> Across the Stars
>
> John Williams and the London Symphony Orchestra

Most programmers write code **compositionally**.[5] They write small lines of code that combine to make small functions that combine to make small files that combine to make a complete software. When studying the semantics of programs, we sometimes like to model these *combination* steps with algebraic operations.

[5] Some don't (e.g. code golfers).

This idea seems to originate in [SS71] and [GTWW77], and it continues to reverberate in current research, e.g., [TP97, HHL22, GMS+23]. It is referred to as **algebraic semantics**. We give only an informal account here to motivate the mathematics behind it.

If P, Q and Q' are programs, we can use $P;Q$ to represent the program that runs P then Q, and $\mathsf{ifte}(P,Q,Q')$ to represent the program that runs P, then runs Q if the Boolean value of the output of P was `True` or Q' if it was `False`. We view the set of programs as an algebra where instead of the well-known operations like addition and multiplication, many new operations are allowed to combine programs. The set of available operations varies with the kind of programs that are studied, it is called the signature, and we say that operations in the signature are interpreted in the algebra of programs.

Furthermore, the set of behaviors of programs[6] is also seen as an algebra for the same signature. Then, semantics is represented by a function from programs to behaviors which preserves the operations, namely, the combination of behaviors is the behavior of the combination. It is a homomorphism of algebras.

[6] The word behavior can be understood in many different ways that depend on what properties of the programs one is interested in.

Oftentimes, one realizes that two different programs have the same behavior, for example $P;(Q;R)$ and $(P;Q);R$ or $P;Q$ and $\mathsf{ifte}(P,Q,Q)$, so they should be considered **equal** (or **equivalent**). The bread and butter of algebraic semanticians is to find a (sound and complete) collection of simple equations (axioms) that make it possible to reason compositionally about program equivalence.[7] Sometimes these axiomatizations help in designing (semi)-automatic procedures to answer the question "is P equal to Q?".

[7] For instance, with the equations above, we can infer that $\mathsf{ifte}(P,Q;(R;S),(Q;R);S)$ and $P;(Q;(R;S))$ are equivalent.

A famous example is combinatory logic, originating in [Cur29], which gives a computational model as powerful as the pure λ-calculus using four operations to combine programs and three equations between small programs.[8] In this thesis, we detail two other well-known examples that model nondeterministic and probabilistic

[8] see, e.g., [Mim20, §3.6.3].

choices in Examples 1.66 and 1.67 respectively.

Much of the work on algebraic semantics relies on the theoretical foundations of **universal algebra**, an old subject popularized by Birkhoff in [Bir33, Bir35]. Three of his major results are:

1. a logical system, called equational logic (Figure 1.3), that allows one to syntactically derive which equations are entailed by a set of axioms,
2. the construction of free algebras (Definition 1.25 and Proposition 1.40), and
3. the HSP (or variety) theorem [Wec12, §3.2, Theorem 21] which characterizes classes of algebras that can be defined with equations.

There is also tight connection between universal algebra and **monads** on **Set** (Definition 1.50) that can be exploited to study semantics with algebraic and categorical reasoning. For instance, nondeterminism can be modelled with the theory of semilattices and the powerset monad (Example 1.66), and probability can be modelled with the theory of convex algebras and the distribution monad (Example 1.67). These two examples and similar ones show up very often in the study of program semantics.[9]

[9] See, e.g., [PP01a, PP01b, PP02, BP15, BSS21, BSV22].

Since computers interact with humans (or the other way around), it makes sense to take into account the quirks of a human mind when studying the behavior of programs. For example, many standard data compression algorithms (in particular for image, audio, and video) are efficient at the cost of losing some small amount of information.[10] In that situation (and others like it), program equivalence is too coarse of a relation, so researchers have to build more sensitive models to handle and compare **approximations** of programs.

[10] Usually, users will not notice nor mind because of the inherent information degradation in the human perception process [SB06].

This makes the case for developing **quantitative algebraic semantics**. We view the set of programs as an algebra (we can still combine them) with a notion of distance (we can now compare them more finely than with equality). Intuitively, the distance between P and Q shall reflect the disparity in their behaviors, hence, the behaviors must come with a notion of distance too. For example, if P is a lossless compression algorithm, and Q is a lossy one, the distance between P and Q may be the fraction of the inputs (picked in a real-world dataset) wherein the outputs of P and Q noticeably differ.[11]

[11] For metrics actually used in practice, see [LJ11].

We most commonly think of a distance as a number, but our formalization of distances (Definitions 2.11 and 2.30) will accomodate a large array of things to call distances, see Examples 2.13–2.15.

If the field of algebraic semantics founds itself on universal algebra, there needs to be a quantitative version of this theoretical basis to support research in quantitative algebraic semantics.

The concept of extending algebraic reasoning to diverse settings is by no means novel, as evidenced by the following (inevitably) non-exhaustive list of references: [Dub70, BD80, KP93, Wea93, GP98, Pow99, Rob02, Pow05, VK11, FH11, AMMU15, LW16, MU19, BG19, FMS21, Ros21, LP23, RT23, Ros24]. While these approaches excel in their generality and abstraction, it is at the cost of usability, even for someone who is already familiar with universal algebra. More concrete solutions exist. We mention two that seem to be of particular interest to computer scientists.

If we equip the algebra of programs with a partial order, the question "is P equal to Q?" becomes "is P less than Q?".[12] There is already a lot of work in universal algebra on partial orders [Blo76, ANR85, KV17, AFMS21, FMS21, ADV22, Sch22a, Sch22b].

[12] The meaning of $P \leq Q$ depends on what kind of programs and properties are studied.

If we equip the algebra of programs with a metric space, the question "is P equal to Q?" becomes "are P and Q closer than ε from each other?", where ε is a real number. There is already a lot of work in universal algebra on metric spaces [Wea95, MPP16, Hin16, MPP17, BMPP18, BBLM18b, MPP18, MV20, BMPP21, MPP21, Ros21, MSV21, Adá22, MSV22, MSV23, ADV23, Ró24].

In this thesis, we make another attempt to genearlize algebraic reasoning without straying too far from the classical setting. Our main inspirations are [MPP16], the seminal paper on quantitative algebras, and [FMS21], a vast generalization.[13]

[13] I gave their references special colors to help recognizing them when reading.

In [MPP16], the authors study algebras equipped with a metric such that the interpretation of operations in the signature are nonexpansive. More precisely, they are metric spaces (A, d) with, for each n-ary operation op in the signature, an interpretation $[\![\mathsf{op}]\!] : A^n \to A$ satisfying

$$\forall a, b \in A^n, d([\![\mathsf{op}]\!](a_1, \dots, a_n), [\![\mathsf{op}]\!](b_1, \dots, b_n)) \leq \max_{1 \leq i \leq n} d(a_i, b_i). \tag{0.1}$$

This is a very natural condition because it is equivalent to saying that $[\![\mathsf{op}]\!]$ is a morphism from $(A, d)^n$ to (A, d) in the category **Met** of metric spaces and nonexpansive maps, where $(A, d)^n$ denotes the n-wise categorical product.[14]

[14] I would say this is the expected definition of "algebra over a metric space", especially to those familiar with functorial semantics [Law63], or subsequent work in categorical algebra and categorical logic.

In [FMS21], the authors view **Met** as an instance of a category $\mathbf{Str}(\mathscr{H})$ of relational structures, see [FMS21, Example 3.5.(3)]. Without going into details, we can mention that the category **Poset** of partially ordered sets and monotone maps is another instance. Therefore, their work is general enough to cover both algebras equipped with a metric and algebras equipped with a partial order. Accordingly, a generalization of (0.1) is imposed on the interpretation of operations, namely, $[\![\mathsf{op}]\!]$ is a morphism from A^n to A, where $A \in \mathbf{Str}(\mathscr{H})$.[15]

[15] It is actually more complicated than that, because in [FMS21], operations come with an arity $\mathsf{ar}(\mathsf{op})$ that is not just a natural number but a whole relational structure itself (with some size conditions). This allows them to handle some *partial* operations, e.g., $x + y$ can be undefined.

In both papers, there is a sound and complete logical system that generalizes Birkhoff's equational logic, [MPP16] replaces equations with *quantitative inferences* and [FMS21] replaces equations with Σ-*relations* (where Σ is the signature). An explicit construction of free algebras equipped with a metric (resp. a relational structure) is given in [MPP16, Theorem 5.3] (resp. [FMS21, Theorem 4.18]). Later papers provided generalizations of the HSP theorem [MPP17, MU19, JMU24], and the connection with monads has been investigated in [FMS21, Adá22, ADV23].

In [MPP16, §8–10], the authors use their logic to axiomatize well-known constructions on metrics. They show that the total variation distance (Example 3.78), the Kantorovich distance (Example 3.5), and the Hausdorff distance (Example 2.16) can all be defined as free algebras for some carefully chosen set of axioms. Ford et al. do the same for the metric completion in [FMS21, Example 4.8]. Many other so-called presentation results are found in, e.g., [MV20, BMPP21, MSV21, MSV22], sometimes with applications to semantics.

While trying to axiomatize other interesting distances, we had to question some assumptions of [MPP16]. We learned about the ŁK distance (3.3) on probability distributions in [CKPR21], where they use it as an easier-to-compute alternative to the Kantorovich distance. We intended to simply adapt that axiomatization, but we quickly faced two obstacles.

First, the ŁK distance is not a metric, it is a diffuse metric [CKPR21, §4.2].[16] In particular, the distance between a distribution and itself can be non-zero. Second, combining probability distributions like it is done for the Kantorovich distance (with convex combinations) is not nonexpansive in the sense of (0.1) with d being the ŁK distance.

[16] That is a relaxation of the usual axioms for metrics (see Definition 0.1). Diffuse metrics are also called dislocated metrics in [HS00].

The generality of [FMS21] is enough to overcome the first problem since the category of diffuse metrics is an instance of $\mathbf{Str}(\mathscr{H})$. However, we already said that they also work with (an analog to) the requirement of (0.1), so the second problem remains. In the present work, we introduce a framework that deals with both 1) distances that are not metrics and 2) operations that do not satisfy (0.1).[17] Rejecting that assumption was previously done in [Wea93, Wea95, Hin16, Hin17, BBLM18a, AFMS21] in various different contexts.

[17] The first time we did this was in [MSV22], and with [MSV23] and this thesis, we aim to simplify and broaden our initial proposal.

We define generalized metrics to be distance functions valued inside an arbitary complete lattice L ($d : A \times A \to \mathsf{L}$) satisfying an arbitrary set of axioms expressed with quantitative equations (a variant of the quantitative inferences in [MPP16]).[18] Then, our quantitative algebras (Definition 3.1) are simply algebras equipped with a generalized metric. Importantly, no further restriction is imposed on the operations in the algebra, and this allows us to axiomatize the ŁK distance in Example 3.88.

[18] In particular, taking $\mathsf{L} = [0, \infty]$ with the axioms of Definition 0.1 translated into quantitative equations yields metrics (see Example 2.32).

With this setting, we recover some of the classical results in universal algebra, and more. The major contributions, Items i., ii., and iv., already appear in [MSV23] with a different presentation and a fixed $\mathsf{L} = [0, 1]$.

i. We define quantitative equational logic (Figure 3.1), a logical system that is sound (Theorem 3.55) and complete (Theorem 3.62) relative to our quantitative algebras. It mirrors equational logic more closely than Mardare et al.'s logic[19] without renouncing their fundamental idea to *merely* change equality with equality up to ε.

[19] See the discussion in §0.3 and Example 3.56.

ii. We construct the free quantitative algebras (Theorem 3.48) relative to any class of quantitative equations.[20] This induces a monad on the category $\mathbf{GMet}$ of generalized metric space, and the quantitative algebras modelling the chosen class of quantitative equations coincide with the algebras for that monad (Theorem 3.66).

[20] We give a semantical and a syntactical construction (Definitions 3.37 and 3.59 respectively), and they are equivalent thanks to soundness and completeness of our logic.

iii. We provide a simple axiomatization of the set of probability distributions with the ŁK distance as a free quantitative algebra in Example 3.88.

iv. In achieving Item iii., we prove a more general result (Theorem 3.84) which states that any monad lifting to $\mathbf{GMet}$ (Definition 3.73) of a monad on $\mathbf{Set}$ with an algebraic presentation also has a quantitative algebraic presentation (Definition 3.68), i.e. it can be axiomatized with quantitative equations.[21] In

[21] This is related to [ADV22, §5], where *strongly finitary* monads on the category of posets are shown to be monad liftings of *finitary* monads on $\mathbf{Set}$.

particular, it yields a presentation for a monad on **Met** that is not captured by the framework of [MPP16] nor that of [FMS21] (Remark 3.87).

Apart from those technical contributions, our approach describes quantitative algebraic reasoning as a cleaner generalization of algebraic reasoning.[22] This guided the outline of this manuscript which is divided in three chapters, one on classical algebraic reasoning, one on our tailored generalization of metric spaces, and one on combining these two chapters, **lifting algebraic reasoning to generalized metric spaces**. Let us now give more detailed introductions for each of these chapters.[23]

[22] This is supported by Item iv. and Examples 3.56 and 3.57.

[23] You could skip these now and come back to each of the following sections when starting to read the corresponding chapter.

0.1 Universal Algebra and Monads

With a bit of experience adding natural numbers together, you quickly notice that addition respects some *rules*. If you add n and m, you get the same thing as if you add m and n, no matter what numbers n and m are. If you add n and 0, you obtain n. If you add n and m, then add k, you get the same thing as if you add n to the sum of m and k. We represent these rules with **equations**:

$$n + m = m + n \qquad n + 0 = n \qquad (n + m) + k = n + (m + k). \tag{0.2}$$

These equations also hold when n, m, and k belong to the integers or the real numbers. We can also replace addition with multiplication and 0 with 1.

Since these rules apply in different contexts, mathematicians came up with an abstract definition of a commutative monoid: a set M with a function $+ : M \times M \to M$ (written infix) and an element $0 \in M$, such that for all $n, m, k \in M$, the equations above are true. The study of these abstract structures (and other variants like groups and rings) is extremely fruitful,[24] so much so that you probably learned about them in a first-year undergraduate mathematics course with "algebra" in its title.

[24] I cannot do better than a euphemism here. Even narrowing to theoretical computer science, algebraic reasoning has many applications — there are two noteworthy international conferences with "algebra" and "computer science" in their names, CALCO [GS21] and RAMiCS [FGSW21]. Our story focuses on algebraic semantics only.

With a bit of experience studying monoids, groups, and rings, you quickly notice the similarities in their definitions, and in the reasoning in proofs about them. The purpose of universal algebra is to formalize what they have in common, in order to investigate them all at once. We study an arbitrary **algebraic theory** instead of doing group theory, ring theory, etc.

An algebraic theory is a syntactic gadget that specifies one kind of algebraic structure with a signature Σ containing operation symbols, and a collection of equations E asserting that some sequences of symbols can be replaced by others. For instance, the theory of commutative monoids contains the symbol $+$, the symbol 0, and the equations in (0.2).

The models of a theory derived from (Σ, E) are called (Σ, E)-algebras. They are sets in which you can combine elements as dictated by the operations in Σ in a way that respects the rules expressed by the equations in E. For instance, the models of the theory of commutative monoids are commutative monoids.

The flexibility of universal algebra was recognized as a powerful tool early on in the history of formal semantics of programming languages (at the least in [SS71]). We already saw that sequential composition ; and conditional branching ifte could be modelled as algebraic operations. Let us mention two additional well-known examples which were the main source of examples for quantitative algebras.[25]

To represent programs that use nondeterminism, we use a binary operation $\oplus$. If P and Q are programs, then $P \oplus Q$ nondeterministically chooses to run P or Q. The equations that govern the behavior of $\oplus$ are

$$P \oplus P = P, \quad P \oplus Q = Q \oplus P, \text{ and } P \oplus (Q \oplus R) = (P \oplus Q) \oplus R.$$

Briefly, they state that a nondeterministic choice is not affected by the order or multiplicity of the possibile outcomes.[26]

To represent programs that make decisions according to some probability distributions, we use a family of binary operations $+_p$ indexed by real numbers $0 < p < 1$. If P and Q are programs, then $P +_p Q$ is the program that runs P with probability p and Q with probability $1 - p$. For example, if P and Q return `HEADS` and `TAILS` respectively, then $P +_{0.5} Q$ is a fair coin. The equations look a lot like those for $\oplus$, for example $P +_p P = P$ for any p.[27]

To fully grasp the last sentence of the paragraph on nondeterminism, it is crucial to note that the three equations we gave entail many more equations (for example $(Q \oplus P) \oplus (P \oplus Q) = P \oplus Q$). We can appreciate this from two equivalent angles. Semantically, an equation ϕ is entailed by a set of equations E if all the models of (Σ, E) satisfy ϕ. Syntactically, ϕ is entailed by E if it can be derived in equational logic (see Figure 1.3).[28]

Yet another take on algebraic theories comes from category theory. Birkhoff [Bir35] had already realized that one can always freely generate (Σ, E)-algebras, and Lawvere [Law63] and Linton [Lin66] recognized this induces a monad $\mathcal{T}_{\Sigma,E}$ on the category of sets. They also showed there is a (partial) converse: any *finitary* monad on **Set** is presented by an algebraic theory.[29]

Moggi first conveyed the applicability of monads (an abstract notion from category theory) in computer science in [Mog89, Mog91]. They became a valuable tool in semantics, and a monad paired with an algebraic presentation allows to combine categorical and equational reasoning. It can be very effective (even outside semantics) as shown in, e.g., [PP01a, PP01b, PP02, BP15, BHKR15, DPS18, PRSW20, BSS21, BSV22, ZM22, RZHE24].

In Chapter 1, we tell the story many times retold of universal algebra. We adopt a somewhat peculiar presentation of the material in order to replicate it more accurately in Chapter 3. We also give some examples of algebras, algebraic theories and algebraic presentations.

0.2 Generalized Metric Spaces

In many applications, deciding whether programs are equivalent or not is overly simplistic. We gave the example of compression algorithms, but let us give three more.

Artificial Intelligence. A lot of models in AI, especially in machine learning, rely on probabilistic reasoning to make decisions.[30] For example, when a classifier is

[25] See [BSV22] for a more detailed account in the classical setting, and [MSV21] for the quantitative setting.

[26] See Example 1.66.

[27] See Example 1.67.

[28] The first account of this logic, and the equivalence between these two points of view are due to Birkhoff [Bir35], and we prove it in Theorems 1.43 and 1.48.

[29] i.e. any finitary monad is isomorphic to $\mathcal{T}_{\Sigma,E}$ for some Σ and E.

[30] See, e.g., [CKPR21] which motivated Example 3.88.

fed an image, before deciding what the image depicts, it produces a probability distribution over things that could possibly be in that image. It goes like this:

$$\texttt{dogpic.jpg} \xrightarrow{\text{classify}} 89\%\ \text{dog} + 6\%\ \text{lion} + 2\%\ \text{cat} + \cdots \xrightarrow{\text{max}} \text{dog}.$$

Consider two different classfiers that consistently give the same (possibly correct) answer on the testing dataset. One might consider them to be equal, but a closer examination could reveal that one classifier is more confident than the other. In other words, the distributions produced by one classifier may be more concentrated than those produced by the other.[31] Therefore, it makes sense to compare classifiers (more generally, AI models) by devising a notion of distance on the probability distributions they produce. We will give two examples of distances between distributions within our framework in Examples 3.71 and 3.88.

[31] In particular, the distributions produced by the perfect classifer (one that knows the correct labels) are always fully concentrated at a single point.

Quantitative Information Flow. When designing software that handles data containing private information, one often wants a balance between the privacy of the users and the utility provided. It makes sense to share the average grade for a class of 100 students, but not for a class of 5 students. With the methods developed in quantitative information flow [QIF20] (especially differential privacy [Dwo06]), we can compare the levels of confidentiality of different programs, before deciding what is the safest (most private) one to roll out.[32]

[32] For now, this is only a potential application, we do not have concrete results in this direction.

Code Optimization. Consider the two pieces of pseudocode in Figure 1.

```
return Bernoulli(0.5)
```

```
do
    x = Bernoulli(0.3)
    y = Bernoulli(0.3)
while (x == y)
return x
```

Figure 1: Simulating a fair coin flip with a biased coin (with a weak guarantee of termination).

For all intents and purposes, they are equivalent.[33] However, there is only a weak guarantee that the second program terminates (it does with probability 1). Still, if you are unable to run `Bernoulli(0.5)` for some reason, you would be perfectly happy to use the second program. If you want to have a strong guarantee of termination, you could interrupt the loop after, say, 1000 iterations and then return an arbitrary value (see Figure 2). Unfortunately, this breaks the equivalence with `Bernoulli(0.5)`, but

[33] If you throw a (possibly biased) coin twice and you get two different outcomes, the probability that the first outcome was `HEADS` is equal to the probability that it was `TAILS`, hence it is 0.5 (assuming throws are independent).

```
i = 0
do
    i = i + 1
    x = Bernoulli(0.3)
    y = Bernoulli(0.3)
while (x == y) OR i >= 1000
return x
```

Figure 2: Simulating a fair coin flip with a biased coin (with a strong guarantee of termination).

it is still appropriate to say that the two programs are close to each other (even if they are not equivalent), and that they would be even closer if we increase the maximum number of iterations. When some features are not available, or more realistically when their implementation is not efficient, it can be convenient to write code that approximates the specification but runs (faster).

A widespread alternative to equality that is inherently more fine-grained is metrics. The first definition of metric space (under the name "(E) classes") is credited to Fréchet's thesis [Fré06]. We give the definition that is now standard.[34]

[34] Up to small variations. It is essentially equivalent to Fréchet's definition, but uses different notation and terminology.

Definition 0.1 (Metric space)**.** A **metric space** is a pair (A, d) comprising a set A and a function $d : A \times A \to [0, \infty)$ called the metric satisfying for all $a, b, c \in A$:

1. separation: $d(a, b) = 0 \Leftrightarrow a = b$,
2. symmetry: $d(a, b) = d(b, a)$, and
3. triangle inequality: $d(a, c) \leq d(a, b) + d(b, c)$.

For more than a 100 years now, metrics have been a good abstract formalization of what we intuitively understand to be **distances**. In particular, $d(a, b)$ is often called the distance between a and b. Therefore, instead of reasoning about program equivalence, we reason about **program distances**.[35]

[35] In semantics, people also use the term behavioral distance/metric.

The study of distances between programs (especially those with probabilistic aspects) began in the previous century (see [vB01] for a (relatively old) survey). While there is no international conference on the subject,[36] it is still a very active area of research (see, e.g., [CDL15, CDL17, BBLM18a, BMPP18, BBLM18b, BBKK18, MSV21, Pis21]).

[36] Work on this definitely fits in QAPL, but the last meeting was in 2019 [AW20].

In this literature, there is a recurring idea that positive real numbers are not always the best space to value distances in. Oftentimes, the value ∞ is allowed, where $d(a, b) = \infty$ means a and b are as far apart as they can be. Sometimes, distances are bounded above by 1, so $[0, 1]$ replaces $[0, \infty)$. In more exotic cases, it makes sense for $d(a, b)$ to not even be a number, it can be a set [ABH$^+$12], a probability distribution [HR13], an element of a continuous semiring [LMMP13], or just a boolean value.

It is also common to remove or modify some axioms of Definition 0.1 to work with, e.g., pseudometric spaces [BBKK18] or ultrametric spaces [Esc99, Pis21].

It would be ideal if we could devise a definition that encapsulates all existing formal notions of distance. That is obviously not possible. Moreover, even the term "generalized metric space" is employed across various research communities with different meanings (see, e.g., [BvBR98, Bra00, LY16, Pis21]).

In [MPP16], the authors propose theoretical foundations for quantitative algebraic semantics. Their work allows to reason equationally about metrics. One of our contributions in [MSV22] was to show that you can handpick any subset of the axioms of metric spaces and carry out all the proofs of the original paper [MPP16] without much trouble.[37] I believe this was known to the authors of [MPP16], especially in light of the results of [FMS21] which morally do the same thing except for an even more general class of structures.

[37] Although there is a subtelty about the equality predicate that we explain in §0.3.

In this thesis, we propose yet another definition of generalized metric spaces that is as general as possible without requiring any additional technical machinery. In fact, if you read the present work being comfortable with the frameworks presented in [MPP16] or [MSV22], I believe you will not feel far from home.

We first define L-spaces (Definition 2.11) which are sets equipped with a distance function into a complete lattice L $(A, d : A \times A \to \mathsf{L})$. The structure of a complete lattice allows to compare distances (say one is smaller or bigger than another), and to define a distance as an infimum of a set of bounds in L. That is enough to do quantitative algebraic reasoning in the sense of [MPP16].[38]

[38] We say more on this in Remark 2.21.

Then, we describe a language to specify axioms one can put on L-spaces. We call such axioms quantitative equations (Definition 2.22). They are a restriction of quantitative equations that we define in Chapter 3, so we will motivate them in §0.3. Examples include separation, symmetry and triangle inequality from Definition 0.1, but also reflexivity, transitivity, and antisymmetry of a binary relation, the strong triangle inequality of ultrametric spaces, and many more. A generalized metric space is then an L-space that satisfies a fixed set of quantitative equations.

In Chapter 2, we give lots of examples including posets, preorders, metrics, pseudometrics, ultrametrics, etc. We also study some properties of the categories of generalized metric spaces[39] in preparation for Chapter 3 which essentially just combines the first and second chapter to do universal algebra on generalized metric spaces.

[39] We get one category **GMet** for each complete lattice L and each collection of quantitative equations we decide to impose.

0.3 Universal Quantitative Algebra

The term *quantitative* is used in this thesis to refer to a notion of distance that *quantifies* how far apart two things are.[40] Universal quantitative algebra is then a framework where one can reason about both equality and distances between algebraic terms (built out of variables and operations in a signature). The first paper on the subject is [MPP16]. Its theoretical contributions are three-fold.[41]

[40] In constrast with the work on Girard's *quantitative semantics* [Gir88, BE99] or Kesner and Ventura's *quantitative types* [KV14] which aim to quantify the resource usage of a program.

[41] The authors also admirably sell their results with several examples combining algebraic and metric reasoning to axiomatize well-known metrics, the Hausdorff distance which we treat more generally in Example 3.69, the Kantorovich distance (Example 3.5), and the total variation distance (Example 3.78).

The authors work in the category **Met** of extended metric spaces (distances valued in $[0, \infty]$) and nonexpansive maps — a function is nonexpansive if it never increases the distance of its inputs (2.3). First, they define a **quantitative algebra** to be a metric space (A, d) equipped with operations that are interpreted as nonexpansive functions $(A, d)^n \to (A, d)$, where $(A, d)^n$ denotes the n-wise categorical product of (A, d) with itself. Second, they develop an analog to Birkhoff's equational logic to reason about properties of quantitative algebras, and they show it is sound and complete. Third, they show that free quantitative algebras always exist.

Let us briefly explain the logic presented in [MPP16]. At its core, there is the neat observation that the data of a metric $d : A \times A \to [0, \infty]$ can be equivalently given as a family of binary relations $\{R^d_\varepsilon \subseteq X \times X\}$ indexed by $\varepsilon \in [0, \infty]$ with some additional properties.[42] This point of view is not completely new, it can be glimpsed in [Wea95], [DLPS07, §1.2], [Ngu10, After Proposition 1], and [Con17]. However, in their quantitative equational logic, the authors of [MPP16] propose to take more seriously the point of view that the relation R^d_ε means "equality up to ε", and thus

[42] We prove a more general version in Proposition 2.20.

that we can reason about it kind of how we do for equality. In particular, they use the symbol $=_\varepsilon$.

Their logic closely ressembles implicational logic (see, e.g., IL1–8 in [Wec12, p. 223–224]) where the equality predicate is replaced by a family of predicates $=_\varepsilon$ where ε is a positive real number.[43] The meaning of $s =_\varepsilon t$ is that for all possible assignments of variables, the interpretations of s and t are at distance at most ε. It is clearly reminiscent of the meaning of $s = t$ in universal algebra (that for all possible assignments of variables, the interpretations of s and t are equal). The shape of a generic judgment, called quantitative inference, is $\{s_i =_{\varepsilon_i} t_i\} \vdash s =_\varepsilon t$. It asserts that whenever the distance between the interpretations of s_i and t_i are below ε_i for each $i \in I$, the distance between the interpretations of s and t is below ε.

[43] It is harmless to restrict to rational numbers if one cares about the size of the formal system.

Here are a few inference rules in that logic.

$$\frac{}{\vdash t =_0 t}\,\textsc{Refl} \qquad \frac{}{s =_\varepsilon t \vdash s =_{\varepsilon+\varepsilon'} t}\,\textsc{Max} \qquad \frac{\forall \phi \in \Gamma', \Gamma \vdash \phi \qquad \Gamma' \vdash \psi}{\Gamma \vdash \psi}\,\textsc{Cut}$$

$$\frac{}{\{s_i =_\varepsilon t_i \mid 1 \leq i \leq n\} \vdash \mathsf{op}(s_1, \ldots, s_n) =_\varepsilon \mathsf{op}(t_1, \ldots, t_n)}\,\textsc{Nexp}$$

The first states that the distance between the interpretation of t and itself is always below 0 (hence equal to 0), this mirrors one side of the separation axiom of metric spaces. The second rule, quantified over all positive reals $\varepsilon, \varepsilon'$, states that if ε is an upper bound for the distance between (the interpretations of) s and t, then you can add any positive quantitity and it will remain an upper bound. The third is a cut rule that you always find in similar deductive systems,[44] and it simply reflects the semantics of $\vdash$ being an implication.

[44] c.f. IL7 in [Wec12, p. 224] and *cut* in [CM22b, Definition 4.1.1].

The last one states that whenever the distance between s_i and t_i is bounded above by ε for each $i \in I$, so is the distance between $\mathsf{op}(s_1, \ldots, s_n)$ and $\mathsf{op}(t_1, \ldots, t_n)$. After unrolling some definitions, one verifies this is equivalent to the interpretation of op being nonexpansive with respect to the product metric (0.1).[45]

[45] We mentioned this property of intrepretations is very natural, but so is the Nexp rule: it says that the relation $=_\varepsilon$ is preserved by the operations (like a congruence, except is is not necessarily an equivalence relation).

The quantitative equational logic that we present in Figure 3.1 is adapted from the one in [MPP16] in three key ways.

1. In order to deal with quantitative algebras on generalized metric spaces, the predicates $=_\varepsilon$ are now indexed with quantities $\varepsilon \in \mathsf{L}$, and the rules like Refl above are removed.[46] Without Refl, there is no predicate $=_\varepsilon$ that corresponds to equality. Thus, we have to reintroduce the predicate $=$, and add rules ensuring that it behaves like equality (it is a congruence).

[46] We also remove rules that ensure the other side of separation, symmetry and triangle inequality.

2. We remove Nexp. As we foreshadowed, this rule and the requirement of (0.1) are not necessary to develop the theory of quantitative algebras. We first showed this in [MSV22], where we replaced these with a technical notion we called lifted signatures [MSV22, Definition 3.6] and a corresponding inference rule. In [MSV23] and here, we do not replace them with anything as it makes the base logic simpler. It is always possible to recover the nonexpansive property (or its variants from [MSV22]) by adding more axioms (see (3.9)).

3. In an effort to make a better parallel with equational logic, we slightly reduce the

expressiveness of the logic. The authors of [MPP16] already identified a special class of judgments whose terms in the premises are all variables, that is, their generic shape is $\{x_i =_{\varepsilon_i} y_i\} \vdash s =_\varepsilon t$. They call these *basic quantitative inferences*,[47] and they crucially rely on them to define free algebras[48] [MPP16, Theorem 5.1], and to prove variants of the HSP theorem [MPP17, Theorem 3.11].

The premises of a basic quantitative inference (predicates to the left of the turnstile $\vdash$) can equivalently be described with an L-space on the variables used.[49] Thus, our generic judgments are now written like $(X, d) \vdash s =_\varepsilon t$ or $(X, d) \vdash s = t$, where (X, d) is an L-space, where d is the largest distance that models the premises of the corresponding basic quantitative inference. We call these judgments quantitative equations as we believe they are the proper counterpart to equations in universal algebra.

Recall that quantitative equations also generalize the axioms of generalized metric spaces from §0.2. More accurately, the quantitative equations of Chapter 2 are instances of the quantitative equations of Chapter 3 when the signature is empty. That is morally the reason why we define generalized metric spaces with them.[50]

The first and third item can both be found, under guise of further abstraction, in [FMS21]. They deal with relational structures which are more general, but harder to link back to the equational reasoning we are used to in universal algebra. Our main advantage is that, while we can handle various notions of distances that are not metrics (e.g. ultrametrics and partial orders), our logic is not more complicated than [MPP16]'s. In fact, in a sense it is simpler because it only deals with basic quantitative inferences, yet it is still sound and complete.[51]

To be impactful, one could say our logic is to [MPP16]'s logic as equational logic is to implicational logic. Indeed, what is a *basic* implication in implicational logic? It is a judgment of shape $\{x_i = y_i\} \vdash s = t$, where the terms in the premises are variables only. But this means the premises are trivial because if two variables are equal, you can use a single variable instead. Thus a basic implication is just an equation, and similarly, a basic quantitative inference is just a quantitative equation.

The second item seems to be novel. Although people had removed the nonexpansive requirement in [Wea93, Wea95, Hin16, Hin17, AFMS21],[52] nobody had done it in the logical apparatus. We were inspired by the ad-hoc approach of [BBLM18a].

Dismissing NEXP is necessary to prove Theorem 3.84, the main theorem in §3.4. The motivating applications of [MPP16] are presentation results for monads on **Met**. Briefly, they show how the distances induced by their logic (with different sets of axioms) coincide with popular distances used in semantics. Similar results were obtained in, e.g., [MSV21, BMPP18, BMPP21, MSV22], and they all have in common that they reuse a known algebraic presentation for a monad on **Set**. We show in Theorem 3.84 that this is always possible when the monad on **Met** is a monad lifting of the monad on **Set** (Definition 3.73).

When working with nonexpansive operations, or equivalently with the NEXP rule, the induced monads are automatically enriched.[53] We exhibit a monad lifting that is not enriched in Example 3.74, and it is presented by a quantitative algebraic

[47] They require the set of premises to be finite, but that is not important for us.

[48] Consequently, their examples of axiomatizations only use basic quantitative inferences.

[49] See the discussion on syntactic sugar before Remark 2.26. This idea also appears in, e.g., [AFMS21, FMS21, Adá22, ADV23].

[50] We took a less elegant but more pragmatic approach in [MSV23, §8].

[51] We say more on this in Remark 3.54.

[52] Unfortunately, we were not aware of these papers when we published [MSV22], and we did not cite them.

[53] This is proved in the metric context in [ADV23, Full version, after Corollary 4.19], in the ordered context in [AFMS21, Proposition 4.6], and in the context of relational structures in [FMS21, Corollary 4.14].

theory thanks to Theorem 3.84. This shows that our approach is more general (in one aspect) than [MPP16] and [FMS21].

A final benefit we can highlight is the way our simplifications make the story of universal quantitative algebra so similar to the story of universal algebra. In Chapter 3, the outline and many proofs from Chapter 1 are reprised to work with quantitative algebras. We also give some examples of quantitative algebras, quantitative algebraic theories, and quantitative algebraic presentations.

1 Universal Algebra

Concerto Al Andalus

Marcel Khalifé

For a comprehensive introduction to the concepts and themes explored in this chapter, please refer to §0.1. Here, we only give a brief overview.

In this chapter, we cover 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,[54] but for completeness (and to be honest my own satisfaction) we take our time with it, although we assume the reader is comfortable with basic category theory (the material in the appendix). In Chapter 3, 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 Chapter 3.[55]

[54] [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*).

[55] I hope this will not make this chapter too terse, but the payback of simply copy-pasting proofs to obtain the generalized results is worth it.

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 algebras for a given signature and class of equations. In §1.3, we give the rules for equational logic to derive equations from other equations, and we show it is sound and complete. In §1.4, we 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 important later.

1.1 Algebras and Equations

We said in §0.1 that groups and rings are both examples of algebras we want to understand. Groups and rings allow different kinds of combinations of elements, you can do $x \cdot (y+z)$ in a ring but not in a group. To specify which combinations are allowed, we use a signature, and essentially all of this chapter will be parametric over a signature Σ.

Definition 1.1 (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:[56]

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

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

An algebra for a signature Σ is a structure where each operation symbol in Σ is associated to a concrete way to combine elements.

Definition 1.2 (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 $\mathbb{A}$[57] or the pair $(A, [\![-]\!]_A)$, where $[\![-]\!]_A : \Sigma(A) \to A$ is the function sending $\mathsf{op}(a_1, \dots, a_n)$ to $[\![\mathsf{op}]\!]_A(a_1, \dots, a_n)$ (it compactly describes the interpretations of all symbols).

[57] 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 $a_1, \dots, a_n \in A$,[58]

$$h([\![\mathsf{op}]\!]_A(a_1, \dots, a_n)) = [\![\mathsf{op}]\!]_B(h(a_1), \dots, h(a_n)). \tag{1.2}$$

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

$$\begin{tikzcd} \Sigma(A) \arrow[r, "\Sigma(f)"] \arrow[d, "[\![-]\!]_A"'] & \Sigma(B) \arrow[d, "[\![-]\!]_B"] \\ A \arrow[r, "f"'] & B \end{tikzcd} \tag{1.1}$$

This amounts to an equivalent and more concise definition of $\mathbf{Alg}(\Sigma)$: it is the category of algebras for the signature functor $\Sigma : \mathbf{Set} \to \mathbf{Set}$ [Aw010, Definition 10.8].

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 $\mathbf{Set}$ 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 1.3. 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 1.4. 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 $[\![\mathsf{p}]\!]_A : A^0 \to A$. Since A^0 is the singleton $\mathbf{1}$, $[\![\mathsf{p}]\!]_A$ is just a choice of element in A,[59] so the objects of $\mathbf{Alg}(\Sigma)$ are pointed sets (sets with a distinguished element). Moreover, instantiating (1.2) 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.

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

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 $[\![\mathsf{f}]\!]_A : A \to A$.

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. $[\![\mathsf{f}]\!]_{\mathbb{Z}}(k) = k + 1$. We also have the integers modulo 2, denoted by $\mathbb{Z}_2$, where $[\![\mathsf{f}]\!]_{\mathbb{Z}_2}(k) = k + 1 (\mathrm{mod}\ 2)$.

The fact that a function $h : A \to B$ satisfies (1.2) for the symbol f is equivalent to the following commutative square.

$$\begin{tikzcd} A \arrow[r, "h"] \arrow[d, "[\![\mathsf{f}]\!]_A"'] & B \arrow[d, "[\![\mathsf{f}]\!]_B"] \\ A \arrow[r, "h"'] & B \end{tikzcd}$$

We conclude that $\mathbf{Alg}(\Sigma)$ is the category whose objects are endofunctions and whose morphisms are commutative squares as above.[60] There is a homomorphism 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.

[60] 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 $1 \in \mathbb{N}$.

3. Let $\Sigma = \{\cdot : 2\}$ be the signature containing a single binary operation symbol. A Σ-algebra is a set A equipped with an interpretation $[\![\cdot]\!]_A : A \times A \to A$. Such 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,[61] not every Σ-algebra underlies a group since $[\![\cdot]\!]_A$ is not required to be associative for example. The next definition will allow us to talk about certain classes of Σ-algebras with some properties like associativity.

[61] In fact, every group has an underlying algebra for the signature $\{\cdot : 2, e : 0, -^{-1} : 1\}$.

If we want to say that $\cdot$ is commutative, we could write

$$\forall a, b \in A, \quad [\![\cdot]\!]_A(a, b) = [\![\cdot]\!]_A(b, a).$$

To say that $\cdot$ is associative, we write

$$\forall a, b, c \in A, \quad [\![\cdot]\!]_A([\![\cdot]\!]_A(a, b), c) = [\![\cdot]\!]_A(a, [\![\cdot]\!]_A(b, c)),$$

and as you can see, it gets hard to read very quickly. We make our life easier by defining the interpretation of Σ-terms which are syntactic gadgets built by iterating the symbols in Σ.

Definition 1.5 (Term)**.** Let Σ be a signature and A be a set.[62] We denote with $\mathcal{T}_\Sigma A$ the set of **Σ-terms** built syntactically from A and the operation symbols in Σ, i.e. the set inductively defined by

$$\frac{a \in A}{a \in \mathcal{T}_\Sigma A} \quad \text{and} \quad \frac{\mathsf{op} : n \in \Sigma \qquad t_1, \dots, t_n \in \mathcal{T}_\Sigma A}{\mathsf{op}(t_1, \dots, t_n) \in \mathcal{T}_\Sigma A}. \tag{1.3}$$

[62] In the sequel, unless otherwise stated, Σ will be an arbitrary signature.

We identify elements $a \in A$ with the corresponding terms $a \in \mathcal{T}_\Sigma A$, and we also identify (as outlined in Footnote 56) elements of $\Sigma(A)$ with terms in $\mathcal{T}_\Sigma A$ containing exactly one occurrence of an operation symbol.[63]

[63] Note that any constant $\mathsf{p} : 0 \in \Sigma$ belongs to all $\mathcal{T}_\Sigma A$ by the second rule defining $\mathcal{T}_\Sigma A$.

The assignment $A \mapsto \mathcal{T}_\Sigma A$ can be turned into a functor $\mathcal{T}_\Sigma : \mathbf{Set} \to \mathbf{Set}$ by inductively defining, for any function $f : A \to B$, the function $\mathcal{T}_\Sigma f : \mathcal{T}_\Sigma A \to \mathcal{T}_\Sigma B$ as follows:[64]

$$\frac{a \in A}{\mathcal{T}_\Sigma f(a) = f(a)} \quad \text{and} \quad \frac{\mathsf{op} : n \in \Sigma \qquad t_1, \dots, t_n \in \mathcal{T}_\Sigma A}{\mathcal{T}_\Sigma f(\mathsf{op}(t_1, \dots, t_n)) = \mathsf{op}(\mathcal{T}_\Sigma f(t_1), \dots, \mathcal{T}_\Sigma f(t_n))}. \tag{1.4}$$

[64] In words, $\mathcal{T}_\Sigma f$ replaces a with $f(a)$ and does nothing to operation symbols nor the structure of the term. In particular, $\mathcal{T}_\Sigma f$ acts as identity on constants.

Proposition 1.6. *The action of $\mathcal{T}_\Sigma$ is functorial, namely, for any $A \xrightarrow{f} B \xrightarrow{g} C$, $\mathcal{T}_\Sigma \mathrm{id}_A = \mathrm{id}_{\mathcal{T}_\Sigma A}$ and $\mathcal{T}_\Sigma(g \circ f) = \mathcal{T}_\Sigma g \circ \mathcal{T}_\Sigma f$.*

Proof. We proceed by induction for both equations.[65] For any $a \in A$, we have $\mathcal{T}_\Sigma \mathrm{id}_A(a) = \mathrm{id}_A(a) = a$ and

$$\mathcal{T}_\Sigma(g \circ f)(a) = (g \circ f)(a) = \mathcal{T}_\Sigma g(\mathcal{T}_\Sigma f(a)).$$

[65] Many proofs in this chapter are by induction until some point where we will have enough results to efficiently use commutative diagrams.

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

$$\mathcal{T}_\Sigma \mathrm{id}_A(\mathsf{op}(t_1, \ldots, t_n)) \stackrel{(1.4)}{=} \mathsf{op}(\mathcal{T}_\Sigma \mathrm{id}_A(t_1), \ldots, \mathcal{T}_\Sigma \mathrm{id}_A(t_n)) \stackrel{\text{I.H.}}{=} \mathsf{op}(t_1, \ldots, t_n),$$

and

$$\begin{aligned}
\mathcal{T}_\Sigma(g \circ f)(t) &= \mathcal{T}_\Sigma(g \circ f)(\mathsf{op}(t_1, \ldots, t_n)) & \\
&= \mathsf{op}(\mathcal{T}_\Sigma(g \circ f)(t_1), \ldots, \mathcal{T}_\Sigma(g \circ f)(t_n)) & \text{by (1.4)} \\
&= \mathsf{op}(\mathcal{T}_\Sigma g(\mathcal{T}_\Sigma f(t_1)), \ldots, \mathcal{T}_\Sigma g(\mathcal{T}_\Sigma f(t_n))) & \text{I.H.} \\
&= \mathcal{T}_\Sigma g(\mathsf{op}(\mathcal{T}_\Sigma f(t_1), \ldots, \mathcal{T}_\Sigma f(t_n))) & \text{by (1.4)} \\
&= \mathcal{T}_\Sigma g \mathcal{T}_\Sigma f(\mathsf{op}(t_1, \ldots, t_n)). & \text{by (1.4)}
\end{aligned}$$ □

Examples 1.7. 1. With $\Sigma = \{\mathsf{p} : 0\}$, a Σ-term over A is either an element of A or the constant p. For a function $f : A \to B$, the function $\mathcal{T}_\Sigma f$ sends a to $f(a)$ and p to itself. The functor $\mathcal{T}_\Sigma$ is then naturally isomorphic to the maybe 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.[66] The functor $\mathcal{T}_\Sigma$ is then naturally isomorphic to the functor sending A to $\mathbb{N} \times A$.

3. With $\Sigma = \{\cdot : 2\}$, a Σ-term is either an element of A or any expression formed by *multiplying* elements of A together like $a \cdot b$, $a \cdot (b \cdot c)$, $((a \cdot a) \cdot c) \cdot (b \cdot c)$ and so on when $a, b, c \in A$.[67]

[66] For a function $f : A \to B$, the function $\mathcal{T}_\Sigma f$ replaces a with $f(a)$ and does not change the number of iterations of f.

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

As we said above, any element in A is a term in $\mathcal{T}_\Sigma A$, we will denote this embedding with $\eta^\Sigma_A : A \to \mathcal{T}_\Sigma A$, in particular, we will write $\eta^\Sigma_A(a)$ 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 $\mathcal{T}_\Sigma f$ in (1.4) becomes

$$\frac{a \in A}{\mathcal{T}_\Sigma f(\eta^\Sigma_A(a)) = \eta^\Sigma_B(f(a))}.$$

This is exactly what it means for the family of maps $\eta^\Sigma_A : A \to \mathcal{T}_\Sigma A$ to be natural in A,[68] in other words that $\eta^\Sigma : \mathrm{id}_{\mathbf{Set}} \Rightarrow \mathcal{T}_\Sigma$ is a natural transformation. We can mention now that it will be part of some additional structure on the functor $\mathcal{T}_\Sigma$ (a monad). The other part of that structure is a natural transformation $\mu^\Sigma : \mathcal{T}_\Sigma \mathcal{T}_\Sigma \Rightarrow \mathcal{T}_\Sigma$, that is more easily described using trees.

[68] As a commutative square:

$$\begin{tikzcd}
A \arrow[r, "f"] \arrow[d, "\eta^\Sigma_A"'] & B \arrow[d, "\eta^\Sigma_B"] \\
\mathcal{T}_\Sigma A \arrow[r, "\mathcal{T}_\Sigma f"'] & \mathcal{T}_\Sigma B
\end{tikzcd} \tag{1.5}$$

For an arbitrary signature Σ, we can think of $\mathcal{T}_\Sigma A$ 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 $\mathcal{T}_\Sigma f$ fairly straightforward: it applies f to the labels of all the leaves as depicted in Figure 1.1.

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 $\mu^\Sigma_A : \mathcal{T}_\Sigma \mathcal{T}_\Sigma A \to \mathcal{T}_\Sigma A$ 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

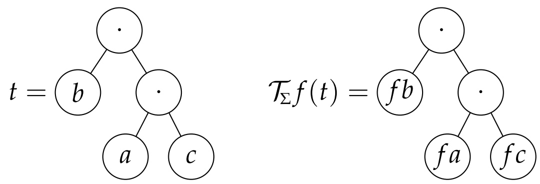


Figure 1.1: Applying $\mathcal{T}_\Sigma f$ to $b \cdot (a \cdot c)$ yields $f(b) \cdot (f(a) \cdot f(c))$.

tree T_i at the leaf labelled T_i). Figure 1.2 shows an example for $\Sigma = \{\cdot : 2\}$. More formally, μ_A^Σ is defined inductively by:

$$\mu_A^\Sigma(\eta^\Sigma_{\mathcal{T}_\Sigma A}(t)) = t \text{ and } \mu_A^\Sigma(\mathsf{op}(t_1, \ldots, t_n)) = \mathsf{op}(\mu_A^\Sigma(t_1), \ldots, \mu_A^\Sigma(t_n)). \tag{1.6}$$

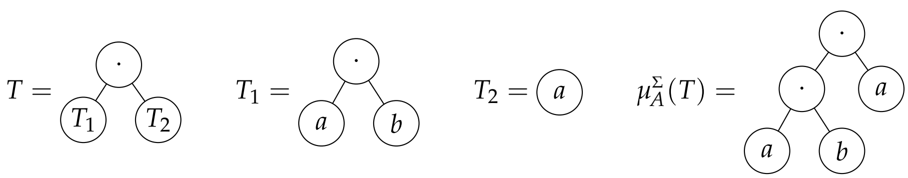


Figure 1.2: Flattening of a term.

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

Proposition 1.8. *The family of maps $\mu_A^\Sigma : \mathcal{T}_\Sigma \mathcal{T}_\Sigma A \to \mathcal{T}_\Sigma A$ is natural in A.*

Proof. We need to prove that for any function $f : A \to B$, $\mathcal{T}_\Sigma f \circ \mu_A^\Sigma = \mu_B^\Sigma \circ \mathcal{T}_\Sigma \mathcal{T}_\Sigma f$.[69] 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.

For the base case (i.e. terms in the image of $\eta^\Sigma_{\mathcal{T}_\Sigma A}$), we have

$$\begin{aligned}
\mu_B^\Sigma(\mathcal{T}_\Sigma \mathcal{T}_\Sigma f(\eta^\Sigma_{\mathcal{T}_\Sigma A}(t))) &= \mu_B^\Sigma(\eta^\Sigma_{\mathcal{T}_\Sigma B}(\mathcal{T}_\Sigma f(t))) && \text{by (1.5)}\\
&= \mathcal{T}_\Sigma f(t) && \text{by (1.6)}\\
&= \mathcal{T}_\Sigma f(\mu_A^\Sigma(\eta^\Sigma_{\mathcal{T}_\Sigma A}(t))). && \text{by (1.6)}
\end{aligned}$$

For the inductive step, we have

$$\begin{aligned}
\mu_B^\Sigma(\mathcal{T}_\Sigma \mathcal{T}_\Sigma f(\mathsf{op}(t_1, \ldots, t_n))) &= \mu_B^\Sigma(\mathsf{op}(\mathcal{T}_\Sigma \mathcal{T}_\Sigma f(t_1), \ldots, \mathcal{T}_\Sigma \mathcal{T}_\Sigma f(t_n))) && \text{by (1.4)}\\
&= \mathsf{op}(\mu_B^\Sigma(\mathcal{T}_\Sigma \mathcal{T}_\Sigma f(t_1)), \ldots, \mu_B^\Sigma(\mathcal{T}_\Sigma \mathcal{T}_\Sigma f(t_n))) && \text{by (1.6)}\\
&= \mathsf{op}(\mathcal{T}_\Sigma f(\mu_A^\Sigma(t_1)), \ldots, \mathcal{T}_\Sigma f(\mu_A^\Sigma(t_n))) && \text{I.H.}\\
&= \mathcal{T}_\Sigma f(\mathsf{op}(\mu_A^\Sigma(t_1), \ldots, \mu_A^\Sigma(t_n))) && \text{by (1.4)}\\
&= \mathcal{T}_\Sigma f(\mu_A^\Sigma(\mathsf{op}(t_1, \ldots, t_n))) && \text{by (1.6)} \quad \square
\end{aligned}$$

[69] As a commutative square:

$$\begin{tikzcd}
\mathcal{T}_\Sigma \mathcal{T}_\Sigma A \arrow[r, "\mathcal{T}_\Sigma \mathcal{T}_\Sigma f"] \arrow[d, "\mu_A^\Sigma"'] & \mathcal{T}_\Sigma \mathcal{T}_\Sigma B \arrow[d, "\mu_B^\Sigma"] \\
\mathcal{T}_\Sigma A \arrow[r, "\mathcal{T}_\Sigma f"'] & \mathcal{T}_\Sigma B
\end{tikzcd} \tag{1.7}$$

By definition, we have that $\mu^\Sigma \cdot \eta^\Sigma \mathcal{T}_\Sigma$ is the identity transformation $\mathbb{1}_{\mathcal{T}_\Sigma} : \mathcal{T}_\Sigma \Rightarrow \mathcal{T}_\Sigma$.[70] 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 if 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:

[70] 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 $\mu^\Sigma \cdot \eta^\Sigma \mathcal{T}_\Sigma$ at A is $\mu^\Sigma_A \circ \eta^\Sigma_{\mathcal{T}_\Sigma A}$ which is $\mathrm{id}_{\mathcal{T}_\Sigma A}$ by (1.6).

Lemma 1.9. *For any set A, $\mu^\Sigma_A \circ \mathcal{T}_\Sigma \eta^\Sigma_A = \mathrm{id}_{\mathcal{T}_\Sigma A}$, hence $\mu^\Sigma \cdot \mathcal{T}_\Sigma \eta^\Sigma = \mathbb{1}_{\mathcal{T}_\Sigma}$.*

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

$$\mu^\Sigma_A(\mathcal{T}_\Sigma \eta^\Sigma_A(\eta^\Sigma_A(a))) \overset{(1.5)}{=} \mu^\Sigma_A(\eta^\Sigma_{\mathcal{T}_\Sigma A}(\eta^\Sigma_A(a))) \overset{(1.6)}{=} \eta^\Sigma_A(a).$$

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

$$\begin{aligned}
\mu^\Sigma_A(\mathcal{T}_\Sigma \eta^\Sigma_A(t)) &= \mu^\Sigma_A(\mathcal{T}_\Sigma \eta^\Sigma_A(\mathsf{op}(t_1, \dots, t_n))) \\
&= \mu^\Sigma_A(\mathsf{op}(\mathcal{T}_\Sigma \eta^\Sigma_A(t_1), \dots, \mathcal{T}_\Sigma \eta^\Sigma_A(t_n))) && \text{by (1.4)} \\
&= \mathsf{op}(\mu^\Sigma_A(\mathcal{T}_\Sigma \eta^\Sigma_A(t_1)), \dots, \mu^\Sigma_A(\mathcal{T}_\Sigma \eta^\Sigma_A(t_n))) && \text{by (1.6)} \\
&= \mathsf{op}(t_1, \dots, t_n) = t && \text{I.H.}
\end{aligned}$$
□

Trees also make the depth of a term a visual concept. A term $t \in \mathcal{T}_\Sigma A$ is said to be of **depth** $d \in \mathbb{N}$ if the tree representing it has depth d.[71] We give an inductive definition:

[71] i.e. the longest path from the root to a leaf has d edges. In Figure 1.2, 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 56).

In any Σ-algebra $\mathbb{A}$, the interpretations of operation symbols give us an element of A for each element of $\Sigma(A)$. Therefore, we get a value in A for all terms in $\mathcal{T}_\Sigma A$ of depth 0 or 1 (the value associated to $\eta^\Sigma_A(a)$ is a). Using the inductive definition of $\mathcal{T}_\Sigma A$, we can extend these interpretations to all terms: abusing notation, we define the function $[\![-]\!]_A : \mathcal{T}_\Sigma A \to A$ by[72]

$$\frac{a \in A}{[\![a]\!]_A = a} \quad \text{and} \quad \frac{\mathsf{op} : n \in \Sigma \qquad t_1, \dots, t_n \in \mathcal{T}_\Sigma A}{[\![\mathsf{op}(t_1, \dots, t_n)]\!]_A = [\![\mathsf{op}]\!]_A([\![t_1]\!]_A, \dots, [\![t_n]\!]_A)}. \tag{1.8}$$

[72] For categorical thinkers, $\mathcal{T}_\Sigma A$ 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.

$$\begin{tikzcd}
\Sigma(\mathcal{T}_\Sigma A) + A \arrow[r, dashed] \arrow[d] & \Sigma(A) + A \arrow[d, "{[[\![-]\!]_A, \mathrm{id}_A]}"] \\
\mathcal{T}_\Sigma A \arrow[r, dashed] & A
\end{tikzcd}$$

The vertical arrow on the left is basically (1.3).

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

$$[\![-]\!]^\iota_A = \mathcal{T}_\Sigma X \xrightarrow{\mathcal{T}_\Sigma \iota} \mathcal{T}_\Sigma A \xrightarrow{[\![-]\!]_A} A. \tag{1.9}$$

Example 1.10. 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 $\mathbb{Z}_2$,[73] we obtain

[73] Recall their Σ-algebra structure given in Example 1.4.

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

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.[74]

$$\begin{tikzcd} \mathcal{T}_\Sigma A \arrow[r, "\mathcal{T}_\Sigma h"] \arrow[d, "[\![-]\!]_A"'] & \mathcal{T}_\Sigma B \arrow[d, "[\![-]\!]_B"] \\ A \arrow[r, "h"'] & B \end{tikzcd} \tag{1.10}$$

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

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

The converse is (almost trivially) true, if (1.10) commutes, then we can quickly see (1.1) commutes by embedding $\Sigma(A)$ into $\mathcal{T}_\Sigma A$ and $\Sigma(B)$ into $\mathcal{T}_\Sigma B$. It follows readily that for all homomorphisms $h : \mathbb{A} \to \mathbb{B}$ and all assignments $\iota : X \to A$,

$$h \circ [\![-]\!]_A^\iota = [\![-]\!]_B^{h\circ\iota}. \tag{1.11}$$

Coming back to associativity, instead of writing $[\![\cdot]\!]_A(a, [\![\cdot]\!]_A(b, c))$, we can now write $[\![a \cdot (b \cdot c)]\!]_A$, and it looks cleaner. Moreover, instead of considering a different term for each choice of $a, b, c \in A$, we can consider the term $x \cdot (y \cdot z)$ over a set of variables $\{x, y, z\}$ and quantify over all the possible assignments $\{x, y, z\} \to A$. We obtain the following definition.

Definition 1.11 (Equation)**.** An **equation** over a signature Σ is a triple comprising a set X of variables called the **context**, and a pair of terms $s, t \in \mathcal{T}_\Sigma X$. 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 $\iota : X \to A$, $[\![s]\!]_A^\iota = [\![t]\!]_A^\iota$. 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 $[\![s]\!]_A^\iota = [\![t]\!]_A^\iota$ holds for a particular assignment $\iota : X \to A$ and not necessarily for all assignments.

Remark 1.12. Our notation for equations is not standard because many authors do not bother writing the context of an equation and suppose it contains exactly the variables used in s and t. That is theoretically sound for universal algebra, but it will not remain so when we generalize to universal quantitative algebras. Thus, we make the context explicit in our equations as is done in [Wec12] or [Bau19] with the notations $\forall X.s = t$ and $X \mid s = t$ respectively.[75] We use the turnstile $\vdash$ to match the convention in the literature on quantitative algebras (e.g. [MPP16] and [FMS21]).

[75] Only finite contexts are used in [Wec12] and [Bau19]. We say a bit more on this in Remark 1.49

Example 1.13 (Associativity)**.** With the signature $\Sigma = \{\cdot : 2\}$ and the context $X = \{x, y, z\}$, the equation $\phi = X \vdash x \cdot (y \cdot z) = (x \cdot y) \cdot z$[76] asserts that the interpretation of $\cdot$ is associative. To prove that, suppose $\mathbb{A} \vDash \phi$, we need to show that for any $a, b, c \in A$,

$$[\![\cdot]\!]_A(a, [\![\cdot]\!]_A(b, c)) = [\![\cdot]\!]_A([\![\cdot]\!]_A(a, b), c). \tag{1.12}$$

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

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

Let $s = x \cdot (y \cdot z)$ and $t = (x \cdot y) \cdot z$. 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 $[\![\cdot]\!]_A(a, [\![\cdot]\!]_A(b, c)) = [\![s]\!]_A^\iota$. Under the same assignment, the interpretation of t is the R.H.S. Since $\mathbb{A} \vDash^\iota X \vdash s = t$, $[\![s]\!]_A^\iota = [\![t]\!]_A^\iota$, and we conclude (1.12) holds.

Examples 1.14. Here are some other simple examples of equations.

- $x, y \vdash x \cdot y = y \cdot x$ states that the interpretation binary operation $\cdot$ is commutative.
- $x, y, z, w \vdash x \cdot y = y \cdot x$ also states that (the interpretation of)$\cdot$ is commutative, but it has some extra unused variables in the context.[77]
- $x \vdash x \cdot x = x$ states that the binary operation $\cdot$ 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).
- $x, y \vdash x = y$ states that all elements in the algebra are equal (this means the algebra is either empty or a singleton).

[77] This is allowed, but it is always possible to remove unused variables in the context (see Remark 1.49).

Using the fact that interpretations are preserved by homomorphisms (1.11), we can describe how satisfaction is also preserved. Very naively, one would want to say that if $h : \mathbb{A} \to \mathbb{B}$ is a homomorphism and $\mathbb{A} \vDash \phi$, then $\mathbb{B} \vDash \phi$. That is not true.[78] It is morally because there can be many more assignments into $\mathbb{B}$ than there are into $\mathbb{A}$. Nevertheless, the naive statement is true on a per-assignment basis.

[78] For any Σ which does not contain constants, there is an initial Σ-algebra $\mathbb{I}$ whose carrier is the empty set $\emptyset$ (the interpretation of operations is completely determined because there $\Sigma(\emptyset) = \emptyset$ and there is only one function $\emptyset^n \to \emptyset$). The unique function $\emptyset \to B$ is always a homomorphism $\mathbb{I} \to \mathbb{B}$ because (1.1) trivially commutes since $\Sigma(\emptyset) = \emptyset$. While $\mathbb{I}$ satisfies all equations (vacuously), it is clearly possible that $\mathbb{B}$ does not.

Lemma 1.15. *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

$$\begin{aligned}
\mathbb{A} \vDash^{\iota} \phi &\iff [\![s]\!]_A^{\iota} = [\![t]\!]_A^{\iota} && \text{definition of } \vDash \\
&\implies h([\![s]\!]_A^{\iota}) = h([\![t]\!]_A^{\iota}) && \\
&\implies [\![s]\!]_B^{h\circ\iota} = [\![t]\!]_B^{h\circ\iota} && \text{by (1.11)} \\
&\iff \mathbb{B} \vDash^{h\circ\iota} \phi. && \text{definition of } \vDash
\end{aligned}$$
□

Another neat fact is that flattening interacts well with interpreting in the following sense.

Lemma 1.16. *For any Σ-algebra $\mathbb{A}$, the following square commutes.*[79]

$$\begin{tikzcd}
\mathcal{T}_\Sigma\mathcal{T}_\Sigma A \arrow[r, "\mathcal{T}_\Sigma[\![-]\!]_A"] \arrow[d, "\mu^\Sigma_A"'] & \mathcal{T}_\Sigma A \arrow[d, "[\![-]\!]_A"] \\
\mathcal{T}_\Sigma A \arrow[r, "[\![-]\!]_A"'] & A
\end{tikzcd} \tag{1.13}$$

[79] In words, given a term in $\mathcal{T}_\Sigma\mathcal{T}_\Sigma A$, 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.

This also generalizes to terms in $\mathcal{T}_\Sigma\mathcal{T}_\Sigma X$. Indeed, given an assignment, $\iota : X \to A$, we can either flatten a term and interpret it under ι, or we can interpret all the inner terms under ι, then interpret the result, as shown in (1.14).

$$\begin{tikzcd}
\mathcal{T}_\Sigma\mathcal{T}_\Sigma X \arrow[r, "\mathcal{T}_\Sigma\mathcal{T}_\Sigma\iota"] \arrow[d, "\mu^\Sigma_X"'] \arrow[rr, bend left, "\mathcal{T}_\Sigma[\![-]\!]^{\iota}_A"] \arrow[dr, phantom, "(1.7)"] & \mathcal{T}_\Sigma\mathcal{T}_\Sigma A \arrow[r, "\mathcal{T}_\Sigma[\![-]\!]_A"] \arrow[d, "\mu^\Sigma_A"'] \arrow[dr, phantom, "(1.13)"] & \mathcal{T}_\Sigma A \arrow[d, "[\![-]\!]_A"] \\
\mathcal{T}_\Sigma X \arrow[r, "\mathcal{T}_\Sigma\iota"'] \arrow[rr, bend right, "[\![-]\!]^{\iota}_A"'] & \mathcal{T}_\Sigma A \arrow[r, "[\![-]\!]_A"'] & A
\end{tikzcd} \tag{1.14}$$

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

$$[\![\mu^\Sigma_A(\eta^\Sigma_A(t))]\!]_A \stackrel{(1.6)}{=} [\![t]\!]_A \stackrel{(1.8)}{=} [\![\eta^\Sigma_A([\![t]\!]_A)]\!]_A \stackrel{(1.5)}{=} [\![\mathcal{T}_\Sigma[\![-]\!]_A(\eta^\Sigma_A(t))]\!].$$

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

$$[\![\mu^\Sigma_A(t)]\!]_A = [\![\mathsf{op}(\mu^\Sigma_A(t_1), \ldots, \mu^\Sigma_A(t_n))]\!]_A \qquad \text{by (1.6)}$$

$$\begin{aligned}
&= [\![\mathsf{op}]\!]_A\left([\![\mu^\Sigma_A(t_1)]\!]_A, \dots, [\![\mu^\Sigma_A(t_n)]\!]_A\right) && \text{by (1.8)}\\
&= [\![\mathsf{op}]\!]_A\left([\![\mathcal{T}_\Sigma[\![-]\!]_A(t_1)]\!]_A, \dots, [\![\mathcal{T}_\Sigma[\![-]\!]_A(t_n)]\!]_A\right) && \text{I.H.}\\
&= [\![\mathsf{op}(\mathcal{T}_\Sigma[\![-]\!]_A(t_1), \dots, \mathcal{T}_\Sigma[\![-]\!]_A(t_n))]\!]_A && \text{by (1.8)}\\
&= [\![\mathcal{T}_\Sigma[\![-]\!]_A(\mathsf{op}(t_1, \dots, t_n))]\!]_A && \text{by (1.4)}\\
&= [\![\mathcal{T}_\Sigma[\![-]\!]_A(t)]\!]_A. && \square
\end{aligned}$$

Remark 1.17. To see Lemma 1.16 in another way, notice that (1.13) looks a lot like (1.10), but the map on the left is not the interpretation on an algebra. Except it is! Indeed, we can give a trivial (or syntactic) interpretation of $\mathsf{op} : n \in \Sigma$ on the set $\mathcal{T}_\Sigma A$ by letting $[\![\mathsf{op}]\!]_{\mathcal{T}_\Sigma A}(t_1, \dots, t_n) = \mathsf{op}(t_1, \dots, t_n)$. Then, we can verify by induction[80] that $[\![-]\!]_{\mathcal{T}_\Sigma A} : \mathcal{T}_\Sigma\mathcal{T}_\Sigma A \to \mathcal{T}_\Sigma A$ is equal to μ^Σ_A. We conclude that Lemma 1.16 says that for any algebra, $[\![-]\!]_A$ is a homomorphism from $(\mathcal{T}_\Sigma A, [\![-]\!]_{\mathcal{T}_\Sigma A})$ to $\mathbb{A}$.

[80] Or we can compare (1.6) and (1.8) 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 $\mathcal{T}_\Sigma\mathcal{T}_\Sigma A$ and $\mathcal{T}_\Sigma A$, and given a function $f : A \to B$, $\mathcal{T}_\Sigma f$ is a homomorphism between $\mathcal{T}_\Sigma A$ and $\mathcal{T}_\Sigma B$.

Lemma 1.18. *For any function $f : A \to B$, the following squares commute.*[81]

$$\begin{tikzcd}
\mathcal{T}_\Sigma\mathcal{T}_\Sigma\mathcal{T}_\Sigma A \arrow[r, "\mathcal{T}_\Sigma\mu^\Sigma_A"] \arrow[d, "\mu^\Sigma_{\mathcal{T}_\Sigma A}"'] & \mathcal{T}_\Sigma\mathcal{T}_\Sigma A \arrow[d, "\mu^\Sigma_A"] \\
\mathcal{T}_\Sigma\mathcal{T}_\Sigma A \arrow[r, "\mu^\Sigma_A"'] & \mathcal{T}_\Sigma
\end{tikzcd} \tag{1.15}$$

$$\begin{tikzcd}
\mathcal{T}_\Sigma\mathcal{T}_\Sigma A \arrow[r, "\mathcal{T}_\Sigma\mathcal{T}_\Sigma B"] \arrow[d, "\mu^\Sigma_A"'] & \mathcal{T}_\Sigma\mathcal{T}_\Sigma B \arrow[d, "\mu^\Sigma_B"] \\
\mathcal{T}_\Sigma \arrow[r, "\mathcal{T}_\Sigma f"'] & \mathcal{T}_\Sigma B
\end{tikzcd} \tag{1.16}$$

[81] *Proof.* We have already shown both these squares commute. Indeed, (1.15) is an instance of (1.13) where we identify μ^Σ_A with the interpretation $[\![-]\!]_{\mathcal{T}_\Sigma A}$ as explained in Remark 1.17, and (1.16) is the naturality square (1.7).

Another consequence of (1.15) is that if you have a term in $\mathcal{T}_\Sigma^n A$ for any $n \in \mathbb{N}$, there are $(n-1)!$ ways to flatten it[82] by successively applying an instance of $\mathcal{T}_\Sigma^i \mu^\Sigma_{\mathcal{T}_\Sigma^j A}$ 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 $\mathcal{T}_\Sigma A$. 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 a consequence of associativity, hence one also says μ^Σ is associative.

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

While the categories $\mathbf{Alg}(\Sigma)$ for different signatures can be interesting to study on their own, the examples we wanted to generalize like **Grp** or **Ring** are not of that kind, they are special subcategories of some $\mathbf{Alg}(\Sigma)$ that are called varieties.

Definition 1.19 (Variety)**.** Given a class E of equations, we say $\mathbb{A}$ satisfies E and write $\mathbb{A} \vDash E$ if $\mathbb{A} \vDash \phi$ for all $\phi \in E$.[83] 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. A **variety** is a category equal to $\mathbf{Alg}(\Sigma, E)$ for some class of equations E.

[83] Similarly for satisfaction under a particular assignment ι:

$$\mathbb{A} \vDash^\iota E \iff \forall \phi \in E, \mathbb{A} \vDash^\iota \phi.$$

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 $U : \mathbf{Alg}(\Sigma) \to \mathbf{Set}$.[84]

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

It is never the case in practice that E is a proper class, it is usually a finite or countable set, even recursively enumerable. Still, nothing breaks when E is a class, and we will need this generality in one our main contributions (Theorem 3.84).

Examples 1.20. 1. With $\Sigma = \{\mathsf{p}:0\}$, there are morally only four different equations:[85]

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

where we write nothing before the turnstile ($\vdash$) instead of the empty set $\emptyset$.

Any algebra $\mathbb{A}$ satisfies the first two equations because $[\![\mathsf{p}]\!]_A^\iota = [\![\mathsf{p}]\!]_A^\iota$, where $\iota : \emptyset \to A$ is the only possible assignment, and $[\![x]\!]_A^\iota = \iota(x) = [\![x]\!]_A^\iota$ for all $\iota : \{x\} \to A$. If $\mathbb{A}$ satisfies the third, it means that A is empty or a singleton because for any $a, b \in A$, the assignments $\iota_a = x \mapsto a$ and $\iota_b = x \mapsto b$ give us[86]

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

If $\mathbb{A}$ satisfies the fourth equation, it is also empty or 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,[87] there are only two varieties in that signature, either $\mathbf{Alg}(\Sigma, E)$ is all of $\mathbf{Alg}(\Sigma)$, or it contains only the empty set and the singletons.

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

$$x, y, z \vdash x + (y + z) = (x + y) + z, \quad x, y \vdash x + y = y + x, \text{ and } \quad x \vdash x + \mathsf{e} = x. \tag{1.17}$$

We already saw in Example 1.13 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 $[\![+]\!]$.[88] Moreover, note that a homomorphism of Σ-algebras from $\mathbb{A}$ to $\mathbb{B}$ is any function $h : A \to B$ that satisfies

$$\forall a, a' \in A, \quad h([\![+]\!]_A(a, a')) = [\![+]\!]_B(h(a), h(a')) \text{ and } h([\![\mathsf{e}]\!]_A) = [\![\mathsf{e}]\!]_B.$$

Namely, a homomorphism preserves the addition and its neutral element. Thus, letting E be the set containing the equations in (1.17), 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 (1.17),[89] and we can show that $\mathbf{Alg}(\Sigma, E)$ is the category **Ab** of abelian groups and group homomorphisms.

4. We could very similarly develop signatures and equations to get **Grp** and **Ring** as varieties. Although we should note that it is possible for (Σ, E) and (Σ', E') to define the same variety (or isomorphic varieties).

Among different classes of equations over the same signature that define the same variety, there is a largest one.

[85] Let us not formally argue about that here, but your intuition on equality and the fact that terms in $\mathcal{T}_\Sigma X$ are either $x \in X$ or p should be enough to convince you.

[86] 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.

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

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

$$[\![a + \mathsf{e}]\!]_A = a.$$

By commutativity, we also get $[\![\mathsf{e} + a]\!]_A = a$.

[89] While the signature has changed between the two examples, the equations of (1.17) can be understood over both signatures because they concern terms constructed using the symbols common to both signatures.

Definition 1.21 (Algebraic theory)**.** Given a class 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:[90]

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

[90] Note that, even if E is a set, there is no guarantee that $\mathfrak{Th}(E)$ is a set (in fact it never is) because the collection of all equations is a proper class (because the contexts can be any set).

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

$$\forall \mathbb{A} \in \mathbf{Alg}(\Sigma), \quad \mathbb{A} \vDash E \implies \mathbb{A} \vDash \phi. \tag{1.18}$$

Of course, $\mathfrak{Th}(E)$ contains all of E,[91] but also many more equations like $x \vdash x = x$ which is satisfied by any algebra. We will see in §1.3 how to find which equations are entailed by others.

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

It is easy to see that $\mathbf{Alg}(\Sigma, E) = \mathbf{Alg}(\Sigma, E')$ implies $\mathfrak{Th}(E) = \mathfrak{Th}(E')$, $E \subseteq \mathfrak{Th}(E)$, and $\mathbf{Alg}(\Sigma, \mathfrak{Th}(E)) = \mathbf{Alg}(\Sigma, E)$. It follows that $\mathfrak{Th}(E)$ is the maximal class of equations defining the variety $\mathbf{Alg}(\Sigma, E)$.

Example 1.22. If E contains the equations in (1.17), then $\mathfrak{Th}(E)$ will contain all the equations that every commutative monoid satisfies. Here is a non-exhaustive list:

- $x \vdash \mathtt{e} + x = x$ says that $[\![\mathtt{e}]\!]$ is a neutral element on the left for $[\![+]\!]$ which is true because, by equations in (1.17), $[\![\mathtt{e}]\!]$ is neutral on the right and $[\![+]\!]$ is commutative.
- $z, w \vdash z + w = w + z$ also states commutativity of $[\![+]\!]$ but with different variable names.
- $x, y, z, w \vdash (x + w) + (x + z) + (x + y) = ((x + x) + x) + (y + (z + (\mathtt{e} + w)))$ is just a random equation that can be shown using the properties of commutative monoids.[92]

[92] We will see in §1.3 how to systematically generate all the equations in $\mathfrak{Th}(E)$.

1.2 Free Algebras

Very briefly, the free (Σ, E)-algebra on X is the least constrained Σ-algebra which "contains" X and satisfies E. It necessarily satisfies all the equations in $\mathfrak{Th}(E)$ as well, but it does not satisfy any other equation $X \vdash s = t$ that is not also satisfied by all (Σ, E)-algebras. We will prove it always exist, and we start with an example.

Example 1.23 (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.[93] There is a natural $\Sigma_{\mathbf{Mon}}$-algebra structure on X^* where $\cdot$ 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 (1.19).[94]

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

[94] It does not satisfy $x, y \vdash x \cdot y = y \cdot x$ asserting commutativity because `ab` and `ba` are two different words.

$$E_{\mathbf{Mon}} = \{x, y, z \vdash x \cdot (y \cdot z) = (x \cdot y) \cdot z, \quad x \vdash x \cdot \mathtt{e} = x, \quad x \vdash \mathtt{e} \cdot x = x\}. \tag{1.19}$$

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 $x \in X \subseteq X^*$.[95] This can be summarized in the following diagram, where X^* denotes both the set of words and the monoid.

[95] f^* sends $x_1 \cdots x_n$ to $[\![f(x_1) \cdot (f(x_2) \cdots f(x_n))]\!]_A$.

$$\begin{tikzcd} \text{in } \mathbf{Set} & & & \text{in } \mathbf{Alg}(\Sigma_{\mathbf{Mon}}, E_{\mathbf{Mon}}) \\ X \arrow[r, hook] \arrow[dr, "f"'] & X^* \arrow[d, dashed, "f^*"] & & X^* \arrow[d, dashed, "f^*"] \\ & A & \arrow[l, "U"'] & \mathbb{A} \end{tikzcd} \tag{1.20}$$

A consequence of (1.20) which makes the idea of freeness more concrete is that X^* satisfies an equation $X \vdash s = t$ if and only if all $(\Sigma_{\mathbf{Mon}}, E_{\mathbf{Mon}})$-algebras satisfy it.[96] In other words, X^* only satisfies the equations it *needs* to satisfy.

[96] The forward direction uses Lemma 1.15 with ι being the inclusion $X \hookrightarrow X^*$ and h being f^*. The converse direction is trivial since we know X^* belongs to $\mathbf{Alg}(\Sigma_{\mathbf{Mon}}, E_{\mathbf{Mon}})$.

The free $(\Sigma_{\mathbf{Mon}}, E_{\mathbf{Mon}})$-algebra over any set is always[97] the set of finite words over that set with $\cdot$ and $\mathtt{e}$ interpreted as concatenation and the empty word respectively.

[97] 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 $\mathcal{T}_{\Sigma_{\mathbf{Mon}}} X$ is the set of all terms constructed with the symbols in $\Sigma_{\mathbf{Mon}}$ and the elements of X.[98] 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 occurrence 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.

[98] 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^*.[99] 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 $E_{\mathbf{Mon}}$.

[99] For instance, both $\mathtt{r} \cdot (\mathtt{s} \cdot \mathtt{m})$ and $(\mathtt{r} \cdot \mathtt{s}) \cdot \mathtt{m}$ become the word $\mathtt{rsm}$ 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 class E of equations over Σ. For any set X, we can define a binary relation $\equiv_E$ on Σ-terms[100] that contains the pair (s, t) whenever the interpretation of s and t coincide in any (Σ, E)-algebra. Formally, we have for any $s, t \in \mathcal{T}_\Sigma X$,

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

$$s \equiv_E t \iff X \vdash s = t \in \mathfrak{Th}(E). \tag{1.21}$$

We now show $\equiv_E$ is a congruence relation on $\mathcal{T}_\Sigma X$.[101]

[101] A **congruence** on a Σ-algebra $\mathbb{A}$ is an equivalence relation $\sim \subseteq A \times A$ on the carrier satisfying for all $\mathsf{op} : n \in \Sigma$ and $a_1, \ldots, a_n, b_1, \ldots, b_n \in A$:

$$(\forall i, a_i \sim b_i) \implies [\![\mathsf{op}]\!]_A(a_1, \ldots, a_n) \sim [\![\mathsf{op}]\!]_A(b_1, \ldots, b_n).$$

Lemma 1.24. *For any set X, the relation $\equiv_E$ is reflexive, symmetric, transitive, and satisfies for any $\mathsf{op} : n \in \Sigma$ and $s_1, \ldots, s_n, t_1, \ldots, t_n \in \mathcal{T}_\Sigma X$,*

$$(\forall 1 \leq i \leq n, s_i \equiv_E t_i) \implies \mathsf{op}(s_1, \ldots, s_n) \equiv_E \mathsf{op}(t_1, \ldots, t_n). \tag{1.22}$$

Proof. Briefly, reflexivity, symmetry, and transitivity all follow from the fact that equality satisfies these properties, and (1.22) follows from the fact that operation symbols are interpreted as *deterministic* functions (a unique output for each input), so they preserve equality. We detail this below.

(*Reflexivity*) For any $t \in \mathcal{T}_\Sigma X$, 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 $s, t \in \mathcal{T}_\Sigma X$ and $\mathbb{A} \in \mathbf{Alg}(\Sigma)$, if $\mathbb{A} \vDash X \vdash s = t$, then $\mathbb{A} \vDash X \vdash t = s$. Indeed, if $[\![s]\!]^\iota_A = [\![t]\!]^\iota_A$ 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 $s, t, u \in \mathcal{T}_\Sigma X$, if all (Σ, E)-algebras satisfy $X \vdash s = t$ and $X \vdash t = u$, then they also satisfy $X \vdash s = u$.[102] Transitivity follows.

[102] 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 $[\![s]\!]^\iota_A = [\![u]\!]^\iota_A$.

(1.22) For any $\mathsf{op} : n \in \Sigma$, $s_1, \dots, s_n, t_1, \dots, t_n \in \mathcal{T}_\Sigma X$, 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,

$$\begin{aligned}
[\![\mathsf{op}(s_1, \dots, s_n)]\!]^\iota_A &= [\![\mathsf{op}]\!]_A([\![s_1]\!]^\iota_A, \dots, [\![s_n]\!]^\iota_A) && \text{by (1.8)} \\
&= [\![\mathsf{op}]\!]_A([\![t_1]\!]^\iota_A, \dots, [\![t_n]\!]^\iota_A) && \forall i, [\![s_i]\!]^\iota_A = [\![t_i]\!]^\iota_A \\
&= [\![\mathsf{op}(s_1, \dots, s_n)]\!]^\iota_A && \text{by (1.8),}
\end{aligned}$$

which means $\mathbb{A} \vDash X \vdash \mathsf{op}(s_1, \dots, s_n) = \mathsf{op}(t_1, \dots, t_n)$. This was true for all Σ-algebras, so we can use the same arguments as above to conclude (1.22). $\square$

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

This yields a functor $\mathcal{T}_{\Sigma,E} : \mathbf{Set} \to \mathbf{Set}$ which sends a function $f : X \to Y$ to the unique function $\mathcal{T}_{\Sigma,E} f$ making (1.23) commute, i.e. satisfying $\mathcal{T}_{\Sigma,E} f([t]_E) = [\mathcal{T}_\Sigma f(t)]_E$. By definition, $[-]_E$ is also a natural transformation from $\mathcal{T}_\Sigma$ to $\mathcal{T}_{\Sigma,E}$.

$$\begin{tikzcd}
\mathcal{T}_\Sigma X \arrow[r, "{[-]_E}"] \arrow[d, "\mathcal{T}_\Sigma f"'] & \mathcal{T}_{\Sigma,E} X \arrow[d, dashed, "\mathcal{T}_{\Sigma,E} f"] \\
\mathcal{T}_\Sigma Y \arrow[r, "{[-]_E}"'] & \mathcal{T}_{\Sigma,E} Y
\end{tikzcd} \tag{1.23}$$

Definition 1.25 (Term algebra, semantically)**.** The **term algebra** for (Σ, E) on X is the Σ-algebra whose carrier is $\mathcal{T}_{\Sigma,E} X$ and whose interpretation of $\mathsf{op} : n \in \Sigma$ is[103]

[103] This is well-defined (i.e. invariant under change of representative) by (1.22).

$$[\![\mathsf{op}]\!]_{\mathbb{T}X}([t_1]_E, \dots, [t_n]_E) = [\mathsf{op}(t_1, \dots, t_n)]_E. \tag{1.24}$$

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 $[-]_E : \mathcal{T}_\Sigma X \to \mathcal{T}_{\Sigma,E} X$ a homomorphism,[104] namely, (1.25) commutes.

[104] Indeed, (1.24) looks exactly like (1.2) with $h = [-]_E$, $\mathbb{A} = \mathcal{T}_\Sigma X$ and $\mathbb{B} = \mathbb{T}X$.

$$\begin{tikzcd}
\mathcal{T}_\Sigma \mathcal{T}_\Sigma X \arrow[r, "\mathcal{T}_\Sigma {[-]_E}"] \arrow[d, "\mu^\Sigma_X"'] & \mathcal{T}_\Sigma \mathcal{T}_{\Sigma,E} X \arrow[d, "{[\![-]\!]_{\mathbb{T}X}}"] \\
\mathcal{T}_\Sigma X \arrow[r, "{[-]_E}"'] & \mathcal{T}_{\Sigma,E} X
\end{tikzcd} \tag{1.25}$$

Remark 1.26. We can understand Definition 1.25 a bit more abstractly. If $\mathbb{A}$ is a Σ-algebra and $\sim \subseteq A \times A$ is a congruence, then the quotient $A/\sim$ inherits a Σ-algebra structure defined as in (1.24) ($[a]$ denotes the equivalence class of a in $A/\sim$):

$$[\![\mathsf{op}]\!]_{A/\sim}([a_1], \ldots, [a_n]) = [[\![\mathsf{op}]\!]_A(a_1, \ldots, a_n)].$$

Then, $\mathbb{T}_{\Sigma,E}X$ is the quotient of the algebra $\mathcal{T}_\Sigma X$ defined in Remark 1.17 by the congruence $\equiv_E$. From this point of view, one can give an equivalent definition of $\equiv_E$ as the smallest congruence on $\mathcal{T}_\Sigma X$ such that the quotient satisfies E.[105]

[105] Namely, if $\mathcal{T}_\Sigma X/\sim$ satisfies E, then $\equiv_E \subseteq \sim$.

It is very easy to *compute* in the term algebra because all operations are realized syntactically, that is, only by manipulating symbols. Let us first look at the interpretation of Σ-terms in $\mathbb{T}X$, i.e. the function $[\![-]\!]_{\mathbb{T}X} : \mathcal{T}_\Sigma \mathcal{T}_{\Sigma,E} X \to \mathcal{T}_{\Sigma,E} X$. It was defined inductively to yield[106]

[106] where $t \in \mathcal{T}_\Sigma X$, $\mathsf{op} : n \in \Sigma$, and $t_1, \ldots, t_n \in \mathcal{T}_\Sigma \mathcal{T}_{\Sigma,E} X$.

$$[\![\eta^\Sigma_{\mathcal{T}_{\Sigma,E}X}([t]_E)]\!]_{\mathbb{T}X} = [t]_E \text{ and } [\![\mathsf{op}(t_1, \ldots, t_n)]\!]_{\mathbb{T}X} = [\![\mathsf{op}]\!]_{\mathbb{T}X}([\![t_1]\!]_{\mathbb{T}X}, \ldots, [\![t_n]\!]_{\mathbb{T}X}). \quad (1.26)$$

Remark 1.27. In particular, when E is empty, the set $\mathcal{T}_{\Sigma,\emptyset}X$ is $\mathcal{T}_\Sigma X$ quotiented by $\equiv_\emptyset$, and one can show that $\equiv_\emptyset$ is equal to equality ($=$), i.e. $\mathfrak{Th}(\emptyset)$ only contains equation of the form $X \vdash t = t$.[107] Therefore, $\mathcal{T}_{\Sigma,\emptyset}X = \mathcal{T}_\Sigma X$. Moreover, since $[-]_\emptyset$ is the identity map, we find that (1.24) becomes the definition of the interpretations given in Remark 1.17, so $\mathbb{T}_{\Sigma,\emptyset}X$ is the algebra on $\mathcal{T}_\Sigma X$ we had defined. Also, we find the interpretation of terms $[\![-]\!]_{\mathbb{T}_{\Sigma,\emptyset}X}$ is the flattening.[108]

[107] For any other equation $X \vdash s = t$ where s and t are not the same term, the Σ-algebra $\mathcal{T}_\Sigma X$ does not satisfy because the assignment $\eta^\Sigma_X : X \to \mathcal{T}_\Sigma X$ yields

$$[\![s]\!]^{\eta^\Sigma_X}_{\mathcal{T}_\Sigma X} = s \neq t = [\![t]\!]^{\eta^\Sigma_X}_{\mathcal{T}_\Sigma X}.$$

[108] By Remark 1.17 or by comparing (1.26) when $E = \emptyset$ and the definition of μ^Σ_X (1.6).

Example 1.28. Let $\Sigma = \Sigma_{\mathbf{Mon}}$ and $E = E_{\mathbf{Mon}}$ be the signature and equations defining monoids as explained in Example 1.23. We saw informally that $\mathcal{T}_{\Sigma,E}X$ is in correspondence with the set X^* of finite words over X, and we already have a monoid structure on X^*.[109] 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 $X^* \cong \mathcal{T}_{\Sigma,E}X$. First we use the inductive definition:

[109] 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 $u, v, w \in \mathcal{T}_{\Sigma,E}X$ 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 (1.24) a couple of 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 $\mathcal{T}_\Sigma \mathcal{T}_{\Sigma,E} X$ 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 (1.24) repeatedly until the result is just an equivalence class in $\mathcal{T}_{\Sigma,E}X$.

Working with terms in $\mathcal{T}_\Sigma\mathcal{T}_{\Sigma,E}X$ as trees whose leaves are labelled in $\mathcal{T}_{\Sigma,E}X$, $[\![-]\!]_{\mathbb{T}X}$ replaces each leaf by the tree corresponding to a representative for the equivalence class of the leaf's label, and then returns the equivalence class of the resulting tree. In this sense, $[\![-]\!]_{\mathbb{T}X}$ looks a lot like the flattening μ^Σ_X except it deals with equivalence classes of terms. This motivates the definition of $\mu^{\Sigma,E}_X$ to be the unique function making (1.27) commute.[110]

[110] This guarantees $\mu^{\Sigma,E}_X$ satisfies the following equations that looks like the inductive definition of μ^Σ_X in (1.6): for any $t \in \mathcal{T}_\Sigma X$, $\mu^{\Sigma,E}_X([[t]_E]_E) = [t]_E$ and for any $\mathsf{op} : n \in \Sigma$ and $t_1, \dots, t_n \in \mathcal{T}_\Sigma X$,

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

Thanks to Remark 1.27, we can immediately see that $\mu^{\Sigma,\emptyset}_X = \mu^\Sigma_X$ because $[-]_\emptyset$ is the identity and $[\![-]\!]_{\mathbb{T}_{\Sigma,\emptyset}X} = \mu^\Sigma_X$.

$$\begin{tikzcd}
\mathcal{T}_\Sigma\mathcal{T}_{\Sigma,E}X \arrow[rr, "{[\![-]\!]_{\mathbb{T}X}}"] \arrow[dr, "{[-]_E}"'] & & \mathcal{T}_{\Sigma,E}X \\
& \mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}X \arrow[ur, dashed, "\mu^{\Sigma,E}_X"'] &
\end{tikzcd} \tag{1.27}$$

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

Proposition 1.29. *The family of maps* $\mu^{\Sigma,E}_X : \mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}X \to \mathcal{T}_{\Sigma,E}X$ *is natural in* X.

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

$$\begin{tikzcd}
\mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}X \arrow[r, "\mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}f"] \arrow[d, "\mu^{\Sigma,E}_X"'] & \mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}Y \arrow[d, "\mu^{\Sigma,E}_Y"] \\
\mathcal{T}_{\Sigma,E}X \arrow[r, "\mathcal{T}_{\Sigma,E}f"'] & \mathcal{T}_{\Sigma,E}Y
\end{tikzcd} \tag{1.28}$$

We can pave the following diagram.[111]

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

$$\begin{tikzcd}
\mathcal{T}_\Sigma\mathcal{T}_{\Sigma,E}X \arrow[rr, "{[-]_E}"] \arrow[dd, "{[-]_E}"'] \arrow[drrr, dashed, "\mathcal{T}_\Sigma\mathcal{T}_{\Sigma,E}f"'] \arrow[ddrr, "{[\![-]\!]_{\mathbb{T}X}}"] & & \mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}X \arrow[rr, "\mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}f"] \arrow[d, phantom, "(a)"] & & \mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}Y \arrow[dd, "\mu^{\Sigma,E}_Y"] \\
{} \arrow[r, phantom, "(b)"] & {} \arrow[r, phantom, "(c)"] & {} & \mathcal{T}_\Sigma\mathcal{T}_{\Sigma,E}Y \arrow[ur, "{[-]_E}"'] \arrow[dr, "{[\![-]\!]_{\mathbb{T}Y}}"'] \arrow[r, phantom, "(d)"] & {} \\
\mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}X \arrow[rr, "\mu^{\Sigma,E}_X"'] & & \mathcal{T}_{\Sigma,E}X \arrow[rr, "\mathcal{T}_{\Sigma,E}f"'] & & \mathcal{T}_{\Sigma,E}Y
\end{tikzcd}$$

All of (a), (b) and (d) commute by definition. In more details, (a) is an instance of (1.23) with X replaced by $\mathcal{T}_{\Sigma,E}X$, Y by $\mathcal{T}_{\Sigma,E}Y$ and f by $\mathcal{T}_{\Sigma,E}f$, and both (b) and (d) are instances of (1.27). To show (c) commutes, we draw another diagram that looks like a cube with (c) as the front face.

$$
\begin{tikzcd}
\mathcal{T}_\Sigma\mathcal{T}_\Sigma X \arrow[rr, "\mathcal{T}_\Sigma\mathcal{T}_\Sigma f"] \arrow[dr, "\mathcal{T}_\Sigma{[-]_E}"'] \arrow[dd, "\mu^\Sigma_X"'] & & \mathcal{T}_\Sigma\mathcal{T}_\Sigma Y \arrow[dr, "\mathcal{T}_\Sigma{[-]_E}"] \arrow[dd, "\mu^\Sigma_Y"' near start] & \\
& \mathcal{T}_\Sigma\mathcal{T}_{\Sigma,E} X \arrow[rr, "\mathcal{T}_\Sigma\mathcal{T}_{\Sigma,E} f" near start] \arrow[dd, "{[\![-]\!]_{\mathbb{T}X}}"' near start] & & \mathcal{T}_\Sigma\mathcal{T}_{\Sigma,E} Y \arrow[dd, "{[\![-]\!]_{\mathbb{T}Y}}"] \\
\mathcal{T}_\Sigma X \arrow[rr, "\mathcal{T}_\Sigma f" near end] \arrow[dr, "{[-]_E}"'] & & \mathcal{T}_\Sigma Y \arrow[dr, "{[-]_E}"] & \\
& \mathcal{T}_{\Sigma,E} X \arrow[rr, "\mathcal{T}_{\Sigma,E} f"'] & & \mathcal{T}_{\Sigma,E} Y
\end{tikzcd}
$$

We can show all the other faces commute, and then use the fact that $\mathcal{T}_\Sigma[-]_E$ is surjective (i.e. epic) to conclude that the front face must also commute.[112] The first diagram we paved implies (1.28) commutes because $[-]_E$ is epic. □

[112] In more details, the left and right faces commute by (1.25), the bottom and top faces commute by (1.23), and the back face commutes by (1.7).
The function $\mathcal{T}_\Sigma[-]_E$ 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 $\mathcal{T}_\Sigma[-]_E$ 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 complete derivation (we write which face was used as justifications for each step).

$$
\begin{aligned}
& \mathcal{T}_{\Sigma,E} f \circ [\![-]\!]_{\mathbb{T}X} \circ \mathcal{T}_\Sigma[-]_E & \\
&= \mathcal{T}_{\Sigma,E} f \circ [-]_E \circ \mu^\Sigma_X & \text{left} \\
&= [-]_E \circ \mathcal{T}_\Sigma f \circ \mu^\Sigma_X & \text{bottom} \\
&= [-]_E \circ \mu^\Sigma_Y \circ \mathcal{T}_\Sigma\mathcal{T}_\Sigma f & \text{back} \\
&= [\![-]\!]_{\mathbb{T}Y} \circ \mathcal{T}_\Sigma[-]_E \circ \mathcal{T}_\Sigma\mathcal{T}_\Sigma f & \text{right} \\
&= [\![-]\!]_{\mathbb{T}Y} \circ \mathcal{T}_\Sigma\mathcal{T}_{\Sigma,E} f \circ \mathcal{T}_\Sigma[-]_E & \text{top}
\end{aligned}
$$

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

$$
\begin{tikzcd}
\mathcal{T}_\Sigma\mathcal{T}_{\Sigma,E} X \arrow[r, "\mathcal{T}_\Sigma\mathcal{T}_{\Sigma,E} f"] \arrow[d, "{[\![-]\!]_{\mathbb{T}X}}"'] & \mathcal{T}_\Sigma\mathcal{T}_{\Sigma,E} Y \arrow[d, "{[\![-]\!]_{\mathbb{T}Y}}"] \\
\mathcal{T}_{\Sigma,E} X \arrow[r, "\mathcal{T}_{\Sigma,E} f"'] & \mathcal{T}_{\Sigma,E} Y
\end{tikzcd}
\tag{1.29}
$$

Stating it like this may remind you of Lemma 1.16 and Remark 1.17. We will need a variant of Lemma 1.16 for $\mathcal{T}_{\Sigma,E}$, 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 (1.30) commutes.

$$
\begin{tikzcd}
\mathcal{T}_\Sigma\mathcal{T}_{\Sigma,E} A \arrow[r, "\mathcal{T}_\Sigma{[\![-]\!]_A}"] \arrow[d, "{[\![-]\!]_{\mathbb{T}A}}"'] & \mathcal{T}_\Sigma A \arrow[d, "{[\![-]\!]_A}"] \\
\mathcal{T}_{\Sigma,E} A \arrow[r, "{[\![-]\!]_A}"'] & A
\end{tikzcd}
\tag{1.30}
$$

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 because if $[s]_E = [t]_E$, then $A \vdash s = t \in \mathfrak{Th}(E)$, so $\mathbb{A}$ satisfies that equation, and taking the assignment $\mathrm{id}_A : A \to A$, we obtain

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

When $\mathbb{A}$ is a (Σ, E)-algebra, we abusively write $[\![-]\!]_A$ for the interpretation of terms and equivalence classes of terms as in (1.31).

$$
\begin{tikzcd}
\mathcal{T}_\Sigma A \arrow[rr, "{[-]_E}"] \arrow[dr, "{[\![-]\!]_A}"'] & & \mathcal{T}_{\Sigma,E} A \arrow[dl, "{[\![-]\!]_A}"] \\
& A &
\end{tikzcd}
\tag{1.31}
$$

Lemma 1.30. *For any (Σ, E)-algebra $\mathbb{A}$, the square (1.30) commutes.*

Proof. Consider the following diagram that we can view as a triangular prism whose front face is (1.30). Both triangles commute by (1.31), the square face at the back and on the left commutes by (1.25), and the square face at the back and on the right commutes by (1.13). With the same trick as in the proof of Proposition 1.29 using the surjectivity of $\mathcal{T}_\Sigma[-]_E$, we conclude that the front face commutes.[113]

[113] Here is the complete derivation.

$$
\begin{aligned}
& [\![-]\!]_A \circ [\![-]\!]_{\mathbb{T}A} \circ \mathcal{T}_\Sigma[-]_E & \\
&= [\![-]\!]_A \circ [-]_E \circ \mu^\Sigma_A & \text{left} \\
&= [\![-]\!]_A \circ \mu^\Sigma_A & \text{bottom} \\
&= [\![-]\!]_A \circ \mathcal{T}_\Sigma[\![-]\!]_A & \text{right} \\
&= [\![-]\!]_A \circ \mathcal{T}_\Sigma[\![-]\!]_A \circ \mathcal{T}_\Sigma[-]_E & \text{top}
\end{aligned}
$$

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

$$
\begin{tikzcd}
& \mathcal{T}_\Sigma \mathcal{T}_\Sigma A \arrow[dl, "\mathcal{T}_\Sigma[-]_E"'] \arrow[dr, "\mathcal{T}_\Sigma[\![-]\!]_A"] \arrow[dd, "\mu^\Sigma_A"] & \\
\mathcal{T}_\Sigma \mathcal{T}_{\Sigma,E} A \arrow[rr, "\mathcal{T}_\Sigma[\![-]\!]_A"] \arrow[dd, "[\![-]\!]_{\mathbb{T}A}"'] & & \mathcal{T}_\Sigma A \arrow[dd, "[\![-]\!]_A"] \\
& \mathcal{T}_\Sigma A \arrow[dl, "[-]_E"'] \arrow[dr, "[\![-]\!]_A"] & \\
\mathcal{T}_{\Sigma,E} A \arrow[rr, "[\![-]\!]_A"'] & & A
\end{tikzcd}
$$

□

An important consequence of Lemma 1.16 was (1.15) saying that flattening is a homomorphism from $\mathbb{T}_{\Sigma,\emptyset}\mathbb{T}_{\Sigma,\emptyset}A$ to $\mathbb{T}_{\Sigma,\emptyset}A$. This is also true when E is not empty, i.e. $\mu_A^{\Sigma,E}$ is a homomorphism from $\mathbb{T}\mathbb{T}A$ to $\mathbb{T}A$.

Lemma 1.31. *For any set A, the following square commutes.*

$$
\begin{tikzcd}
\mathcal{T}_\Sigma \mathcal{T}_{\Sigma,E} \mathcal{T}_{\Sigma,E} A \arrow[r, "\mathcal{T}_\Sigma \mu_A^{\Sigma,E}"] \arrow[d, "[\![-]\!]_{\mathbb{T}\mathbb{T}A}"'] & \mathcal{T}_\Sigma \mathcal{T}_{\Sigma,E} A \arrow[d, "[\![-]\!]_{\mathbb{T}A}"] \\
\mathcal{T}_{\Sigma,E} \mathcal{T}_{\Sigma,E} A \arrow[r, "\mu_A^{\Sigma,E}"'] & \mathcal{T}_{\Sigma,E} A
\end{tikzcd}
\tag{1.32}
$$

Proof. We prove it exactly like Lemma 1.30 with the following diagram.[114]

$$
\begin{tikzcd}
& \mathcal{T}_\Sigma \mathcal{T}_\Sigma \mathcal{T}_{\Sigma,E} A \arrow[dl, "\mathcal{T}_\Sigma[-]_E"'] \arrow[dr, "\mathcal{T}_\Sigma[\![-]\!]_{\mathbb{T}A}"] \arrow[dd, "\mu^\Sigma_{\mathcal{T}_{\Sigma,E}A}"] & \\
\mathcal{T}_\Sigma \mathcal{T}_{\Sigma,E} \mathcal{T}_{\Sigma,E} A \arrow[rr, "\mathcal{T}_\Sigma \mu_A^{\Sigma,E}"] \arrow[dd, "[\![-]\!]_{\mathbb{T}\mathbb{T}A}"'] & & \mathcal{T}_\Sigma \mathcal{T}_{\Sigma,E} A \arrow[dd, "[\![-]\!]_{\mathbb{T}A}"] \\
& \mathcal{T}_\Sigma \mathcal{T}_{\Sigma,E} A \arrow[dl, "[-]_E"'] \arrow[dr, "[\![-]\!]_{\mathbb{T}A}"] & \\
\mathcal{T}_{\Sigma,E} \mathcal{T}_{\Sigma,E} A \arrow[rr, "\mu_A^{\Sigma,E}"'] & & \mathcal{T}_{\Sigma,E} A
\end{tikzcd}
$$

□

[114] The top and bottom faces commute by definition of $\mu_A^{\Sigma,E}$ (1.27), the back-left face by (1.25), and the back-right face by (1.13).
Then, $\mathcal{T}_\Sigma[-]_E$ is epic, so the following derivation suffices.

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

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 $\mathcal{T}_\Sigma Y$ under an assignment $\sigma : Y \to \mathcal{T}_{\Sigma,E}X$.[115] By the definition $[\![-]\!]^\sigma_{\mathbb{T}X} = [\![-]\!]_{\mathbb{T}X} \circ \mathcal{T}_\Sigma\sigma$, 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 $\sigma : X \to \mathcal{T}_{\Sigma,E}X$ that sends a variable x to its equivalence class $[x]_E$, the interpretation of a term $t \in \mathcal{T}_\Sigma X$ is $[t]_E$.[116] We prove this formally below.

[115] We used ι before for assignments, but when considering assignments into (equivalence classes of) terms, we prefer using σ because we will adopt a different attitude with them (see Definition 1.35).

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

Lemma 1.32. *Let $\sigma = X \xrightarrow{\eta^{\Sigma}_X} \mathcal{T}_{\Sigma}X \xrightarrow{[-]_E} \mathcal{T}_{\Sigma,E}X$ be an assignment. Then, $[\![-]\!]^{\sigma}_{\mathbb{T}X} = [-]_E$.*

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

$$\begin{aligned}
[\![\eta^{\Sigma}_X(x)]\!]^{\sigma}_{\mathbb{T}X} &= [\![\mathcal{T}_{\Sigma}\sigma(\eta^{\Sigma}_X(x))]\!]_{\mathbb{T}X} && \text{by (1.9)}\\
&= [\![\mathcal{T}_{\Sigma}[-]_E(\mathcal{T}_{\Sigma}\eta^{\Sigma}_X(\eta^{\Sigma}_X(x)))]\!]_{\mathbb{T}X} && \text{Proposition 1.6}\\
&= [\![\mathcal{T}_{\Sigma}[-]_E(\eta^{\Sigma}_{\mathcal{T}_{\Sigma}X}(\eta^{\Sigma}_X(x)))]\!]_{\mathbb{T}X} && \text{by (1.5)}\\
&= [\![\eta^{\Sigma}_{\mathcal{T}_{\Sigma,E}X}([\eta^{\Sigma}_X(x)]_E)]\!]_{\mathbb{T}X} && \text{by (1.5)}\\
&= [\eta^{\Sigma}_X(x)]_E && \text{by (1.26)}
\end{aligned}$$

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

$$\begin{aligned}
[\![t]\!]^{\sigma}_{\mathbb{T}X} &= [\![\mathcal{T}_{\Sigma}\sigma(t)]\!]_{\mathbb{T}X} && \text{by (1.9)}\\
&= [\![\mathcal{T}_{\Sigma}\sigma(\mathsf{op}(t_1, \dots, t_n))]\!]_{\mathbb{T}X} && \\
&= [\![\mathsf{op}(\mathcal{T}_{\Sigma}\sigma(t_1), \dots, \mathcal{T}_{\Sigma}\sigma(t_n))]\!]_{\mathbb{T}X} && \text{by (1.4)}\\
&= [\![\mathsf{op}]\!]_{\mathbb{T}X}\left([\![\mathcal{T}_{\Sigma}\sigma(t_1)]\!]_{\mathbb{T}X}, \cdots, [\![\mathcal{T}_{\Sigma}\sigma(t_n)]\!]_{\mathbb{T}X}\right) && \text{by (1.26)}\\
&= [\![\mathsf{op}]\!]_{\mathbb{T}X}\left([t_1]_E, \cdots, [t_n]_E\right) && \text{I.H.}\\
&= [\mathsf{op}(t_1, \dots, t_n)]_E. && \text{by (1.24)} \qquad \square
\end{aligned}$$

We will denote that special assignment $\eta^{\Sigma,E}_X = [-]_E \circ \eta^{\Sigma}_X : X \to \mathcal{T}_{\Sigma,E}X$.[117] A quick corollary of the previous lemma is that for any equation ϕ with context X, ϕ belongs to $\mathfrak{Th}(E)$ if and only if the algebra $\mathbb{T}_{\Sigma,E}X$ satisfies it under the assignment $\eta^{\Sigma,E}_X$. This comes back to Example 1.23 where we said that freeness of X^* means it satisfies all and only the equations in $\mathfrak{Th}(E_{\mathbf{Mon}})$. Instead here, we do not know yet that $\mathbb{T}X$ is free (we have not even proved it satisfies E yet), but we can already show it satisfies only the necessary equations, and freeness will follow.

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

Lemma 1.33. *Let $s, t \in \mathcal{T}_{\Sigma}X$, $X \vdash s = t \in \mathfrak{Th}(E)$ if and only if $\mathbb{T}_{\Sigma,E}X \vDash^{\eta^{\Sigma,E}_X} X \vdash s = t$.*[118]

[118] *Proof.* By Lemma 1.32, 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 $\equiv_E$, $X \vdash s = t \in \mathfrak{Th}(E)$ 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 1.34. *The following diagram commutes.*

$$\begin{tikzcd}
\mathcal{T}_{\Sigma,E}X \arrow[r, "\eta^{\Sigma,E}_{\mathcal{T}_{\Sigma,E}X}"] \arrow[dr, "\mathrm{id}_{\mathcal{T}_{\Sigma,E}X}"'] & \mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}X \arrow[d, "\mu^{\Sigma,E}_X"] & \mathcal{T}_{\Sigma,E}X \arrow[l, "\mathcal{T}_{\Sigma,E}\eta^{\Sigma,E}_X"'] \arrow[dl, "\mathrm{id}_{\mathcal{T}_{\Sigma,E}X}"] \\
& \mathcal{T}_{\Sigma,E}X &
\end{tikzcd}$$

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

$$\begin{tikzcd}
\mathcal{T}_{\Sigma,E}X \arrow[r, "\eta^{\Sigma}_{\mathcal{T}_{\Sigma,E}X}"] \arrow[rr, bend left=50, "\eta^{\Sigma,E}_{\mathcal{T}_{\Sigma,E}X}"] \arrow[drr, bend right, "\mathrm{id}_{\mathcal{T}_{\Sigma,E}X}"'] & \mathcal{T}_{\Sigma}\mathcal{T}_{\Sigma,E}X \arrow[r, "{[-]_E}"] \arrow[dr, "{[\![-]\!]_{\mathbb{T}X}}"'] \arrow[u, phantom, "\text{(a)}"] \arrow[d, phantom, "\text{(b)}"] & \mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}X \arrow[d, "\mu^{\Sigma,E}_X"] \arrow[dl, phantom, "\text{(c)}"] \\
& & \mathcal{T}_{\Sigma,E}X
\end{tikzcd} \tag{1.33}$$

Showing (1.33) commutes:
(a) Definition of $\eta^{\Sigma,E}_X$.
(b) Definition of $[\![-]\!]_{\mathbb{T}X}$ (1.26).
(c) Definition of $\mu^{\Sigma,E}_X$ (1.27).

For the triangle on the right, we show that $[-]_E = \mu_X^{\Sigma,E} \circ \mathcal{T}_{\Sigma,E}\eta_X^{\Sigma,E} \circ [-]_E$ by paving (1.34), and we can conclude since $[-]_E$ is epic that $\mathrm{id}_{\mathcal{T}_{\Sigma,E}X} = \mu_X^{\Sigma,E} \circ \mathcal{T}_{\Sigma,E}\eta_X^{\Sigma,E}$.

$$\begin{tikzcd}
\mathcal{T}_\Sigma X \arrow[r, "{[-]_E}"] \arrow[drr, "\mathcal{T}_\Sigma\eta_X^\Sigma"'] \arrow[ddrr, "\mathrm{id}_{\mathcal{T}_\Sigma X}"'] & \mathcal{T}_{\Sigma,E}X \arrow[r, "\mathcal{T}_{\Sigma,E}\eta_X^\Sigma"] \arrow[rrr, bend left, "\mathcal{T}_{\Sigma,E}\eta_X^{\Sigma,E}"] \arrow[d, phantom, "\text{(b)}"] & \mathcal{T}_{\Sigma,E}\mathcal{T}_\Sigma X \arrow[rr, "{\mathcal{T}_{\Sigma,E}[-]_E}"] \arrow[rr, phantom, bend left, "\text{(a)}"] \arrow[dr, phantom, "\text{(c)}"] & & \mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}X \arrow[dd, "\mu_X^{\Sigma,E}"] \\
\arrow[dr, phantom, "\text{(d)}"] & & \mathcal{T}_\Sigma\mathcal{T}_\Sigma X \arrow[u, "{[-]_E}"'] \arrow[r, "{\mathcal{T}_\Sigma[-]_E}"'] \arrow[d, "\mu_X^\Sigma"] \arrow[dr, phantom, "\text{(e)}"] & \mathcal{T}_\Sigma\mathcal{T}_{\Sigma,E}X \arrow[ur, "{[-]_E}"'] \arrow[dr, "{[\![-]\!]_{\mathbb{T}X}}"'] \arrow[r, phantom, "\text{(f)}"] & \\
& & \mathcal{T}_\Sigma X \arrow[rr, "{[-]_E}"'] & & \mathcal{T}_{\Sigma,E}X
\end{tikzcd} \tag{1.34}$$

Showing (1.34) commutes:

(a) Definition of $\eta_X^{\Sigma,E}$ and functoriality of $\mathcal{T}_{\Sigma,E}$.

(b) Naturality of $[-]_E$ (1.23).

(c) Naturality of $[-]_E$ again.

(d) Definition of μ_X^Σ (1.6).

(e) By (1.25).

(f) By (1.27).

□

We single out another special case of interpretation in a term algebra when E is empty (recall from Remark 1.27 that $\mathbb{T}_{\Sigma,\emptyset}X$ is the algebra on $\mathcal{T}_\Sigma X$ whose interpretation of op applies op syntactically).

Definition 1.35 (Substitution)**.** Given a signature Σ, an empty set of equations, and an assignment $\sigma : Y \to \mathcal{T}_\Sigma X$,[119] we call $[\![-]\!]^\sigma_{\mathbb{T}X}$ the **substitution** map, and we denote it by $\sigma^* : \mathcal{T}_\Sigma Y \to \mathcal{T}_\Sigma X$. We saw in Remark 1.27 that $[\![-]\!]_{\mathbb{T}X} = \mu_X^\Sigma$, thus substitution is

$$\sigma^* = \mathcal{T}_\Sigma Y \xrightarrow{\mathcal{T}_\Sigma\sigma} \mathcal{T}_\Sigma\mathcal{T}_\Sigma X \xrightarrow{\mu_X^\Sigma} \mathcal{T}_\Sigma X. \tag{1.35}$$

In words, σ^* replaces the occurrences of a variable y by $\sigma(y)$.[120]

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.[121]

Lemma 1.36. *Let $Y \vdash s = t$ be an equation, $\sigma : Y \to \mathcal{T}_\Sigma X$ 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.

$$\begin{aligned}
[\![-]\!]^{\iota_\sigma}_A &= [\![-]\!]_A \circ \mathcal{T}_\Sigma(\iota_\sigma) && \text{by (1.9)} \\
&= [\![-]\!]_A \circ \mathcal{T}_\Sigma([\![\sigma(-)]\!]^\iota_A) && \text{definition of } \iota_\sigma \\
&= [\![-]\!]_A \circ \mathcal{T}_\Sigma\left([\![-]\!]_A \circ \mathcal{T}_\Sigma\iota \circ \sigma\right) && \text{by (1.9)} \\
&= [\![-]\!]_A \circ \mathcal{T}_\Sigma[\![-]\!]_A \circ \mathcal{T}_\Sigma\mathcal{T}_\Sigma\iota \circ \mathcal{T}_\Sigma\sigma && \text{Proposition 1.6} \\
&= [\![-]\!]_A \circ \mu_A^\Sigma \circ \mathcal{T}_\Sigma\mathcal{T}_\Sigma\iota \circ \mathcal{T}_\Sigma\sigma && \text{by (1.13)} \\
&= [\![-]\!]_A \circ \mathcal{T}_\Sigma\iota \circ \mu_Y^\Sigma \circ \mathcal{T}_\Sigma\sigma && \text{by (1.7)}
\end{aligned}$$

[119] We can identify $\mathcal{T}_\Sigma X$ with $\mathcal{T}_{\Sigma,\emptyset}X$ because $\equiv_\emptyset$ is the equality relation.

[120] 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

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

[121] We will give more intuition on Lemma 1.36 when we define equational logic.

$$
\begin{aligned}
&= [\![-]\!]_A \circ \mathcal{T}_\Sigma \iota \circ \sigma^* && \text{by (1.35)}\\
&= [\![\sigma^*(-)]\!]^\iota_A. && \text{by (1.9)} \qquad \square
\end{aligned}
$$

We are finally ready to show that $\mathbb{T}_{\Sigma,E}A$ is a (Σ, E)-algebra.[122]

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

Proposition 1.37. *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 $\iota : X \to \mathcal{T}_{\Sigma,E}A$ be an assignment. We need to show that $[\![s]\!]^\iota_{\mathbb{T}A} = [\![t]\!]^\iota_{\mathbb{T}A}$. We factor ι into[123]

$$\iota = X \xrightarrow{\eta^{\Sigma,E}_X} \mathcal{T}_{\Sigma,E}X \xrightarrow{\mathcal{T}_{\Sigma,E}\iota} \mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}A \xrightarrow{\mu^{\Sigma,E}_A} \mathcal{T}_{\Sigma,E}A.$$

[123] This factoring is correct because

$$
\begin{aligned}
\iota &= \mathrm{id}_{\mathcal{T}_{\Sigma,E}A} \circ \iota\\
&= \mu^{\Sigma,E}_A \circ \eta^{\Sigma,E}_{\mathcal{T}_{\Sigma,E}A} \circ \iota && \text{Lemma 1.34}\\
&= \mu^{\Sigma,E}_A \circ \mathcal{T}_{\Sigma,E}\iota \circ \eta^{\Sigma,E}_X. && \text{naturality of } \eta^{\Sigma,E}
\end{aligned}
$$

Now, Lemma 1.33 says that the equation is satisfied in $\mathbb{T}X$ under the assignment $\eta^{\Sigma,E}_X$, i.e. that $[\![s]\!]^{\eta^{\Sigma,E}_X}_{\mathbb{T}X} = [\![t]\!]^{\eta^{\Sigma,E}_X}_{\mathbb{T}X}$. We also know by Lemma 1.15 that homomorphisms preserve satisfaction, so we can apply it twice using the facts that $\mathcal{T}_{\Sigma,E}\iota$ and $\mu^{\Sigma,E}_A$ are homomorphisms (by (1.29) and (1.32) respectively) to conclude that

$$[\![s]\!]^\iota_{\mathbb{T}A} = [\![s]\!]^{\mu^{\Sigma,E}_A \circ \mathcal{T}_{\Sigma,E}\iota \circ \eta^{\Sigma,E}_X}_{\mathbb{T}A} = [\![t]\!]^{\mu^{\Sigma,E}_A \circ \mathcal{T}_{\Sigma,E}\iota \circ \eta^{\Sigma,E}_X}_{\mathbb{T}A} = [\![t]\!]^\iota_{\mathbb{T}A}. \qquad \square$$

We now know that $\mathbb{T}_{\Sigma,E}X$ belongs to $\mathbf{Alg}(\Sigma, E)$. In order to tie up the parallel with Example 1.23, we will show that $\mathbb{T}_{\Sigma,E}X$ is the free (Σ, E)-algebra over X.

Definition 1.38 (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 (relative 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.[124]

[124] This is almost a copy of (1.20).

$$
\begin{tikzcd}
\text{in } \mathbf{C} & & & \text{in } \mathbf{D}\\
X \arrow[r, "i"] \arrow[dr, "f"'] & UY \arrow[d, dashed, "Uf^*"] & & Y \arrow[d, dashed, "f^*"]\\
& UA & \arrow[l, "U"'] & A
\end{tikzcd}
\tag{1.36}
$$

Proposition 1.39. *Free objects are unique up to isomorphism, namely, if Y and Y' are free objects on X, then $Y \cong Y'$.*[125]

[125] 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 1.40. *For any set X, the term algebra $\mathbb{T}_{\Sigma,E}X$ 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 $f^* = [\![-]\!]_A \circ \mathcal{T}_{\Sigma,E}f$ is the unique homomorphism making the following commute.

$$
\begin{tikzcd}
\text{in } \mathbf{Set} & & & \text{in } \mathbf{Alg}(\Sigma,E)\\
X \arrow[r, "\eta^{\Sigma,E}_X"] \arrow[dr, "f"'] & \mathcal{T}_{\Sigma,E}X \arrow[d, dashed, "f^*"] & & \mathbb{T}X \arrow[d, dashed, "f^*"]\\
& A & \arrow[l, "U"'] & \mathbb{A}
\end{tikzcd}
$$

First, f^* is a homomorphism because it is the composite of two homomorphisms $\mathcal{T}_{\Sigma,E}f$ (by (1.29)) and $[\![-]\!]_A$ (by Lemma 1.30 since $\mathbb{A}$ satisfies E). Next, the triangle commutes by the following derivation.

$$[\![-]\!]_A \circ \mathcal{T}_{\Sigma,E}f \circ \eta^{\Sigma,E}_X = [\![-]\!]_A \circ \eta^{\Sigma,E}_A \circ f \qquad \text{naturality of } \eta^{\Sigma,E}$$

$$
\begin{aligned}
&= [\![-]\!]_A \circ [-]_E \circ \eta_A^\Sigma \circ f && \text{definition of } \eta^{\Sigma,E} \\
&= [\![-]\!]_A \circ \eta_A^\Sigma \circ f && \text{by (1.31)} \\
&= f && \text{definition of } [\![-]\!]_A \text{ (1.8)}
\end{aligned}
$$

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.[126] □

[126] 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 $t \in \mathcal{T}_\Sigma X$, we showed in Lemma 1.32 that $[t]_E = [\![t]\!]_{\mathbb{T}X}^{\eta_X^{\Sigma,E}}$. Then using (1.11), 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, the most wonderful mathematical object in the world (objectively). Unfortunately, our universal 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. Everything is done modulo $\equiv_E$ which was defined in (1.21) to basically contain all the equations in $\mathfrak{Th}(E)$, that is, all the equations semantically entailed by E. Thanks to Lemma 1.33, if we want to know whether $X \vdash s = t$ is in $\mathfrak{Th}(E)$, it is enough to check if the free (Σ, E)-algebra $\mathbb{T}X$ satisfies it, but that is a circular argument since the carrier $\mathcal{T}_{\Sigma,E}X$ is defined via $\equiv_E$.

Equational logic is a deductive system which produces an alternative definition of the free algebra, relying only on syntax. In short, the rules of equational logic allow to syntactically derive all of $\mathfrak{Th}(E)$ starting from E.

In Lemma 1.24, we proved that $\equiv_E$ is a congruence (i.e. reflexive, symmetric, transitive, and invariant under operations), and in Lemma 1.36 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. If we know $x, y \vdash x = y \in E$, then we can conclude $X \vdash s = t \in \mathfrak{Th}(E)$, i.e. all terms are equal modulo E, by substituting x with s and y with t. This can be summarized with the inference rules of **equational logic** in Figure 1.3.

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

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

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

Figure 1.3: 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 term in $\mathcal{T}_\Sigma X$ (or $\mathcal{T}_\Sigma Y$ for Sub). 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.

The first four rules are fairly simple, and they essentially say that equality is an equivalence relation that is preserved by operations. The Sub rule looks a bit more complicated, it is named after the function σ^* used in the conclusion which we called substitution. Intuitively, it reflects the fact that variables in the context Y are universally quantified. If you know $Y \vdash s = t$ holds, then you can replace each variable $y \in Y$ by $\sigma(y)$ (which may even be a complex term using new variables in X), and you can prove that $X \vdash \sigma^*(s) = \sigma^*(t)$ holds. We did this in Lemma 1.36, and the argument to extract from there is that the interpretation of $\sigma^*(t)$ under some assignment $\iota : X \to A$ is equal to the interpretation of t under the assignment ι_σ sending $y \in Y$ to the interpretation of $\sigma(y)$ under ι. Since satisfaction of $Y \vdash s = t$ means satisfaction under any assignment (this is where universal quantification comes in), we conclude that $X \vdash \sigma^*(s) = \sigma^*(t)$ must be satisfied.

If you have written sequences of computations to solve a mathematical problem, you are already familiar with the essence of doing proofs in equational logic. The rigorous details of such proofs can be formalized with the following definition.

Definition 1.41 (Derivation)**.** A **derivation**[127] of $X \vdash s = t$ in equational logic with axioms E (a class of equations) is a finite rooted tree such that:

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

- all nodes are labelled by equations,
- the root is labelled by $X \vdash s = t$,
- if an internal node (not a leaf) is labelled by ϕ and its children are labelled by $\phi_1, \dots, \phi_n$, then there is a rule in Figure 1.3 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 $u \in \mathcal{T}_\Sigma Y$.

Example 1.42. We write a derivation with the same notation used to specify the inference rules in Figure 1.3. Consider the signature $\Sigma = \{+:2, \mathsf{e}:0\}$ with E containing the equations defining commutative monoids in (1.17). Here is a derivation of $x, y, z \vdash x + (y + z) = z + (x + y)$ in equational logic with axioms E.

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

Given any class 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 now 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, ϕ is satisfied by all (Σ, E)-algebras, and completeness says that whenever an equation ϕ is satisfied by all (Σ, E)-algebras, there is a derivation of ϕ in equational logic with axioms E.

Soundness is a straightforward consequence of earlier results.[128]

[128] In the story we are telling, 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 1.43 (Soundness)**.** *If $\phi \in \mathfrak{Th}'(E)$, then $\phi \in \mathfrak{Th}(E)$.*

Proof. In the proof of Lemma 1.24, 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 1.36 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.[129]

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. $\phi \in \mathfrak{Th}(E)$. $\square$

[129] This is a standard theorem of first order logic:

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

Completeness is the harder direction, and there are many ways to prove it.[130] We will define an algebra exactly like $\mathbb{T}X$ but using the equality relation induced by $\mathfrak{Th}'(E)$ instead of $\equiv_E$ which was induced by $\mathfrak{Th}(E)$. We then show that algebra is a (Σ, E)-algebra, and by construction, it will imply $\mathfrak{Th}(E) \subseteq \mathfrak{Th}'(E)$.

[130] The original proof of Birkhoff [Bir35, Theorem 10] relies on constructing free algebras. Several later proofs (e.g. [Wec12, Theorem 29]) rely on a theory of congruences.

Fix a signature Σ and a class E of equations over Σ. For any set X, we can define a binary relation $\equiv'_E$ on Σ-terms[131] that contains the pair (s,t) whenever $X \vdash s = t$ can be proven in equational logic. Formally, we have for any $s, t \in \mathcal{T}_\Sigma X$ (c.f. (1.21)),

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

$$s \equiv'_E t \iff X \vdash s = t \in \mathfrak{Th}'(E). \tag{1.37}$$

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

Lemma 1.44. *For any set X, the relation $\equiv'_E$ is reflexive, symmetric, transitive, and for any $\mathsf{op} : n \in \Sigma$ and $s_1, \dots, s_n, t_1, \dots, t_n \in \mathcal{T}_\Sigma X$,*[132]

[132] i.e. $\equiv'_{\hat{E}}$ is a congruence on the Σ-algebra $\mathcal{T}_\Sigma X$ defined in Remark 1.17.

$$(\forall 1 \leq i \leq n, s_i \equiv'_E t_i) \implies \mathsf{op}(s_1, \dots, s_n) \equiv'_E \mathsf{op}(t_1, \dots, t_n). \tag{1.38}$$

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

We write $\wr - \int_E : \mathcal{T}_\Sigma X \to \mathcal{T}_\Sigma X/\equiv'_E$ 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 1.45 (Term algebra, syntactically)**.** The *new* term algebra for (Σ, E) on X is the Σ-algebra whose carrier is $\mathcal{T}_\Sigma X/\equiv'_E$ and whose interpretation of $\mathsf{op} : n \in \Sigma$ is defined by[133]

[133] This is well-defined (i.e. invariant under change of representative) by (1.38).

$$[\![\mathsf{op}]\!]_{\mathbb{T}'X}(\wr t_1 \int_E, \dots, \wr t_n \int_E) = \wr \mathsf{op}(t_1, \dots, t_n) \int_E. \tag{1.39}$$

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

With soundness (Theorem 1.43) of equational logic, completeness would mean this alternative definition of the term algebra coincides with $\mathbb{T}X$. First, we have to show that $\mathbb{T}'X$ belongs to $\mathbf{Alg}(\Sigma, E)$ like we did for $\mathbb{T}X$ in Proposition 1.37, and we prove a technical lemma before that.

Lemma 1.46. *Let $\iota : Y \to \mathcal{T}_\Sigma X/\equiv'_E$ be an assignment. For any function $\sigma : Y \to \mathcal{T}_\Sigma X$ satisfying $\wr \sigma(y) \int_E = \iota(y)$ for all $y \in Y$, we have $[\![-]\!]^\iota_{\mathbb{T}'X} = \wr \sigma^*(-) \int_E$.*[134]

[134] This result looks like a stronger version of Lemma 1.32 for $\mathbb{T}'X$. Morally, they are both saying that interpretation of terms in $\mathbb{T}X$ or $\mathbb{T}'X$ is just a syntactical matter.

Proof. We proceed by induction. For the base case, we have by definition of the interpretation of terms (1.8), definition of σ, and definition of σ^* (1.35),

$$[\![\eta_Y^\Sigma(y)]\!]_{\mathbb{T}'X}^{\iota} \overset{(1.8)}{=} \iota(y) = \wr \sigma(y) \int_E \overset{(1.35)}{=} \wr \sigma^*(\eta_Y^\Sigma(y)) \int_E.$$

For the inductive step, we have

$$\begin{aligned}
[\![\mathsf{op}(t_1,\dots,t_n)]\!]_{\mathbb{T}'X}^{\iota} &= [\![\mathsf{op}]\!]_{\mathbb{T}'X}([\![t_1]\!]_{\mathbb{T}'X}^{\iota},\dots,[\![t_n]\!]_{\mathbb{T}'X}^{\iota}) && \text{by (1.8)}\\
&= [\![\mathsf{op}]\!]_{\mathbb{T}'X}(\wr \sigma^*(t_1) \int_E,\dots,\wr \sigma^*(t_n) \int_E) && \text{I.H.}\\
&= \wr \mathsf{op}(\sigma^*(t_1),\dots,\sigma^*(t_n)) \int_E && \text{by (1.39)}\\
&= \wr \sigma^*(\mathsf{op}(t_1,\dots,t_n)) \int_E. && \text{definition of } \sigma^*
\end{aligned}$$

□

Proposition 1.47. *For any set X, $\mathbb{T}'X$ satisfies all the equations in E.*

Proof. Let $Y \vdash s = t$ belong to E and $\iota : Y \to \mathcal{T}_\Sigma X/{\equiv'_E}$ be an assignment. By the axiom of choice,[135] there is a function $\sigma : Y \to \mathcal{T}_\Sigma X$ satisfying $\wr \sigma(y) \int_E = \iota(y)$ for all $y \in Y$. Thanks to Lemma 1.46, it is enough to show $\wr \sigma^*(s) \int_E = \wr \sigma^*(t) \int_E$.[136] Equivalently, by definition of $\wr - \int_E$ and $\mathfrak{Th}'(E)$, we can just 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 $\sigma : Y \to \mathcal{T}_\Sigma X$ and the equation $Y \vdash s = t$ which is an axiom. □

[135] Choice implies the quotient map $\wr - \int_E$ has a right inverse $r : \mathcal{T}_\Sigma X/{\equiv'_E} \to \mathcal{T}_\Sigma X$, and we can then set $\sigma = r \circ \iota$.

[136] By Lemma 1.46, it implies

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

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 1.48 (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 1.47 and definition of $\mathfrak{Th}(E)$, we know that $\mathbb{T}'X \vDash \phi$. In particular, $\mathbb{T}'X$ satisfies ϕ under the assignment

$$\iota = X \xrightarrow{\eta_X^\Sigma} \mathcal{T}_\Sigma X \xrightarrow{\wr - \int_E} \mathcal{T}_\Sigma X/{\equiv'_E},$$

namely, $[\![s]\!]_{\mathbb{T}'X}^{\iota} = [\![t]\!]_{\mathbb{T}'X}^{\iota}$. Moreover with $\sigma = \eta_X^\Sigma$, we can show σ satisfies the hypothesis of Lemma 1.46 and $\sigma^* = \mathrm{id}_{\mathcal{T}_\Sigma X}$,[137] thus we conclude

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

This implies $s \equiv'_E t$ which in turn means $X \vdash s = t$ belongs to $\mathfrak{Th}'(E)$. □

[137] We defined ι precisely to have $\wr \sigma(x) \int_E = \iota(x)$. To show $\sigma^* = \eta_X^{\Sigma *}$ is the identity, use (1.35) and the fact that $\mu^\Sigma \cdot \mathcal{T}_\Sigma \eta^\Sigma = \mathbb{1}_{\mathcal{T}_\Sigma}$ (Lemma 1.9).

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.[138]

[138] It is good to keep in mind these two equivalent definitions of the free (Σ, E)-algebra on X. It means you can prove s equals t in $\mathbb{T}X$ by exhibiting a derivation of $X \vdash s = t$ in equational logic, or you can prove $s \neq t$ by exhibiting an algebra that satisfies E but not $X \vdash s = t$.

Remark 1.49. We have used the axiom of choice in proving completeness of equational logic, but that is only an artifact of our presentation that deals with arbitrary contexts. Since terms are finite and operation symbols have finite arities, we can 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.[139] You can also verify semantically that $\mathbb{A}$ satisfies $X \vdash s = t$ if and only if it satisfies $\mathrm{FV}\{s,t\} \vdash s = t$ essentially because the extra variables have no effect on the quantification of the free variables in s and t nor on the interpretation.

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

$$\begin{aligned}
\mathrm{FV}\{\eta_X^\Sigma(x)\} &= \{x\}\\
\mathrm{FV}\{\mathsf{op}(t_1,\dots,t_n)\} &= \mathrm{FV}\{t_1\} \cup \cdots \cup \mathrm{FV}\{t_n\}\\
\mathrm{FV}\{t_1,\dots,t_n\} &= \mathrm{FV}\{t_1\} \cup \cdots \cup \mathrm{FV}\{t_n\}.
\end{aligned}$$

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. First, for any set X and variable y, the following inference rules are derivable in equational logic.

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

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[140], then since y is not in the free variables of s and t, $\sigma^*(s) = s$ and $\sigma^*(t) = t$.

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

Second, we allowed the collection of equations E generating an algebraic theory $\mathfrak{Th}(E)$ to be a proper class, and that is really not common. Oftentimes, a countable set of variables $\{x_1, x_2, \dots\}$ is assumed, and equations are defined only when with a context contained in that set. With this assumption, the collection of all equations is a set, and so are E and $\mathfrak{Th}(E)$. This has no effect on expressiveness since for any equation $X \vdash s = t$, there is an equivalent equation $X' \vdash s' = t'$ with $X' \subseteq \{x_1, x_2, \dots\}$.[141]

[141] We already know $X \vdash s = t$ is equivalent to $\mathrm{FV}\{s,t\} \vdash s = t$, and since the context of the latter is finite, we have a bijection $\sigma : \mathrm{FV}\{s,t\} \cong \{x_1, \dots, x_n\}$. Then the SUB rule instantiated with σ and σ^{-1} proves the desired equivalence.

1.4 Monads

Our presentation of universal algebra used the language of category theory, e.g. functors, natural transformations, commutative diagrams. Both these fields of mathematics were born within a decade of each other[142] with a similar goal: abstracting the way mathematicians use mathematical objects in order to apply one general argument to many specific cases.[143] One could argue (looking at today's practicing mathematicians) that category theory was more successful. This is why a portion of this manuscript is spent on monads, a more categorical formulation of the content in universal algebra which became popular in computer science after Moggi's work [Mog89, Mog91] using monads to abstract computational effects.

[142] [Bir33, Bir35] and [EM45] were the seminal papers for universal algebra and category theory respectively. Birkhoff and MacLane even wrote an undergraduate textbook together [MB99].

[143] This is very close to a goal of mathematics as a whole: abstracting the way nature works in order to apply one general argument to many specific cases, c.f. Cheng calling category theory the "mathematics of mathematics" [Che16].

There is another categorical approach to universal algebra introduced by Lawvere [Law63] and first popularized in the computer science community by Hyland, Plotkin, and Power [PP01a, PP01b, HPP06, HP07]. We will stick to monads because most of the literature on quantitative algebras does, and because I am not sure yet how the generalizations we contributed port to Lawvere's approach.[144]

[144] In the paper introducing quantitative algebra [MPP16], the authors already mentioned enriched Lawvere theories [Pow99]. The works of Power [Pow05], Lucyshyn-Wright and Parker [LW16, LP23] and Rosický [Ros24] are also relevant.

Definition 1.50 (Monad)**.** A **monad** on a category $\mathbf{C}$ is a triple (M, η, μ) made up 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 (1.40) and (1.41) commute in $[\mathbf{C}, \mathbf{C}]$.[145]

[145] I also recommend Marsden's series of blog posts on monads for a relatively light and comprehensive survey: `https://stringdiagram.com/2022/05/17/hello-monads/`.

$$\begin{tikzcd} M \arrow[r, "M\eta"] \arrow[dr, "\mathbb{1}_M"'] & M^2 \arrow[d, "\mu"] & M \arrow[l, "\eta M"'] \arrow[dl, "\mathbb{1}_M"] \\ & M & \end{tikzcd} \tag{1.40}$$

$$\begin{tikzcd} M^3 \arrow[r, "\mu M"] \arrow[d, "M\mu"'] & M^2 \arrow[d, "\mu"] \\ M^2 \arrow[r, "\mu"'] & M \end{tikzcd} \tag{1.41}$$

We often refer to the monad (M, η, μ) simply with M.

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 computer scientists and especially for this manuscript).

Example 1.51 (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** (the name "option monad" is also common).[146] We write inl^{X+Y} (resp. inr^{X+Y}) for the coprojection of X (resp. Y) into $X + Y$.[147] First, note that for a morphism $f : X \to Y$,

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

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 (1.40) and (1.41) 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 $\{*\}$:[148]

- $X + \mathbf{1}$ 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 (1.40) and (1.41) commute by hand because (briefly) $x \in X$ is always sent to $x \in X$ and $*$ is always sent to $*$.

More often than not, the fresh element $*$ is seen as a terminating state, so the maybe monad models the most basic computational effect. Even when no other observation can be made on states of a program, one can distinguish between states by looking at their execution traces which may or may not contain $*$.[149]

Example 1.52 (Powerset). The covariant **non-empty finite powerset** functor $\mathcal{P}_{\mathsf{ne}} : \mathbf{Set} \to \mathbf{Set}$ sends a set X to the set of non-empty finite subsets of X which we denote by $\mathcal{P}_{\mathsf{ne}}X$. It acts on functions just like the usual powerset functor, i.e. given a function $f : X \to Y$, $\mathcal{P}_{\mathsf{ne}}f$ is the direct image function, it sends $S \subseteq X$ to $f(S) = \{f(x) \mid x \in S\}$.[150]

One can show $\mathcal{P}_{\mathsf{ne}}$ is a monad with the following unit and multiplication:[151]

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

Again, as early as in Moggi's papers, the powerset monad was used to model nondeterministic computations (see also [VW06, KS18, BSV19, GPA21]). A set

[146] It is also called the lift monad in [Jac16, Example 5.1.3.2].

[147] These notations are common in the community of programming language research, they stand for *injection left* (resp. *right*). We may omit the superscript.

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

[149] This was already known to Moggi who used different terminology in [Mog91, Example 1.1].

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

[151] Note that $\{x\}$ is non-empty and finite, and so is $\cup_{s \in F} s$ whenever F and all $s \in F$ are non-empty and finite. Thus, we can define $\mathcal{P}_{\mathsf{ne}}$ as a submonad of the *full* powerset monad in, e.g., [Jac16, Example 5.1.3.1].

$S \in \mathcal{P}_{\mathrm{ne}}X$ is seen as all the possible states at a point in the execution. We assume that S is finite for convenience, and that it is non-empty because an empty set of possible states would mean termination which can already be modelled with the maybe monad.[152]

[152] Also, the maybe monad can be *combined* with any other monad, see for example [MSV21, Corollary 5].

Example 1.53 (Distributions)**.** The functor $\mathcal{D} : \mathbf{Set} \to \mathbf{Set}$ sends a set X to the set of **finitely supported distributions** on X:[153]

[153] We will simply call them distributions.

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

We call $\varphi(x)$ the **weight** of φ at x and let $\mathrm{supp}(\varphi)$ denote the **support** of φ, that is, $\mathrm{supp}(\varphi)$ contains all the elements $x \in X$ such that $\varphi(x) \neq 0$.[154] On morphisms, $\mathcal{D}$ sends a function $f : X \to Y$ to the function between sets of distributions defined by

[154] We often write $\varphi(S)$ for the total weight of φ on all of $S \subseteq X$.

$$\mathcal{D}f : \mathcal{D}X \to \mathcal{D}Y = \varphi \mapsto \left(y \mapsto \sum_{x \in X, f(x) = y} \varphi(x) \right). \tag{1.42}$$

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.[155]

[155] The distribution $\mathcal{D}f(\varphi)$ is sometimes called the **pushforward** of φ.

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

$$\mu_X = \Phi \mapsto \left(x \mapsto \sum_{\varphi \in \mathrm{supp}(\Phi)} \Phi(\varphi)\varphi(x) \right). \tag{1.43}$$

In words, the weight $\mu_X(\Phi)$ at x is the average of $\varphi(x)$ weighted by $\Phi(\varphi)$ for all distributions in the support of Φ.[156]

[156] It was Giry [Gir82] who first studied probabilities through the categorical lens with a monad with inspiration from Lawvere [Law62], $\mathcal{D}$ is a discrete version of Giry's original construction. (See [Jac16, Example 5.1.3.4].)

Moggi only hinted at the distribution monad being a good model for computations that rely on random/probabilistic choices. For fleshed out research see, e.g., [VW06, SW18, BSV19].

Monads have been a popular categorical approach to universal algebra[157] thanks to a result of Linton [Lin66, Proposition 1] stating that any algebraic theory gives rise to a monad. Given a signature Σ and a class E of equations, we already implicitly described the monad Linton constructed, it is the triple $(\mathcal{T}_{\Sigma,E}, \eta^{\Sigma,E}, \mu^{\Sigma,E})$.

[157] See [HP07] for a thorough survey on categorical approaches to universal algebra.

Proposition 1.54. *The functor $\mathcal{T}_{\Sigma,E} : \mathbf{Set} \to \mathbf{Set}$ defines a monad on $\mathbf{Set}$ 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.[158] We showed that $\eta^{\Sigma,E}$ and $\mu^{\Sigma,E}$ are natural transformations of the right type in Footnote 117 and Proposition 1.29 respectively, and we showed the appropriate instance of (1.40) commutes in Lemma 1.34. It remains to prove (1.41) commutes which, instantiated here, means proving the following diagram commutes for every set A.

[158] 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. $\mathcal{T}_{\Sigma,E}A$.

$$\begin{tikzcd}
\mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}A \arrow[r, "\mathcal{T}_{\Sigma,E}\mu_A^{\Sigma,E}"] \arrow[d, "\mu^{\Sigma,E}_{\mathcal{T}_{\Sigma,E}A}"'] & \mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}A \arrow[d, "\mu_A^{\Sigma,E}"] \\
\mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}A \arrow[r, "\mu_A^{\Sigma,E}"'] & \mathcal{T}_{\Sigma,E}A
\end{tikzcd}$$

It follows from the following paved diagram.[159]

$$\begin{tikzcd}
\mathcal{T}_\Sigma\mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}A \arrow[rrr, "\mathcal{T}_\Sigma\mu_A^{\Sigma,E}"] \arrow[dr, "{[-]_E}"] \arrow[dd, "{[\![-]\!]_{\mathbb{T}\mathbb{T}A}}"'] & & & \mathcal{T}_\Sigma\mathcal{T}_{\Sigma,E}A \arrow[dl, "{[-]_E}"'] \arrow[dd, "{[\![-]\!]_{\mathbb{T}A}}"] \\
& \mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}A \arrow[r, "\mathcal{T}_{\Sigma,E}\mu_A^{\Sigma,E}"'] \arrow[dl, "\mu^{\Sigma,E}_{\mathcal{T}_{\Sigma,E}A}"] & \mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}A \arrow[dr, "\mu_A^{\Sigma,E}"'] & \\
\mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}A \arrow[rrr, "\mu_A^{\Sigma,E}"'] & & & \mathcal{T}_{\Sigma,E}A
\end{tikzcd}$$

(a) (b) (c) (d)

Note that when E is empty, we get a monad $(\mathcal{T}_\Sigma, \eta^\Sigma, \mu^\Sigma)$.[160] □

[159] We know that (a), (b) and (c) commute by (1.27), (1.23), and (1.27) respectively. This means that (d) pre-composed by the epimorphism $[-]_E$ yields the outer square. Moreover, we know the outer square commutes by (1.32), therefore, (d) must also commute.

[160] Here is an alternative proof that $\mathcal{T}_\Sigma$ is a monad. We showed η^Σ and μ^Σ are natural in (1.5) and (1.7) respectively. The right triangle of (1.40) commutes by definition of μ^Σ (1.6), the left triangle commutes by Lemma 1.9, and the square (1.41) commutes by (1.15).

Linton also showed that from a monad M, you can build a theory whose corresponding term monad is isomorphic to M [Lin69, Lemma 10.1]. This however relied on a more general notion of theory. We will not go over the details here, rather we will introduce the necessary concepts to talk about our main examples on **Set**: $(-+\mathbf{1})$, $\mathcal{P}_{\mathrm{ne}}$, and $\mathcal{D}$. 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 1.55 (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 (1.44) and (1.45) commute.

$$\begin{tikzcd}
A \arrow[r, "\eta_A"] \arrow[dr, "\mathrm{id}_A"'] & MA \arrow[d, "\alpha"] \\
& A
\end{tikzcd} \qquad (1.44) \qquad\qquad
\begin{tikzcd}
MMA \arrow[r, "\mu_A"] \arrow[d, "M\alpha"'] & MA \arrow[d, "\alpha"] \\
MA \arrow[r, "\alpha"'] & A
\end{tikzcd} \qquad (1.45)$$

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

Definition 1.56 (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 (1.46) commute.

$$\begin{tikzcd}
MA \arrow[r, "Mh"] \arrow[d, "\alpha"'] & MB \arrow[d, "\beta"] \\
A \arrow[r, "h"'] & B
\end{tikzcd} \qquad (1.46)$$

The composition of two M-homomorphisms is an M-homomorphism and id_A is an M-homomorphism from (A, α) to itself, thus we get a category of M-algebras and M-homomorphisms called the **Eilenberg–Moore category** of M and denoted by $\mathbf{EM}(M)$.[161] 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.

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

Example 1.57. We will see some more concrete examples in a bit, but we can mention now that the similarities between the squares in the definitions of a monad (1.41), of an algebra (1.45), and of a homomorphism (1.46) have profound consequences. First, for any A, the pair (MA, μ_A) is an M-algebra because (1.47) and (1.48) commute by the properties of a monad.[162]

$$\begin{tikzcd} MA \arrow[r, "\eta_{MA}"] \arrow[dr, "\mathrm{id}_{MA}"'] & MMA \arrow[d, "\mu_A"] \\ & MA \end{tikzcd} \qquad (1.47) \qquad \begin{tikzcd} MMMA \arrow[r, "\mu_{MA}"] \arrow[d, "M\mu_A"'] & MMA \arrow[d, "\mu_A"] \\ MMA \arrow[r, "\mu_A"'] & MA \end{tikzcd} \qquad (1.48)$$

Furthermore, for any M-algebra $\alpha : MA \to A$, (1.45) (reflected through the diagonal) precisely says that α is a M-homomorphism from (MA, μ_A) to (A, α). After a bit more work[163] we conclude that (MA, μ_A) is the free M-algebra (relative to $U^M : \mathbf{EM}(M) \to \mathbf{Set}$).

[162] (1.47) is the component at A of the right triangle in (1.40), and (1.48) is the component at A of (1.41).

[163] Given an M-algebra (A', α') and a function $f : A \to A'$, we can show $\alpha' \circ Mf$ is the unique M-homomorphism such that $\alpha' \circ Mf \circ \eta_A = f$.

The terminology suggests that (Σ, E)-algebras and $\mathcal{T}_{\Sigma,E}$-algebras are the same thing.[164] Let us check this, obtaining a large family of examples at the same time.

[164] Also, Example 1.57 starts to confirm this if we compare it with Remark 1.17, and Lemma 1.18.

Proposition 1.58. *There is an isomorphism* $\mathbf{Alg}(\Sigma, E) \cong \mathbf{EM}(\mathcal{T}_{\Sigma,E})$.

Proof. Given a (Σ, E)-algebra $\mathbb{A}$, we already explained in (1.31) how to obtain a function $\llbracket - \rrbracket_A : \mathcal{T}_{\Sigma,E}A \to A$ which sends $[t]_E$ to the interpretation of the term t under the trivial assignment $\mathrm{id}_A : A \to A$.[165] Let us verify that $\llbracket - \rrbracket_A$ is a $\mathcal{T}_{\Sigma,E}$-algebra. We need to show the following instances of (1.44) and (1.45) commutes.

[165] That is well-defined because $\mathbb{A}$ satisfies all the equations in $\mathfrak{Th}(E)$.

$$\begin{tikzcd} A \arrow[r, "\eta^{\Sigma,E}_A"] \arrow[dr, "\mathrm{id}_A"'] & \mathcal{T}_{\Sigma,E}A \arrow[d, "\llbracket - \rrbracket_A"] \\ & A \end{tikzcd} \qquad \begin{tikzcd} \mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}A \arrow[r, "\mu^{\Sigma,E}_A"] \arrow[d, "\mathcal{T}_{\Sigma,E}\llbracket - \rrbracket_A"'] & \mathcal{T}_{\Sigma,E}A \arrow[d, "\llbracket - \rrbracket_A"] \\ \mathcal{T}_{\Sigma,E}A \arrow[r, "\llbracket - \rrbracket_A"'] & A \end{tikzcd}$$

The triangle commutes by definitions,[166] and the square commutes by the following diagram.

[166] We have $\llbracket \eta^{\Sigma,E}_A(a) \rrbracket_A = \llbracket [a]_E \rrbracket_A = \llbracket a \rrbracket_A = a$.

$$\begin{tikzcd} \mathcal{T}_\Sigma\mathcal{T}_{\Sigma,E}A \arrow[rrr, "\mathcal{T}_\Sigma\llbracket - \rrbracket_A"] \arrow[dr, "{[-]_E}"] \arrow[dd, "\llbracket - \rrbracket_{\mathbb{T}A}"'] & & & \mathcal{T}_\Sigma A \arrow[dl, "{[-]_E}"'] \arrow[dd, "\llbracket - \rrbracket_A"] \\ \arrow[r, phantom, "\text{(b)}"] & \mathcal{T}_{\Sigma,E}\mathcal{T}_{\Sigma,E}A \arrow[r, "\mathcal{T}_{\Sigma,E}\llbracket - \rrbracket_A"] \arrow[dl, "\mu^{\Sigma,E}_A"] \arrow[u, phantom, "\text{(a)}"] \arrow[d, phantom, "\text{(c)}"] & \mathcal{T}_{\Sigma,E}A \arrow[dr, "\llbracket - \rrbracket_A"] \arrow[r, phantom, "\text{(d)}"] & \\ \mathcal{T}_{\Sigma,E}A \arrow[rrr, "\llbracket - \rrbracket_A"'] & & & A \end{tikzcd}$$

Since the outer rectangle commutes by Lemma 1.30, (a) commutes by naturality of $[-]_E$ (1.23), (b) commutes by definition of $\mu^{\Sigma,E}_A$ (1.27), and (d) commutes by (1.31), we can conclude that (c) commutes because $[-]_E$ is epic.

We also already explained in Footnote 74 that any homomorphism $h : \mathbb{A} \to \mathbb{B}$

makes the outer rectangle below commute.

$$
\begin{tikzcd}
\mathcal{T}_\Sigma A \arrow[rrr, "\mathcal{T}_\Sigma h"] \arrow[dr, "{[-]_E}"] \arrow[dd, "{[\![-]\!]_A}"'] & & & \mathcal{T}_\Sigma B \arrow[dl, "{[-]_E}"'] \arrow[dd, "{[\![-]\!]_B}"] \\
\arrow[r, phantom, "\text{(b)}"] & \mathcal{T}_{\Sigma,E} A \arrow[r, "\mathcal{T}_{\Sigma,E} h"] \arrow[dl, "{[\![-]\!]_A}"] \arrow[u, phantom, "\text{(a)}"] \arrow[d, phantom, "\text{(c)}"] & \mathcal{T}_{\Sigma,E} A \arrow[dr, "{[\![-]\!]_A}"'] \arrow[r, phantom, "\text{(d)}"] & \\
A \arrow[rrr, "h"'] & & & B
\end{tikzcd}
$$

Since (a), (b), and (d) commute by naturality of $[-]_E$, (1.31), and (1.31) respectively, we conclude that (c) commutes again because $[-]_E$ is epic. This means h is a $\mathcal{T}_{\Sigma,E}$-homomorphism.

We obtain a functor[167] $P : \mathbf{Alg}(\Sigma, E) \to \mathbf{EM}(\mathcal{T}_{\Sigma,E})$ sending $\mathbb{A} = (A, [\![-]\!]_A)$ to $(A, \alpha_{\mathbb{A}})$ where $\alpha_{\mathbb{A}} = [\![-]\!]_A : \mathcal{T}_{\Sigma,E} A \to A$ (we give it a different name to make the sequel easier to follow).

[167] Checking functoriality is trivial because P acts like the identity on morphisms.

In the other direction, given an algebra $\alpha : \mathcal{T}_{\Sigma,E} A \to A$, we define an algebra $\mathbb{A}_\alpha$ with the interpretation of $\mathsf{op} : n \in \Sigma$ given by

$$[\![\mathsf{op}]\!]_\alpha(a_1, \dots, a_n) = \alpha[\mathsf{op}(a_1, \dots, a_n)]_E, \tag{1.49}$$

and we can prove by induction that $[\![t]\!]_\alpha = \alpha[t]_E$ for any Σ-term t over A (note that we use the $\mathcal{T}_{\Sigma,E}$-algebra properties of α).[168] Now, if $h : (A, \alpha) \to (B, \beta)$ is a $\mathcal{T}_{\Sigma,E}$-homomorphism, then h is a homomorphism from $\mathbb{A}_\alpha$ to $\mathbb{B}_\beta$ because for any $\mathsf{op} : n \in \Sigma$ and $a_1, \dots, a_n \in A$, we have

[168] For the base case, we have

$$[\![a]\!]_\alpha \overset{(1.8)}{=} a \overset{(1.44)}{=} \alpha[\eta^\Sigma_A(a)]_E = \alpha[a]_E.$$

For the inductive step, let $t = \mathsf{op}(t_1, \dots, t_n) \in \mathcal{T}_\Sigma A$:

$$
\begin{aligned}
[\![t]\!]_\alpha &= [\![\mathsf{op}(t_1, \dots, t_n)]\!]_\alpha \\
&= [\![\mathsf{op}]\!]_\alpha([\![t_1]\!]_\alpha, \dots, [\![t_n]\!]_\alpha) && (1.8) \\
&= [\![\mathsf{op}]\!]_\alpha(\alpha[t_1]_E, \dots, \alpha[t_n]_E) && \text{I.H.} \\
&= \alpha[\mathsf{op}(\alpha[t_1]_E, \dots, \alpha[t_n]_E)]_E && (1.49) \\
&= \alpha[\mathcal{T}_\Sigma \alpha(\mathsf{op}([t_1]_E, \dots, [t_n]_E))]_E && (1.4) \\
&= \alpha(\mathcal{T}_{\Sigma,E} \alpha[\mathsf{op}([t_1]_E, \dots, [t_n]_E)]_E) && (1.23) \\
&= \alpha(\mu^{\Sigma,E}_A[\mathsf{op}([t_1]_E, \dots, [t_n]_E)]_E) && (1.44) \\
&= \alpha[\mathsf{op}(t_1, \dots, t_n)]_E && (1.27) \\
&= \alpha[t]_E.
\end{aligned}
$$

$$
\begin{aligned}
h([\![\mathsf{op}]\!]_\alpha(a_1, \dots, a_n)) &= h(\alpha[\mathsf{op}(a_1, \dots, a_n)]_E) && \text{by (1.49)} \\
&= \beta(\mathcal{T}_{\Sigma,E} h[\mathsf{op}(a_1, \dots, a_n)]_E) && \text{by (1.46)} \\
&= \beta[\mathcal{T}_\Sigma h(\mathsf{op}(a_1, \dots, a_n))]_E && \text{by (1.23)} \\
&= \beta[\mathsf{op}(h(a_1), \dots, h(a_n))]_E && \text{by (1.4)} \\
&= [\![\mathsf{op}]\!]_\beta(h(a_1), \dots, h(a_n)). && \text{by (1.49)}
\end{aligned}
$$

We obtain a functor $P^{-1} : \mathbf{EM}(\mathcal{T}_{\Sigma,E}) \to \mathbf{Alg}(\Sigma, E)$ sending (A, α) to $\mathbb{A}_\alpha$.

Finally, we need to check that P and P^{-1} are inverses to each other, i.e. that $\alpha_{\mathbb{A}_\alpha} = \alpha$ and $\mathbb{A}_{\alpha_{\mathbb{A}}} = \mathbb{A}$. For the former, $\alpha_{\mathbb{A}_\alpha}$ is defined to be the interpretation $[\![-]\!]_\alpha$ extended to terms modulo E, which we showed in Footnote 168 acts just like α. For the latter, we need to show that $[\![-]\!]_{\alpha_{\mathbb{A}}}$ and $[\![-]\!]_A$ coincide. Using Footnote 168 for the first equation and the definition of $\alpha_{\mathbb{A}}$ for the second, we have

$$[\![t]\!]_{\alpha_{\mathbb{A}}} = \alpha_{\mathbb{A}}[t]_E = [\![t]\!]_A.$$

Therefore, P and P^{-1} are inverses, thus $\mathbf{Alg}(\Sigma, E)$ and $\mathbf{EM}(\mathcal{T}_{\Sigma,E})$ are isomorphic.[169] □

[169] Observe that the functors P and P^{-1} commute with the forgetful functors because they do not change the carriers of the algebras.

Remark 1.59. This result (along with the construction of free (Σ, E)-algebras in Proposition 1.40) means that $U : \mathbf{Alg}(\Sigma, E) \to \mathbf{Set}$ is a (strictly) **monadic** functor. I decided not to define or discuss monadic functors in this document in order to have less prerequisites,[170] and because I like to exhibit the explicit isomorphism between categories of algebras. MacLane proves Proposition 1.58 using a monadicity theorem in [Mac71, §VI.8, Theorem 1].

[170] I became comfortable with monadicity theorems relatively late into my PhD, so I think avoiding them keeps things more accessible.

What about algebras for other monads? Are they algebras for some signature Σ and equations E?

Example 1.60 (Maybe)**.** In **Set**, a $(-+\mathbf{1})$-algebra is a function $\alpha : A+\mathbf{1} \to A$ making the following diagrams commute.

$$\begin{tikzcd} A \arrow[r, "\eta_A"] \arrow[dr, "\mathrm{id}_A"'] & A+\mathbf{1} \arrow[d, "\alpha"] \\ & A \end{tikzcd} \qquad \begin{tikzcd} A+\mathbf{1}+\mathbf{1} \arrow[r, "\mu_A"] \arrow[d, "\alpha+\mathbf{1}"'] & A+\mathbf{1} \arrow[d, "\alpha"] \\ A+\mathbf{1} \arrow[r, "\alpha"'] & A \end{tikzcd}$$

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 $\alpha = [\mathrm{id}_A, \alpha(*)]$.[171] The square commuting adds no 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 (1.50) is equivalent to $h(\alpha(*)) = \beta(*)$. Hence, a $(-+\mathbf{1})$-homomorphism is a function that preserves the distinguished point.

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

$$\begin{tikzcd} A+\mathbf{1} \arrow[r, "h+\mathbf{1}"] \arrow[d, "{[\mathrm{id}_A,\alpha(*)]}"'] & B+\mathbf{1} \arrow[d, "{[\mathrm{id}_B,\beta(*)]}"] \\ A \arrow[r, "h"'] & B \end{tikzcd} \tag{1.50}$$

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.[172]

[172] Notice, again, that this isomorphism would commute with the forgetful functors to **Set** because the carriers are unchanged.

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

Definition 1.61 (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 (1.51) and (1.52) commute.[173]

$$\begin{tikzcd} \mathrm{id}_{\mathbf{C}} \arrow[d, "\eta^M"'] \arrow[dr, "\eta^N"] & \\ M \arrow[r, "\rho"'] & N \end{tikzcd} \tag{1.51}$$

$$\begin{tikzcd} MM \arrow[r, "\rho\diamond\rho"] \arrow[d, "\mu^M"'] & NN \arrow[d, "\mu^N"] \\ M \arrow[r, "\rho"'] & N \end{tikzcd} \tag{1.52}$$

[173] 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.[174]

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

Example 1.62. For the signature $\Sigma = \{\mathsf{p} : 0\}$, the term monad $\mathcal{T}_\Sigma$ 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 $\rho_A : \mathcal{T}_\Sigma A \to A+\mathbf{1}$ that sends any element of A to itself and p to $* \in \mathbf{1}$. To verify that ρ is a monad morphism, we check these diagrams commute.[175]

[175] All of them commute essentially because ρ_A and both multiplications act like the identity on A.

$$\begin{tikzcd} \mathcal{T}_\Sigma A \arrow[r, "\rho_A"] \arrow[d, "\mathcal{T}_\Sigma f"'] & A+\mathbf{1} \arrow[d, "f+\mathbf{1}"] \\ \mathcal{T}_\Sigma B \arrow[r, "\rho_B"'] & B+\mathbf{1} \end{tikzcd} \tag{1.53}$$

$$\begin{tikzcd} A \arrow[d, "\eta^\Sigma_A"'] \arrow[dr, "\eta_A"] & \\ \mathcal{T}_\Sigma A \arrow[r, "\rho_A"'] & A+\mathbf{1} \end{tikzcd} \tag{1.54}$$

$$\begin{tikzcd} \mathcal{T}_\Sigma \mathcal{T}_\Sigma A \arrow[r, "\rho_{\mathcal{T}_\Sigma A}\circ(\rho_A+\mathbf{1})"] \arrow[d, "\mu^\Sigma_A"'] & A+\mathbf{1}+\mathbf{1} \arrow[d, "\mu_A"] \\ \mathcal{T}_\Sigma A \arrow[r, "\rho_A"'] & A+\mathbf{1} \end{tikzcd} \tag{1.55}$$

We obtain a monad isomorphism between the maybe monad and the term monad for the signature $\Sigma = \{\mathsf{p}:0\}$. We can recover the isomorphism between the categories of algebras from Example 1.60 with the following result.

Proposition 1.63. *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.

$$\begin{tikzcd} A \arrow[r, "\eta^M_A"] \arrow[dr, "\eta^N_A"] \arrow[ddr, "\mathrm{id}_A"'] & MA \arrow[d, "\rho_A"] \\ & NA \arrow[d, "\alpha"] \\ & A \end{tikzcd} \qquad \begin{tikzcd} MMA \arrow[rr, "\mu^M_A"] \arrow[d, "M\rho_A"'] & \text{(c)} & MA \arrow[d, "\rho_A"] \\ MNA \arrow[r, "\rho_{NA}"] \arrow[d, "M\alpha"'] & NNA \arrow[r, "\mu^N_A"] \arrow[d, "N\alpha"'] & NA \arrow[d, "\alpha"] \\ MA \arrow[r, "\rho_A"'] & NA \arrow[r, "\alpha"'] & A \end{tikzcd} \tag{1.56}$$

Showing (1.56) commutes:
(a) By (1.51).
(b) By (1.44) for $\alpha : NA \to A$.
(c) By (1.52), noting that $(\rho \diamond \rho)_A = \rho_{NA} \circ M\rho_A$.
(d) Naturality of ρ.
(e) By (1.45) 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.[176]

$$\begin{tikzcd} MA \arrow[r, "Mh"] \arrow[d, "\rho_A"'] & MB \arrow[d, "\rho_B"] \\ NA \arrow[r, "Nh"'] \arrow[d, "\alpha"'] & NB \arrow[d, "\beta"] \\ A \arrow[r, "h"'] & B \end{tikzcd}$$

[176] The top square commutes by naturality of ρ and the bottom square commutes because h is an N-homomorphism (1.46).

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 $\rho' : N \Rightarrow L$, $-(\rho' \cdot \rho) = (-\rho) \circ (-\rho')$.[177] Thus, when ρ is a monad isomorphism with inverse ρ^{-1}, $-\rho^{-1}$ is the inverse of $-\rho$, so $-\rho$ is an isomorphism. □

[177] 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 (ignoring size issues).

With the monad isomorphism $\mathcal{T}_\Sigma \cong - + \mathbf{1}$ of Example 1.62, we obtain an isomorphism $\mathbf{EM}(- + \mathbf{1}) \cong \mathbf{EM}(\mathcal{T}_\Sigma)$, and composing it with the isomorphism of Proposition 1.58 $\mathbf{EM}(\mathcal{T}_\Sigma) \cong \mathbf{Alg}(\Sigma)$ (instantiating $E = \emptyset$), we get back the result from Example 1.60 that algebras for the maybe monad are the same thing as algebras for the signature with a single constant.

In general, we now know that $\mathcal{T}_{\Sigma,E} \cong M$ implies $\mathbf{EM}(M) \cong \mathbf{Alg}(\Sigma, E)$, but constructing a monad isomorphism (and showing it is one) is not always the easiest thing to do.[178] There is a converse implication, but it requires a restriction to isomorphisms of categories that commute with the forgetful functors to $\mathbf{Set}$. Anyways, that is a mild condition we foreshadowed.

[178] For instance, the isomorphism of categories of algebras in Example 1.60 is definitely clearer than the isomorphism of monads in Example 1.62.

Proposition 1.64. *If $P : \mathbf{EM}(N) \to \mathbf{EM}(M)$ is a functor such that $U^M \circ P = U^N$, then there is a monad morphism $\rho : M \to N$. If P is an isomorphism, then so is ρ.*

Proof. Quick corollary of [BW05, Chapter 3, Theorem 6.3]. □

This motivates the following definition which states that a monad M is presented by (Σ, E) when it is isomorphic to the term monad $\mathcal{T}_{\Sigma,E}$ or, thanks to Proposition 1.64 and Proposition 1.58, when M-algebras on A and (Σ, E)-algebras on A are identified.

Definition 1.65 (**Set** presentation)**.** Let M be a monad on **Set**, an **algebraic presentation** of M is signature Σ and a class of equations E along with a monad isomorphism $\rho : \mathcal{T}_{\Sigma,E} \cong M$. We also say M is presented by (Σ, E).

We chose to state the definition with the monad isomorphism it makes some arguments in §3.4 quicker. Showing that a monad is presented by (Σ, E) can be done in many ways that are equivalent to building a monad isomorphism.[179]

We have proven in Example 1.62 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 1.66 (Powerset)**.** The powerset monad $\mathcal{P}_{\mathrm{ne}}$ is presented by the theory of **semilattices** $(\Sigma_{\mathbf{S}}, E_{\mathbf{S}})$,[180] where $\Sigma_{\mathbf{S}} = \{\oplus : 2\}$ and $E_{\mathbf{S}}$ contains the following equations stating that $\oplus$ is idempotent, commutative and associative respectively.

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

This means there is a monad isomorphism $\mathcal{T}_{\Sigma_{\mathbf{S}}, E_{\mathbf{S}}} \cong \mathcal{P}_{\mathrm{ne}}$.

Another thing we obtain from this isomorphism is that for any set X, interpreting $\oplus$ as union on $\mathcal{P}_{\mathrm{ne}}X$ (i.e. $(S, T) \mapsto S \cup T$) yields the free semilattice on X.[181]

Example 1.67 (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 $p, q \in (0, 1)$.

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

The free convex algebra on X can now be seen as $\mathcal{D}X$ with $+_p$ interpreted as the usual convex combination, that is,[182]

$$[\![\varphi +_p \psi]\!]_{\mathcal{D}X} = p\varphi + (1-p)\psi = (x \mapsto p\varphi(x) + (1-p)\psi(x)) . \tag{1.57}$$

Remark 1.68*.* Not all monads on **Set** have an algebraic presentation.[183] The monads that can be presented by a signature with finitary operation symbols are aptly called **finitary monads**. They can be characterized as the monads whose underlying functor preserve limits of a certain shape and size, see e.g. [Bor94, Proposition 4.6.2].

In Chapter 3, we will need to relate monads on different categories, we give some background on that here.

Definition 1.69 (Monad functor)**.** Let (M, η^M, μ^M) be a monad on $\mathbf{C}$, and (T, η^T, μ^T) be a monad on $\mathbf{D}$. A **monad functor** from M to T is a pair (F, λ) comprising a functor $F : \mathbf{C} \to \mathbf{D}$, and a natural transformation $\lambda : TF \Rightarrow FM$ making (1.58) and

[179] We already gave one with Proposition 1.64, and you can also read some great discussions in Remark 3.6 and §4.2 in [BSV22].

[180] Usually, when we say "theory of X", we mean that Xs are the algebras for that theory. For instance, semilattices are the $(\Sigma_{\mathbf{S}}, E_{\mathbf{S}})$-algebras. After some unrolling, we get the more common definition of a semilattice, that is, a set with a binary operation that is idempotent, commutative, and associative.

[181] It is relatively easy to show that union is idempotent, commutative, and associative, freeness is more difficult but follows from the algebraic presentation, and the fact that $(\mathcal{P}_{\mathrm{ne}}X, \mu_X)$ is the free $\mathcal{P}_{\mathrm{ne}}$-algebra (recall Example 1.57).

[182] For later, we will write $\overline{p}$ for $1 - p$.

[183] For example, the *full* powerset monad does not, although it still has an algebraic flavor as its algebras are in correspondence with complete sup-lattices, see e.g. [Bor94, Proposition 4.6.5].

(1.59) commute.[184]

$$\begin{tikzcd} F \arrow[d, "\eta^T F"'] \arrow[dr, "F\eta^M"] & \\ TF \arrow[r, "\lambda"'] & FM \end{tikzcd} \qquad (1.58) \qquad \begin{tikzcd} TTF \arrow[r, "T\lambda"] \arrow[d, "\mu^T F"'] & TFM \arrow[r, "\lambda M"] & FMM \arrow[d, "F\mu^M"] \\ TF \arrow[rr, "\lambda"'] & & FM \end{tikzcd} \qquad (1.59)$$

[184] Note the similarities with Definition 1.61, monad functors generalize monad morphisms to monads on different base categories.

Proposition 1.70. *If $(F, \lambda) : M \to T$ is a monad functor, then there is a functor $F - \circ\lambda$: $\mathbf{EM}(M) \to \mathbf{EM}(T)$ sending an M-algebra $\alpha : MA \to A$ to $F\alpha \circ \lambda_A : TFA \to A$, and an M-homomorphism $h : A \to B$ to $Fh : FA \to FB$.*[185]

[185] By definition, the functor $F - \circ\lambda$ lifts F along the forgetful functors, namely, it makes (1.60) commute.

$$\begin{tikzcd} \mathbf{EM}(M) \arrow[r, "F-\circ\lambda"] \arrow[d, "U^M"'] & \mathbf{EM}(T) \arrow[d, "U^T"] \\ \mathbf{C} \arrow[r, "F"'] & \mathbf{D} \end{tikzcd} \qquad (1.60)$$

Proof. We need to show that $F\alpha \circ \lambda$ is a T-algebra whenever α is an M-algebra. We pave the following diagrams showing (1.44) and (1.45) commute respectively.

$$\begin{tikzcd} FA \arrow[r, "\eta^T_{FA}"] \arrow[dr, "F\eta^M_A"] \arrow[ddr, "\mathrm{id}_{FA}"'] & TFA \arrow[d, "\lambda_A"] \arrow[dl, phantom, "\text{(a)}"] \\ & FMA \arrow[d, "F\alpha"] \arrow[dl, phantom, "\text{(b)}"] \\ & FA \end{tikzcd} \qquad \begin{tikzcd} TTFA \arrow[rr, "\mu^T_{FA}"] \arrow[d, "T\lambda_A"'] \arrow[drr, phantom, "\text{(c)}"] & & TFA \arrow[d, "\lambda_A"] \\ TFMA \arrow[r, "\lambda_{MA}"] \arrow[d, "TF\alpha"'] \arrow[dr, phantom, "\text{(d)}"] & FMMA \arrow[r, "F\mu^M_A"] \arrow[d, "FM\alpha"'] \arrow[dr, phantom, "\text{(e)}"] & FMA \arrow[d, "F\alpha"] \\ TFA \arrow[r, "\lambda_A"'] & FMA \arrow[r, "F\alpha"'] & FA \end{tikzcd} \qquad (1.61)$$

Showing (1.61) commutes:

(a) By (1.58).

(b) Apply F to (1.44).

(c) By (1.59).

(d) Naturality of λ.

(e) Apply F to (1.45).

Next, we need to show that when $h : A \to B$ is an M-homomorphism from α to β, then Fh is a T-homomorphism from $F\alpha \circ \lambda_A$ to $F\alpha \circ \lambda_B$. We pave the following diagram where (a) commutes by naturality of λ and (b) by applying F to (1.46).

$$\begin{tikzcd} TFA \arrow[r, "TFh"] \arrow[d, "\lambda_A"'] \arrow[dr, phantom, "\text{(a)}"] & TFB \arrow[d, "\lambda_B"] \\ FMA \arrow[r, "FMh"] \arrow[d, "F\alpha"'] \arrow[dr, phantom, "\text{(b)}"] & FMB \arrow[d, "F\beta"] \\ FA \arrow[r, "Fh"'] & FB \end{tikzcd}$$

□

There are two special cases of monad functors. When M and T are on the same category $\mathbf{C}$ and $F = \mathrm{id}_{\mathbf{C}}$, a monad functor is just a monad morphism,[186] and then the proof above reduces to the proof of Proposition 1.63. When λ_A is an identity morphism for every A, i.e. $TF = FM$, we say that M is a monad lifting of T along F. That notion is central to §3.4, where we redefine it in a more specific setting.

[186] Sometimes, authors introduce monad functors with the name monad morphism, and take our notion of monad morphism as a particular instance. Some authors also use the name monad map for either notion.

Our goal for the next two chapters is to make all the results here more general by considering carriers to be generalized metric spaces, i.e. sets with a notion of distance. In Chapter 2 we define what we mean by distance, and in Chapter 3, we define quantitative algebras, quantitative equational logic, and quantitative algebraic presentations analogously to the definitions above.

2 Generalized Metric Spaces

The Homeless Wanderer

Emahoy Tsegué-Maryam Guèbrou

For a comprehensive introduction to the concepts and themes explored in this chapter, please refer to §0.2. Here, we only give a brief overview.

In this chapter, we give our definition of generalized metric spaces which is different from the many definitions already in the literature.[187] Once again, we take our time with this material in preparation for the next chapter, introducing many examples and disseminating some insights along the way. While the content of Chapter 1 can safely be skipped before reading the current chapter, our main point here is the definition of quantitative equation (Definition 2.22) as an answer to the question "How do we impose constraints on distances with the familiar syntax of equations?", thus it makes sense to be comfortable with equational reasoning before reading what follows.

[187] See e.g. [BvBR98, Bra00, Pis21].

Outline: In §2.1, we define complete lattices and relations valued in a complete lattice, we also give an equivalent definition that justifies the syntax of quantitative equations. In §2.2, we defined quantitative equations and the categories of generalized metric spaces which are defined by collections of quantitative equations. In §2.3, we study the properties that all categories of generalized metric spaces have.

2.1 L-Spaces

Chapter 1 is titled *Universal Algebra* and Chapter 3 is titled *Universal Quantitative Algebra*. In order to go from the former to the latter, we will explain what we mean by *quantitative*. In the original paper on quantitative algebras [MPP16], and in many other works on quantitative program semantics,[188] the **quantities** considered are, more often than not, real numbers. In [MSV22, MSV23], we worked with quantities inside $[0,1]$. In this document, we will abstract away from real numbers, thinking of quantities as things you can compare and say whether one is bigger or smaller than another. You can do that with real numbers thanks to the usual ordering $\leq$, but it has a crucial property that we exploit, it is *complete* in the (informal) sense that you can always find the smallest quantity of a set of real numbers. Formally it is a complete lattice.[189]

[188] e.g. [Kwi07, vBW01, KyKK+21, ZK22].

[189] Small caveat: we need to add ∞ to the real numbers or work with an upper bound (see Example 2.3).

Definition 2.1 (Complete lattice)**.** A **complete lattice** is a partially ordered set[190] (or

[190] 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.

poset) $(\mathsf{L}, \leq)$ where all subsets $S \subseteq \mathsf{L}$ have an infimum and a supremum denoted by $\inf S$ and $\sup S$ respectively. In particular, L has a **bottom element** $\bot = \inf \mathsf{L} = \sup \emptyset$ and a **top element** $\top = \sup \mathsf{L} = \inf \emptyset$ 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, and we call its elements **quantities**.[191]

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

Example 2.2 (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 $\mathbb{R}$ 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 $\top = 1$.

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.[192]

Example 2.3 (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.[193] In our papers [MSV21, MSV22, MSV23], we worked with the unit interval, but in theory, there is no difference since $[0, 1]$ and $[0, \infty]$ are isomorphic as complete lattices.[194] In practice, one can use additional structure and properties that are not preserved by this isomorphism (like adding quantities).

Remark 2.4. The first two examples are both **quantales** [HST14, §II.1.10], informally, complete lattices where quantities can be added together in a way that preserves the order and the "smallest" quantities. It is also quite common in the literature on quantitative programming semantics to generalize from real numbers to elements of a quantale.[195] Since none of the results we establish require dealing with addition, we will work at the level of generality complete lattices (no difficulty arises from this abstraction), even though many of the following examples are quantales.

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

Example 2.5 (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 2.6 (Extended natural numbers)**.** The set $\mathbb{N}_\infty$ of natural numbers extended with ∞ is a sublattice of $[0, \infty]$.[196] Indeed, it is a poset with the usual order and

[191] That is not standard, we use this terminology only in the context of our work.

[192] 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 $[0, \infty]$.

[193] In fact, $[0, \infty]$ is also famous under the name *Lawvere quantale* because of Lawvere's seminal paper [Law02]. In that work, he used the quantale structure on $[0, \infty]$ to give a categorical definition very close to that of a metric.

[194] Take the mapping $x \mapsto \frac{1}{1-x} - 1$ from $[0, 1]$ to $[0, \infty]$ with $\frac{1}{0} - 1 = \infty$. It is monotone and preserves infimums.

[195] e.g. [DGY19, GP21, GD23, FSW+23].

[196] 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.

the infimum and supremum of a subset of natural numbers is either itself a natural number or ∞ (when the subset is empty or unbounded respectively).

Example 2.7 (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 $X \to \{\bot, \top\}$.[197] 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.

[197] 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 2.8 (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. 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 $\top$.

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 $\mathsf{L} \times \mathsf{L}$.

Example 2.9 (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

$$(\varepsilon, \delta) \leq_{\mathsf{L}\times\mathsf{K}} (\varepsilon', \delta') \iff \varepsilon \leq_\mathsf{L} \varepsilon' \text{ and } \delta \leq_\mathsf{K} \delta'. \tag{2.1}$$

It is a complete lattice where the infimums and supremums are computed coordinatewise, namely, for any $S \subseteq \mathsf{L} \times \mathsf{K}$,[198]

[198] Where π_L and π_K are the projections from $\mathsf{L} \times \mathsf{K}$ to L and K respectively.

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

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 $\top_{\mathsf{L}\times\mathsf{K}} = (\top_\mathsf{L}, \top_\mathsf{K})$.

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

Example 2.10 (CDF)**.** A **cumulative distribution function**[199] (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

$$f(\delta) = \sup\{f(\varepsilon) \mid \varepsilon < \delta\}. \tag{2.2}$$

Intuitively, (2.2) says that f cannot abruptly change value at some $x \in [0,\infty]$, but it can do that "after" some x.[200] For instance, out of the two functions below, only $f_{>1}$ is a CDF.

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

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.[201]

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 [DP02, Theorem 2.31].

- 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 (2.2).

Definition 2.11 (L-space)**.** Given a complete lattice L and a set A, an L**-relation** on A is a function $d : A \times A \to \mathsf{L}$. We call the pair (A, d) an L**-space**, and A its **carrier** or **underlying** set. We will also use a single bold-face symbol $\mathbf{A}$ to refer to an L-space with underlying set A and L-relation $d_{\mathbf{A}}$.[202]

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

$$\forall x, x' \in A, \quad d_{\mathbf{B}}(f(x), f(x')) \leq d_{\mathbf{A}}(x, x'). \tag{2.3}$$

The identity maps $\mathrm{id}_A : A \to A$ and the composition of two nonexpansive maps are always nonexpansive[203], therefore we have a category whose objects are L-spaces and morphisms are nonexpansive maps. We denote it by $\mathsf{L}\mathbf{Spa}$.

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

Remark 2.12. In the sequel, we will not distinguish between the morphism $f : \mathbf{A} \to \mathbf{B}$ and the underlying function $f : A \to B$. Although, we may write Uf for the latter, when disambiguation is necessary.

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

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

[201] 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 $\inf_{n>0} f_n = f_{>1}$.

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

[203] Fix three L-spaces $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ with two nonexpansive maps $f : A \to B$ and $g : B \to C$, we have by nonexpansiveness of g then f:

$$\begin{aligned} d_{\mathbf{C}}(gf(a), gf(a')) &\leq d_{\mathbf{B}}(f(a), f(a')) \\ &\leq d_{\mathbf{A}}(a, a'). \end{aligned}$$

Instantiating L for different complete lattices, we can get a feel for what the categories $\mathsf{L}\mathbf{Spa}$ look like. We also give concrete examples of L-spaces.

Examples 2.13 (Binary relations)**.** When $\mathsf{L} = \mathsf{B}$, a function $d : A \times A \to \mathsf{B}$ is the same thing as a subset of $A \times A$, which is the same thing as a binary relation on A.[204] Then, a B-space is a set equipped with a binary relation and we choose to have, as a convention, $d(a, a') = \bot$ when a and a' are related and $d(a, a') = \top$ when they are not.[205] A nonexpansive map from $\mathbf{A}$ to $\mathbf{B}$ is a function $f : A \to B$ such that for any $a, a' \in A$, $f(a)$ and $f(a')$ are related when a and a' are. When a and a' are not related, $f(a)$ and $f(a')$ might still be related.[206] The category $\mathsf{B}\mathbf{Spa}$ 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 chessboard (`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 $\mathsf{B}\mathbf{Spa}$. 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 $\mathsf{B}\mathbf{Spa}$. 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.[207] One can check that any rotation of the chessboard is nonexpansive from (P, d_B) to itself and from (P, d_Q) to itself. And since nonexpansive maps compose, any rotation is also nonexpansive from (P, d_B) to (P, d_Q).

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.[208] The pair (H, d_S) is an object of $\mathsf{B}\mathbf{Spa}$. 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 $\mathsf{B}\mathbf{Spa}$. 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 biological mother.

Examples 2.14 (Distances)**.** The main examples of L-spaces in this thesis are $[0,1]$-spaces or $[0,\infty]$-spaces. These are sets A equipped with a function $d : A \times A \to [0,1]$ or $d : A \times A \to [0,\infty]$, and we can usually understand $d(a, a')$ as the distance between two points $a, a' \in A$. With this interpretation, a function is nonexpansive when applying it never increases the distances between points.[209] Let us give several examples of $[0,1]$- and $[0,\infty]$-spaces:

1. **Euclidean.** Probably the most famous 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 $[0,\infty]\mathbf{Spa}$.[210] 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

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

[205] This convention might look backwards, but it makes sense with the morphisms.

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

[207] 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.

[208] Full siblings share the same biological parents.

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

[210] It is also very common to study subsets of $\mathbb{R}$, like $\mathbb{Q}$ or $[0,1]$, with the Euclidean distance appropriately restricted. We say that $(\mathbb{Q}, d)$ and $([0,1], d)$ are subspaces of $(\mathbb{R}, d)$. In general, a **subspace** of a L-space $\mathbf{A}$ is a subset $B \subseteq A$ equipped with the L-relation $d_{\mathbf{A}}$ restricted to B, i.e. $d_{\mathbf{B}} = B \times B \hookrightarrow A \times A \xrightarrow{d_{\mathbf{A}}} \mathsf{L}$.

$f : (\mathbb{R}, d) \to (\mathbb{R}, d)$ is nonexpansive when its derivative at any point is between -1 and 1.[211]

2. **Collaboration.** Let H be the set of humans again. A **collaboration chain** between two humans h and k is a sequence of scientific papers $P_1, \dots, P_n$ such that h is a coauthor of P_1, k is a coauthor of P_n and P_i and P_{i+1} always have at least one common coauthor. The collaboration distance d between two humans h and k is the length of a shortest collaboration chain.[212] For instance $d(\text{me}, \text{Paul Erdős}) = 4$ as computed by `csauthors.net` on February 20th 2024:
$$\text{me} \xrightarrow{\text{[PS21]}} \text{D. Petrişan} \xrightarrow{\text{[GPR16]}} \text{M. Gehrke} \xrightarrow{\text{[EGP07]}} \text{M. Erné} \xrightarrow{\text{[EE86]}} \text{P. Erdős}$$
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.[213] 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.)

As Example 2.14 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,[214] even when L is neither $[0, 1]$ nor $[0, \infty]$.

Examples 2.15. We give more examples of L-spaces to showcase the potential of our abstract framework.

1. **Diversion.**[215] 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 $[0, 1]\mathbf{Spa}$.[216]

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 compiled into a function $d_R : P \times P \to \mathsf{B}$ which sends (p, q) to $\bot$ if and only if $R(p) \leq R(q)$, so d_R compares the ranks of web pages. This yields a B-space (P, d_R).[217]

   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 time, we can define $d'_R(p, q)$ to be the function of type $T \to \mathsf{B}$ which sends t to the rank Boolean value of $R(p) \leq R(q)$ computed at time t. This makes (P, d'_R) into a B^T-space.

   In order to create a search engine, we also need to consider the input of the user looking for some web page.[218] If U is the set of possible user inputs, we can define $d''_R(p, q)$ to depend on U and T, so that (P, d''_R) is a $\mathsf{B}^{U \times T}$-space.

[211] The derivatives might not exist, so this is just an informal explanation.

[212] As conventions, the length of a chain is the number of papers, 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).

[213] 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 $\texttt{o} \neq \texttt{e}$).

[214] The asymmetry in the terminology "distance from x to y" is justified because, in general, nothing guarantees $d(x, y) = d(y, x)$. Since language is processed in a sequential order, we cannot even get rid of this asymmetry, but I feel like "distance *between* x and y" would be more appropriate if we required $d(x, y) = d(y, x)$.

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

[216] Even though 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.

[217] The set P equipped with the function $R : P \to [0, 1]$ is not a $[0, 1]$-space, but it is a *fuzzy set* in the sense of Castelnovo and Miculan [CM22a]. Their work shows how to reason with algebraic structures on fuzzy sets instead of L-spaces like we do here.

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

3. **Collaboration (bis).** In Example 2.14, we defined the collaboration distance $d : H \times H \to \mathbb{N}_\infty$ that measures how far two people are from collaborating on a scientific paper. We can define a finer measure by taking into account the total number of people involved in the collaboration. It allows us to say you are closer to Erdös if you wrote a paper with him and no one else than if you wrote a paper with him and two additional coauthors. The distance d' is now valued in $\mathbb{N}_\infty \times \mathbb{N}_\infty$,[219] the first coordinate of $d'(h,k)$ is $d(h,k)$ the length of the shortest collaboration chain between h and k, and the second coordinate of $d'(h,k)$ is the smallest total number of authors in a collaboration chain of length $d(h,k)$. For instance, according to csauthors.net on February 20th 2024, there are only two chains of length four between me and Erdös, both involving (the same) seven people, hence $d'(\text{me}, \text{Paul Erdös}) = (4,7)$.

4. **Bisimulation for CTS.** A conditional transition system (CTS) [ABH+12, Example 2.5] is a labelled transition system with a semantics different than the usual one. Instead of following transitions when the label matches an input, some label is chosen before the execution, and only those transitions which have the chosen label remain possible. Formulated differently, it is a family of transition systems on the same set of states indexed by a set of labels. If X is the set of states, and L is the set of labels, we can define a $\mathcal{P}(L)$-relation $d : X \times X \to \mathcal{P}(L)$ by[220]

$$d(x,y) = \{\ell \in L \mid x \text{ and } y \text{ are not bisimilar when } \ell \text{ is chosen}\}.$$

[219] There may be cases where $d'(h,k) = (4,7)$ (a long chain with few authors) and $d'(h,k') = (2,16)$ (a short chain with many authors). Then, with the product of complete lattices defined in Example 2.9, we could not compare the two distances. This is unfortunate in this application, so we may want to consider a different kind of product of complete lattices. The **lexicographical order** on $\mathbb{N}_\infty \times \mathbb{N}_\infty$ is

$$(\varepsilon,\delta) \leq_{\text{lex}} (\varepsilon',\delta') \Leftrightarrow \varepsilon \leq \varepsilon' \text{ or } (\varepsilon = \varepsilon' \text{ and } \delta \leq \delta').$$

In words, you use the order on the first coordinates, and only when they are equal, you use the order on the second coordinates.

If L and K are complete lattices, $(\mathsf{L} \times \mathsf{K}, \leq_{\text{lex}})$ is a complete lattice where the infimum is not computed pointwise, but rather

$$\inf S = (\inf \pi_{\mathsf{L}} S, \sup\{\varepsilon \mid \forall s \in S, (\inf \pi_{\mathsf{L}} S, \varepsilon) \leq s\}).$$

[220] More details in [ABH+12, §Definitions C.1 and C.2].

Here is one last example further making the case for working over an abstract complete lattice. We also revisit it in Examples 3.4 and 3.69.

Example 2.16 (Hausdorff distance)**.** Given an L-relation d on a set X, we define the L-relation $d^\uparrow$ on non-empty finite subsets of X:

$$\forall S,T \in \mathcal{P}_{\text{ne}}X, \quad d^\uparrow(S,T) = \sup\left\{\sup_{x\in S}\inf_{y\in T} d(x,y), \sup_{y\in T}\inf_{x\in S} d(x,y)\right\}.$$

This distance is a variation of a metric defined by Hausdorff in [Hau14].[221] It measures how far apart two subsets are in three steps. First, we postulate that a point $x \in S$ and T are as far apart as x and the closest point $y \in T$. Then, the distance from S to T is as big as the distance between the point $x \in S$ furthest from T. Finally, to obtain a symmetric distance, we take the maximum of the distance from S to T and from T to S. As we expect from any interesting optimization problem, there is a dual formulation given by the L-relation $d^\downarrow$.[222]

[221] Hausdorff considered positive real valued distances and compact subsets.

[222] The notation was inspired by [BBKK18]. We write $\pi_S(C)$ for $\{x \in S \mid \exists (x,y) \in C\}$ and similarly for π_T. (We should really write $\mathcal{P}_{\text{ne}}\pi_S(C)$ and $\mathcal{P}_{\text{ne}}\pi_T(C)$.)

$$\forall S,T \in \mathcal{P}_{\text{ne}}X,\ d^\downarrow(S,T) = \inf\left\{\sup_{(x,y)\in C} d(x,y) \mid C \subseteq X \times X, \pi_1(C) = S, \pi_2(C) = T\right\}$$

To compare two sets with the second method, you first need a binary relation C on X that covers all and only the points of S and T in the first and second coordinate respectively. Borrowing the terminology from probability theory, we call C a **coupling** of S and T, it is a subset of $X \times X$ whose *marginals* are S and T. According to a coupling C, the distance between S and T is the biggest distance between a pair in C. Amongst all couplings of S and T, we take the one achieving the smallest distance to define $d^{\downarrow}(S,T)$.

The first punchline of this example is that the two L-relations $d^{\uparrow}$ and $d^{\downarrow}$ coincide.

Lemma 2.17. *For any* $S,T \in \mathcal{P}_{\mathrm{ne}}X$, $d^{\uparrow}(S,T) = d^{\downarrow}(S,T)$.[223]

[223] Hardly adapted from [Mé11, Proposition 2.1].

Proof. ($\leq$) For any coupling $C \subseteq X \times X$, for each $x \in S$, there is at least one $y_x \in T$ such that $(x,y_x) \in C$ (because $\pi_1(C) = S$) so

$$\sup_{x\in S}\inf_{y\in T} d(x,y) \leq \sup_{x\in S} d(x,y_x) \leq \sup_{(x,y)\in C} d(x,y).$$

After a symmetric argument, we find that $d^{\uparrow}(S,T) \leq \sup_{(x,y)\in C} d(x,y)$ for all couplings, the first inequality follows.

($\geq$) For any $x \in S$, let $y_x \in T$ be a point in T that attains the infimum of $d(x,y)$,[224] and note that our definition ensures $d(x,y_x) \leq d^{\uparrow}(S,T)$. Symmetrically define x_y for any $y \in T$ and let $C = \{(x,y_x) \mid x \in S\} \cup \{(x_y,y) \mid y \in T\}$. It is clear that C is a coupling of S and T, and by our choices of y_x and x_y, we ensured that

[224] It exists because T is non-empty and finite.

$$\sup_{(x,y)\in C} d(x,y) \leq d^{\uparrow}(S,T),$$

therefore we found a coupling witnessing that $d^{\downarrow}(S,T) \leq d^{\uparrow}(S,T)$ as desired. □

The second punchline of this example comes from instantiating it with the complete lattice B. Recall that a B-relation d on X corresponds to a binary relation $R_d \subseteq X \times X$ where x and y are related if and only if $d(x,y) = \bot$. This seemingly backwards convention makes it so that nonexpansive functions are those that preserve the relation. Let us be careful about it while describing $R_{d^{\uparrow}}$ and $R_{d^{\downarrow}}$.

Given $S,T \in \mathcal{P}_{\mathrm{ne}}X$ and $x \in S$, notice that $\inf_{y\in T} d(x,y) = \bot$ if and only if $d(x,y) = \bot$ for at least one y, or equivalently, if x is related by R_d to at least one $y \in T$. This means the infimum behaves like an existential quantifier. Dually, the supremum acts like a universal quantifier yielding[225]

[225] Symmetrically,
$$\sup_{y\in T}\inf_{x\in S} d(x,y) = \bot \Leftrightarrow \forall y \in T, \exists x \in S, (x,y) \in R_d.$$

$$\sup_{x\in S}\inf_{y\in T} d(x,y) = \bot \iff \forall x \in S, \exists y \in T, (x,y) \in R_d.$$

Combining with its symmetric counterpart, and noting that a binary universal quantification is just an `AND`, we find that (S,T) belongs to $R_{d^{\uparrow}}$ if and only if

$$\forall x \in S, \exists y \in T, (x,y) \in R_d \text{ and } \forall y \in T, \exists x \in S, (x,y) \in R_d. \tag{2.4}$$

We call $R_{d^{\uparrow}}$ the Egli–Milner extension of R_d as in, e.g., [WS20, GPA21].

Given a coupling C of S and T, $\sup_{(x,y)\in C} d(x,y)$ can only equal $\bot$ when all pairs $(x,y) \in C$ are related by R_d. Then, if a coupling $C \subseteq R_d$ exists, the infimum of $d^{\downarrow}$ will be $\bot$. Therefore, S and T are related by $R_{d^{\downarrow}}$ if and only if

$$\exists C \subseteq R_d, \pi_S(C) = S \text{ and } \pi_T(C) = T. \tag{2.5}$$

The relation $R_{d\downarrow}$ is sometimes called the Barr lifting of R_d [Bar06].

Our proof above yields the equivalence between (2.4) and (2.5).[226]

[226] That equivalence is folklore and has probably been given as exercise to many students in a class on bisimulation or coalgebras.

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. The syntax is heavily inspired by the syntax of equations in universal algebra, the binary predicate $=$ for equality is joined by a family of binary predicates $=_\varepsilon$ indexed by the quantities in L. That idea comes from the original work of Mardare, Panangaden, and Plotkin on quantitative algebras [MPP16], and it implicitly relies on the following equivalent definition of L-spaces (the equivalent definition is not due to Mardare et al., see the discussion in §0.3).

Definition 2.18 (L-structure)**.** Given a complete lattice L, an L**-structure**[227] is a set X equipped with a family of binary relations $R_\varepsilon \subseteq X \times X$ indexed by $\varepsilon \in \mathsf{L}$ satisfying

- **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 $\varepsilon_i \in \mathsf{L}$,[228]

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

[227] We borrow the name "structure" from model theorists. The more general notion of relational structure is used in [FMS21, Par22, Par23]. Also, our L-structures are both more and less general than the $\mathcal{L}_S$-structures of [Con17].

[228] By monotonicity, $R_\delta \subseteq R_{\varepsilon_i}$ so the inclusion $R_\delta \subseteq \cap_{i\in I} R_{\varepsilon_i}$ always holds. Also, continuity implies monotonicity because $\varepsilon \leq \varepsilon'$ implies

$$R_\varepsilon \cap R_{\varepsilon'} = R_{\inf\{\varepsilon,\varepsilon'\}} = R_\varepsilon,$$

which means $R_\varepsilon \subseteq R_{\varepsilon'}$. Still, we keep monotonicity explicit for better exposition.

Intuitively $(x,y) \in R_\varepsilon$ 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 $\inf_{i\in I} \varepsilon_i$.

The names for these conditions come from yet another equivalent definition.[229] Organizing 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 2.7. 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.[230]

[229] This time more directly equivalent.

[230] 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

$$\forall \varepsilon \in \mathsf{L}, \forall x, x' \in X, (x,x') \in R_\varepsilon \implies (f(x), f(x')) \in S_\varepsilon. \tag{2.6}$$

This should feel similar to nonexpansive maps.[231] Let us call $\mathsf{L}\mathbf{Str}$ the category of L-structures.

[231] In words, (2.6) reads as: if x and x' are at a distance below ε' then so are $f(x)$ and $f(x')$.

We give one trivial example, before proving that L-structures are just L-spaces.

Example 2.19. A consequence of continuity (take $I = \emptyset$) is that $R_\top$ is the full binary relation $X \times X$. Therefore, taking $\mathsf{L} = 1$ to be a singleton where $\bot = \top$, a 1-structure is only a set (there is no choice for R), and a morphism is only a function (the implication in (2.6) is always true because $S_\varepsilon = Y \times Y$). In other words, $1\mathbf{Str}$ is isomorphic to **Set**. Instantiating the next result (Proposition 2.20) means that $1\mathbf{Spa}$ is also isomorphic to **Set**, this is clear because there is only one function $d : X \times X \to 1$ for any set X. This example is relatively important because it means the theory we develop later over an arbitrary category of L-spaces specializes to the case of **Set**.[232]

[232] See Example 3.56.

Proposition 2.20. *For any complete lattice* L*, the categories* $\mathsf{L}\mathbf{Spa}$ *and* $\mathsf{L}\mathbf{Str}$ *are isomorphic.*[233]

[233] This result is a stripped down version of [MPP17, Theorem 4.3]. A more general version also appears in [FMS21, Example 3.5.(4)]. Another similar result is shown in [Par22, Appendix]. The core idea, (2.7) and (2.8), also appears in [Con17, Theorem A].

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

$$(x, x') \in R^d_\varepsilon \iff d(x, x') \leq \varepsilon. \tag{2.7}$$

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

$$(x, x') \in R^d_\varepsilon \overset{(2.7)}{\iff} d(x, x') \leq \varepsilon \implies d(x, x') \leq \varepsilon' \overset{(2.7)}{\iff} (x, x') \in R^d_{\varepsilon'}.$$

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 definition 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.

Any nonexpansive map $f : (X, d) \to (Y, \Delta)$ in $\mathsf{L}\mathbf{Spa}$ 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 $x, x' \in X$,

$$(x, x') \in R^d_\varepsilon \overset{(2.7)}{\iff} d(x, x') \leq \varepsilon \overset{(2.3)}{\implies} \Delta(f(x), f(x')) \leq \varepsilon \overset{(2.7)}{\iff} (f(x), f(x')) \in R^\Delta_\varepsilon.$$

Taking L = B, Proposition 2.20 gives back our interpretation of $\mathsf{B}\mathbf{Spa}$ as the category $2\mathbf{Rel}$ from Example 2.13. Indeed, a B-structure is just a set X equipped with a binary relation $R_\bot \subseteq X \times X$ (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 $d(x, y) = \bot$ meaning x and y are related.

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[234]

$$d_R(x, x') \leq \varepsilon \iff (x, x') \in R_\varepsilon. \tag{2.8}$$

[234] 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

$$d_R(x, x') \leq \varepsilon \overset{(2.8)}{\iff} (x, x') \in R_\varepsilon \overset{(2.6)}{\implies} (f(x), f(x')) \in S_\varepsilon \overset{(2.8)}{\iff} d_S(f(x), f(x')) \leq \varepsilon,$$

hence putting $\varepsilon = d_R(x, x')$, 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 (2.7) and (2.8) 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 $\mathsf{L}\mathbf{Spa} \cong \mathsf{L}\mathbf{Str}$. $\square$

This result is central in our treatment of L-spaces because it allows us to specify an L-relation through the (binary) truth value of a family of predicates $=_\varepsilon$. In other words, we can reason equationally about L-spaces.

Remark 2.21. The upshot of Proposition 2.20 is that the structure of a complete lattice is enough to do quantitative algebraic reasoning.[235] Still, in practice, L often has more structure. If you need to state the triangular inequality (2.12), then you need a way of adding distances/quantities. A frequent choice made by researchers is to let L be a quantale see e.g., [CH06, Pis21]. Often, this is for the theoretical convenience of seeing a metric space as an enriched category as suggested in [Law02].[236] In closely related work [CM22a], Castelnovo and Miculan require L to be a frame (or complete lattice with some distributivity properties).

[235] This point will be strenghtened when we develop the theory of quantitative algebras over an arbitary complete lattice in Chapter 3.

[236] The book [HST14] explores the theoretical foundations of this approach.

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 to enforce on distance functions. For example, we can restate (with a slight modification[237]) the axioms defining metric spaces (Definition 0.1).

[237] The separation axiom is now divided in two, (2.10) and (2.11).

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

$$\forall x, y \in X, \quad d(x,y) = d(y,x). \tag{2.9}$$

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

$$\forall x \in X, \quad d(x,x) = 0. \tag{2.10}$$

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:

$$\forall x, y \in X, \quad d(x,y) = 0 \implies x = y. \tag{2.11}$$

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:

$$\forall x, y, z \in X, \quad d(x,z) \leq d(x,y) + d(y,z). \tag{2.12}$$

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. They are abstraction of properties of the "smaller or equal" order $\leq$ on, say, real numbers.

First, reflexivity says that any element x is related to itself, i.e. $x \leq x$. Translating back to the B-relation, this is equivalent to:

$$\forall x \in X, \quad d(x,x) = \bot. \tag{2.13}$$

Antisymmetry says that if both (x, y) and (y, x) are in the order relation, then they must be equal:[238]

[238] i.e., if $x \leq y$ and $y \leq x$, then $x = y$.

$$\forall x, y \in X, \quad d(x, y) = \bot = d(y, x) \implies x = y. \tag{2.14}$$

Finally, transitivity says that if (x, y) and (y, z) belong to the order relation, then so does (x, z):[239]

[239] i.e. if $x \leq y$ and $y \leq z$, then $x \leq z$.

$$\forall x, y, z \in X, \quad d(x, y) = \bot = d(y, z) \implies d(x, z) = \bot. \tag{2.15}$$

We can immediately notice that all the axioms (2.9)–(2.15) start with a universal quantification of variables. Another thing to note is that we never actually needed to talk about equality between distances. For instance, the equation $d(x, y) = d(y, x)$ in the axiom of symmetry (2.9) can be replaced by two inequalities $d(x, y) \leq d(y, x)$ and $d(y, x) \leq d(x, y)$, and moreover since x and y are universally quantified, only one of these inequalities is necessary:

$$\forall x, y \in X, \quad d(x, y) \leq d(y, x). \tag{2.16}$$

If we rely on the equivalence between L-spaces and L-structures (Proposition 2.20), we can transform (2.16) into a family of implications indexed by all $\varepsilon \in \mathsf{L}$:[240]

[240] Recall that $(x, y) \in R^d_\varepsilon$ is the same thing as $d(x, y) \leq \varepsilon$. Hence, (2.16) and (2.17) 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 $d(x, y)$.

$$\forall x, y \in X, \quad (y, x) \in R^d_\varepsilon \implies (x, y) \in R^d_\varepsilon. \tag{2.17}$$

Starting from the triangle inequality (2.12) and applying the same transformations that got us from (2.9) to (2.17), we obtain a family of implications indexed by two values $\varepsilon, \delta \in \mathsf{L}$:[241]

[241] You can try proving how (2.12) and (2.18) are equivalent if the process of going from the former to the latter was not clear to you.

$$\forall x, y, z \in X, \quad (x, y) \in R^d_\varepsilon \text{ and } (y, z) \in R^d_\delta \implies (x, z) \in R^d_{\varepsilon+\delta}. \tag{2.18}$$

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 you can keep in mind the examples $\mathsf{L} = [0, 1]$ and $\mathsf{L} = [0, \infty]$.

Definition 2.22 (Quantitative equation).[242] 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 a quantity $\varepsilon \in \mathsf{L}$. 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.

[242] The name quantitative equation will be reclaimed in Definition 3.6 for a more general notion. See also Remark 3.7.

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

- $\mathbf{X} \vdash x = y$ if for any nonexpansive assignment $\hat{\imath} : \mathbf{X} \to \mathbf{A}$, $\hat{\imath}(x) = \hat{\imath}(y)$.
- $\mathbf{X} \vdash x =_\varepsilon y$ if for any nonexpansive assignment $\hat{\imath} : \mathbf{X} \to \mathbf{A}$, $d_{\mathbf{A}}(\hat{\imath}(x), \hat{\imath}(y)) \leq \varepsilon$.[243]

[243] Viewing it in the L-structure $(A, \{R^{d_{\mathbf{A}}}_\varepsilon\})$, we want that $\hat{\imath}(x) \; R^{d_{\mathbf{A}}}_\varepsilon \; \hat{\imath}(y)$ which looks a lot like $x =_\varepsilon y$.

We use ϕ and ψ to refer to a quantitative equation, and we sometimes call them simply equations. We write $\mathbf{A} \vDash \phi$ when $\mathbf{A}$ satisfies ϕ,[244] and we also write $\mathbf{A} \vDash^{\hat{\imath}} \phi$ when the equality $\hat{\imath}(x) = \hat{\imath}(y)$ or the bound $d_{\mathbf{A}}(\hat{\imath}(x), \hat{\imath}(y)) \leq \varepsilon$ holds for a particular assignment $\hat{\imath} : \mathbf{X} \to \mathbf{A}$ (and not necessarily for all assignments).

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

Let us illustrate this definition with an example.

Example 2.23 (Symmetry)**.** We want to translate (2.17) 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 mentioned 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 $\hat{\imath} : \mathbf{X} \to \mathbf{A}$. 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 $\mathbf{X}$ be the L-space with two elements x and y such that $d_{\mathbf{X}}(y, x) = \varepsilon$ and all other distances are $\top$. A nonexpansive assignment $\hat{\imath} : \mathbf{X} \to \mathbf{A}$ is just a choice of two elements $\hat{\imath}(x), \hat{\imath}(y) \in A$ satisfying $d_{\mathbf{A}}(\hat{\imath}(y), \hat{\imath}(x)) \leq \varepsilon$.[245] For all of these, we have to impose the condition $d_{\mathbf{A}}(\hat{\imath}(x), \hat{\imath}(y)) \leq \varepsilon$. Therefore, our quantitative equation is

$$\mathbf{X} \vdash x =_\varepsilon y. \tag{2.19}$$

[245] 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 (2.19) if and only if it satisfies (2.17). Hence,[246] if $\mathbf{A}$ satisfies that quantitative equation for all $\varepsilon \in \mathsf{L}$, then it satisfies (2.9), i.e. the distance $d_{\mathbf{A}}$ is symmetric.

[246] Recall our argument in Footnote 240.

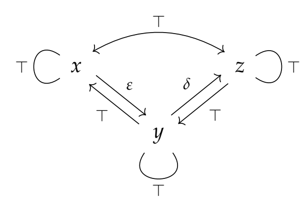
In practice, defining the context like this is more cumbersome than need be, so we will define some syntactic sugar to remedy this. Before that, we take the time to do another example.

Example 2.24 (Triangle inequality)**.** With $\mathsf{L} = [0, 1]$ or $\mathsf{L} = [0, \infty]$, let the context $\mathbf{X}$ be the L-space with three elements x, y and z such that $d_{\mathbf{X}}(x, y) = \varepsilon$ and $d_{\mathbf{X}}(y, z) = \delta$, and all other distances are $\top$. [247] A nonexpansive assignment $\hat{\imath} : \mathbf{X} \to \mathbf{A}$ is just a choice of three elements $a = \hat{\imath}(x), b = \hat{\imath}(y), c = \hat{\imath}(z) \in A$ such that $d_{\mathbf{A}}(a, b) \leq \varepsilon$ and $d_{\mathbf{A}}(b, c) \leq \delta$. Hence, if $\mathbf{A}$ satisfies

$$\mathbf{X} \vdash x =_{\varepsilon+\delta} z, \tag{2.20}$$

it means that for any such assignment, $d_{\mathbf{A}}(a, c) \leq \varepsilon + \delta$ also holds. We conclude that $\mathbf{A}$ satisfies (2.18). If $\mathbf{A}$ satisfies $\mathbf{X} \vdash x =_{\varepsilon+\delta} z$ for all $\varepsilon, \delta \in \mathsf{L}$, then $\mathbf{A}$ satisfies the triangle inequality (2.12).

[247] Here is a depiction of $\mathbf{X}$, where the label on an arrow is the distance from the source to the target of that arrow:



Remark 2.25. There is a small caveat above. If we are in $\mathsf{L} = [0, 1]$ and $\varepsilon = 1$ and $\delta = 1$, then $\varepsilon + \delta = 2 \notin [0, 1]$, so the predicate $x =_{\varepsilon+\delta} z$ is not allowed. There are two easy fixes that we never explicit. You can either define a truncated addition so that $\varepsilon + \delta = 1$ whenever their sum is really above 1, or you can quantify over ε and δ such that $\varepsilon + \delta \leq 1$. Indeed, every $[0, 1]$-space satisfies $\mathbf{X} \vdash x =_1 z$ because 1 is a global upper bound for the distance between points, thus when $\varepsilon + \delta > 1$, there is no difference between having that equation or not as an axiom.

Notice that in the contexts of Examples 2.23 and 2.24, 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, (2.19) will be written[248]

$$y =_\varepsilon x \vdash x =_\varepsilon y, \tag{2.21}$$

and (2.20) will be written

$$x =_\varepsilon y, y =_\delta z \vdash x =_{\varepsilon+\delta} z. \tag{2.22}$$

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[249]

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

and the L-relation is defined for $u, v \in X$ by[250]

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

Remark 2.26. The judgments (or quantitative inferences) in the logic of [MPP16] with an empty signature coincide with our syntactic sugar. We showed those are a formally equivalent to quantitative equationsin [MSV23, Lemma 8.4], but there is a special case we want to discuss.

In [MPP16, Definition 2.1], their axiom (Arch) is equivalent, in the presence of their axiom (Max), to

$$\{x =_{\varepsilon_i} y \mid i \in I\} \vdash x =_{\inf_{i\in I} \varepsilon_i} y.$$

Now, if we apply our translation to obtain a quantitative equation as in Definition 2.22, we get $\mathbf{X} \vdash x =_\varepsilon y$, where $d_{\mathbf{X}}(x, y) = \varepsilon = \inf_{i\in I} \varepsilon_i$ and all other distances are $\top$. This quantitative equation is obviously always satisfied,[251] so it makes sense to have it as an axiom, but it seems we are loosing a bit of information. That is, the original axiom looks like it ensures the continuity property of Definition 2.18. In fact, that axiom has several names in different papers, one of which is CONT. In the version of quantitative equational logic we propose in this thesis (Figure 3.1), there is an inference rule CONT (rather than an axiom) that ensures continuity.

Here are some more translations of famous properties into quantitative equations written with the syntactic sugar:

- reflexivity (of a metric) (2.10) becomes $x \vdash x =_0 x$,[252]
- Leibniz's law (2.11) becomes $x =_0 y \vdash x = y$,
- reflexivity (of an order) (2.13) becomes $x \vdash x =_\perp x$,

[248] We can understand this syntax as putting back the information in the context into an implication. For instance, you can read (2.21) as "**if** the distance from y to x is bounded above by ε, **then** so is the distance from x to y". You can read (2.22) as "if the distance from x to y is bounded above by ε **and** the distance from y to z is bounded above by δ, then the distance from x to z is bounded above by $\varepsilon + \delta$".

[249] 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.

[250] In words, the distance from u to v is the smallest value ε such that $u =_\varepsilon v$ was a premise. If no such premise occurs, the distance from u to v is $\top$. It is rare that u and v appear several times together (because $u =_\varepsilon v$ and $u =_\delta v$ can be replaced with $u =_{\inf\{\varepsilon,\delta\}} v$), but our definition allows it.

[251] For any nonexpansive assignment $\hat{\imath} : \mathbf{X} \to \mathbf{A}$, $d_{\mathbf{A}}(\hat{\imath}(x), \hat{\imath}(y)) \leq d_{\mathbf{X}}(x, y) = \varepsilon$.

[252] As further sugar, we also write x instead of $x =_\top x$ to the left of the turnstile $\vdash$ to say that the variable x is in the context without imposing any constraint. For instance, the context of $x, y \vdash x = y$ has two variables x and y and all distances are $\top$. Thus, if $\mathbf{A}$ satisfies $x, y \vdash x = y$, then $\mathbf{A}$ is either empty or a singleton.

- antisymmetry (2.14) becomes $x =_\perp y, y =_\perp x \vdash x = y$, and
- transitivity (2.15) becomes $x =_\perp y, y =_\perp z \vdash x =_\perp z$.

Remark 2.27. The translations of (2.10) and (2.13) look very close. In fact, noting that 0 is the bottom element of $[0,1]$ and $[0,\infty]$, the quantitative equation $x \vdash x =_\perp x$ 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 (2.22), if we let ε and δ range over B and interpret $+$ as an `OR`, we get three vacuous quantitative equations[253] and the translation of (2.15) above. So transitivity and triangle inequality are the same under this abstract point of view.[254]

Let us emphasize one thing about contexts of quantitative equations: they only give constraints that are upper bounds for distances.[255] In particular, it can be very hard to operate on the quantities in L non-monotonically. For instance, we will see (after Definition 2.38) that we cannot read $x =_{\varepsilon_1} y, y =_{\varepsilon_2} z, y =_{\varepsilon_3} y \vdash x =_{\varepsilon_1+\varepsilon_2-\varepsilon_3} z$ as saying that $d(x,z) \leq d(x,y) + d(y,z) - d(y,y)$, and one quick explanation is that subtraction is not a monotone operation on $[0,\infty] \times [0,\infty]$.[256] Another consequence is that an equation ϕ will always entail ψ when the latter has a *stricter* context (i.e. when the upper-bounds in the premises are smaller).[257] We prove a more general version of this below.

[253] 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.

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

[255] Well, if you consider the opposite order on L, they now give lower bounds. What is important is that they only speak about one of them.

[256] Assume $\mathsf{L} = [0,\infty]$ and $d(y,y)$ may be non-zero.

[257] For example, if $\mathbf{A}$ satisfies $x =_{1/2} y \vdash x = y$, then it satisfies $x =_{1/3} y \vdash x = y$. This says that if all distances between distinct points are above $1/2$, then they are also above $1/3$.

Lemma 2.28. *Let $f : \mathbf{X} \to \mathbf{Y}$ be a nonexpansive map. If $\mathbf{A}$ satisfies $\mathbf{X} \vdash x = y$ (resp. $\mathbf{X} \vdash x =_\varepsilon y$), then $\mathbf{A}$ satisfies $\mathbf{Y} \vdash f(x) = f(y)$ (resp. $\mathbf{Y} \vdash f(x) =_\varepsilon f(y)$).*

Proof. Any nonexpansive assignment $\hat{\imath} : \mathbf{Y} \to \mathbf{A}$ yields a nonexpansive assignment $\hat{\imath} \circ f : \mathbf{X} \to \mathbf{A}$. By hypothesis, we have

$$\mathbf{A} \vDash^{\hat{\imath}\circ f} \mathbf{X} \vdash x = y \qquad (\text{resp. } \mathbf{A} \vDash^{\hat{\imath}\circ f} \mathbf{X} \vdash x =_\varepsilon y),$$

which means $\hat{\imath}(f(x)) = \hat{\imath}(f(y))$ (resp. $d_{\mathbf{A}}(\hat{\imath}(f(x)), \hat{\imath}(f(y))) \leq \varepsilon$). Thus, we conclude

$$\mathbf{A} \vDash^{\hat{\imath}} \mathbf{Y} \vdash f(x) = f(y) \qquad (\text{resp. } \mathbf{A} \vDash^{\hat{\imath}} \mathbf{Y} \vdash f(x) =_\varepsilon f(y)). \qquad \square$$

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 2.29. For any complete lattice L.

1. The **strong triangle inequality** states that $d(x,z) \leq \max\{d(x,y), d(y,z)\}$,[258] it is equivalent to the satisfaction of the following family of quantitative equations

$$\forall \varepsilon, \delta \in \mathsf{L}, \quad x =_\varepsilon y, y =_\delta z \vdash x =_{\sup\{\varepsilon,\delta\}} z. \tag{2.23}$$

2. We can impose that all distances are below a **global upper bound** $\varepsilon \in \mathsf{L}$ (i.e. $d(x,y) \leq \varepsilon$) with the quantitative equation[259]

$$x, y \vdash x =_\varepsilon y. \tag{2.24}$$

[258] This property is used in defining ultrametrics [Rut96].

[259] For instance $[0,1]$-spaces are $[0,\infty]$-spaces that satisfy $x, y \vdash x =_1 y$.

3. We can *almost* impose a **global lower bound** $\varepsilon \in \mathsf{L}$ on distances. What we can do instead is impose a strict lower bound on distances that are not self-distances (i.e. $\forall x \neq y, d(x,y) > \varepsilon$).[260] To achieve this with an equation, we ensure the equivalent property that whenever $d(x,y)$ is smaller than ε, then $x = y$:

$$x =_\varepsilon y \vdash x = y. \tag{2.25}$$

[260] We can also do a non-strict lower bound (i.e. $\forall x \neq y, d(x,y) \geq \varepsilon$) by considering the family of equations $x =_\delta y \vdash x = y$ for all $\delta < \varepsilon$.

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

1. Given a positive number $b > 0$, the ***b*-triangle inequality** states that $d(x,z) \leq b(d(x,y) + d(y,z))$,[261] it is equivalent to the satisfaction of

$$\forall \varepsilon, \delta \in \mathsf{L}, \quad x =_\varepsilon y, y =_\delta z \vdash x =_{b(\varepsilon+\delta)} z. \tag{2.26}$$

[261] This property is used in defining b-metrics [KP22, Definition 1.1].

2. The **rectangle inequality** states that $d(x,w) \leq d(x,y) + d(y,z) + d(z,w)$,[262] it is equivalent to the satisfaction of

$$\forall \varepsilon_1, \varepsilon_2 \in \mathsf{L}, \quad x =_{\varepsilon_1} y, y =_{\varepsilon_2} z, z =_{\varepsilon_3} w \vdash x =_{\varepsilon_1+\varepsilon_2+\varepsilon_3} w. \tag{2.27}$$

[262] This property is used in defining g.m.s. in [Bra00, Definition 1.1].

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

$$x =_\perp y, x =_\perp y' \vdash y = y'. \tag{2.28}$$

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 $y \in X$.[263] This is equivalent to satisfying

$$x =_\perp y, x' =_\perp y \vdash x = x'. \tag{2.29}$$

[263] Equivalently, the opposite (or converse) of R is functional. You may want to formulate totality or surjectivity of a binary relation with quantitative equations, but you will find that difficult. We show in Example 2.45 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 (2.15)). This is equivalent to satisfying

$$x =_\perp y, y =_\perp z \vdash z =_\perp x. \tag{2.30}$$

We now turn to the study of subcategories of L**Spa** that are defined via (sets of) quantitative equations. Given a class $\hat{E}$ of quantitative equations, we can define a full subcategory of L**Spa** that contains only those L-spaces that satisfy $\hat{E}$, this is the category $\mathbf{GMet}(\mathsf{L}, \hat{E})$ whose objects we call generalized metric spaces or spaces for short. We also write $\mathbf{GMet}(\hat{E})$ or **GMet** when the complete lattices L or the class $\hat{E}$ are fixed or irrelevant. There is an evident forgetful functor $U : \mathbf{GMet} \to \mathbf{Set}$ which is the composition of the inclusion functor $\mathbf{GMet} \to \mathsf{L}\mathbf{Spa}$ and $U : \mathsf{L}\mathbf{Spa} \to \mathbf{Set}$.[264]

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

The terminology generalized metric space appears quite a lot in the literature with different meanings (e.g. [BvBR98, Bra00, Pis21]), so I expect many will navigate to this definition before reading what is above. Catering to these readers, let us redefine what we mean by generalized metric space.

Definition 2.30 (Generalized metric space)**.** A **generalized metric space** or **space** is a set X along with a function $d : X \times X \to \mathsf{L}$ into a complete lattice L such that (X, d) satisfies some constraints expressed by a fixed collection of quantitative equations.

When $\mathsf{L} = [0, \infty]$, examples include metrics [Fré06], ultrametrics [Rut96], pseudometrics, quasimetrics [Wil31a], semimetrics [Wil31b], b-metrics [KP22], the generalized metric spaces of [Bra00], dislocated metrics [HS00] also called diffuse metrics in [CKPR21], the generalized metric spaces of [BvBR98] which are the metric spaces of [Law02], etc.

When $\mathsf{L} = \mathsf{B}$ (the Boolean lattice), examples include posets, preorders, equivalence relations, partial (or restricted) equivalence relations [Sco76], graphs, etc.

The most notable examples of generalized metric spaces are posets and metric spaces, they form the categories **Poset** and **Met**.

Example 2.31 (**Poset**)**.** The category of partially ordered sets and monotone maps is the full subcategory of $\mathsf{B}\mathbf{Spa}$ with all B-spaces satisfying reflexivity, antisymmetry, and transitivity stated as quantitative equations:[265]

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

[265] Examples of posets include any set of numbers (e.g. $\mathbb{N}$, $\mathbb{Q}$, $\mathbb{R}$) equipped with the usual (non-strict) order $\leq$, and $\mathcal{P}_{\mathrm{ne}}X$ with the inclusion order.

In practice, it would be useful to replace the symbol for $=_\bot$ with $\leq$ so the axioms become the more familiar

$$\hat{E}_{\mathbf{Poset}} = \{x \vdash x \leq x, x \leq y, y \leq x \vdash x = y, x \leq y, y \leq z \vdash x \leq z\}.$$

Example 2.32 (**Met**)**.** The category of **metric spaces** and nonexpansive maps is the full subcategory of $[0,1]\mathbf{Spa}$ (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:[266] $\hat{E}_{\mathbf{Met}}$ contains all the following

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

[266] Examples of metric spaces include $[0,1]$ with the Euclidean distance from Example 2.14, the Kantorovich distance from Example 3.5, and the total variation distance from Example 3.78.

2.3 The Categories GMet

In this section, we prove some basic results about the categories of generalized metric spaces. We fix a complete lattice L and a class of quantitative equations $\hat{E}$ throughout, and denote by **GMet** the category of L-spaces that satisfy $\hat{E}$. The goal here is mainly to become familiar with L-spaces and quantitative equations, so not everything will be useful later. This also means we will avoid using abstract results (that we prove later) which can (sometimes drastically) simplify some proofs.[267]

We also take some time to identify some (well-known) conditions on L-spaces that cannot be expressed via quantitative equations.[268] These proofs are always in the same vein, we know **GMet** has some property, we show the class of L-spaces with a

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

[268] Again, we cannot make an exhaustive list.

condition does not have that property, hence that condition is not expressible as a class of quantitative equations.

In order to keep all the information about **GMet** in the same place, we will quickly summarize at the end the things we know about these categories (including things that will come from results in Chapter 3).

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 2.33. *The category* **GMet** *has a terminal object.*

Proof. The terminal object $\mathbf{1}$ in L**Spa** is relatively easy to find,[269] 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 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[270], ! is the unique morphism of this type.

Since **GMet** is a full subcategory of L**Spa**, it is enough to show $\mathbf{1}$ is in **GMet** to conclude it is the terminal object in this subcategory. We can do this by showing $\mathbf{1}$ satisfies absolutely all quantitative equations, and in particular those of $\hat{E}$.[271] 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 $\hat{\imath} : \mathbf{X} \to \mathbf{1}$, and it sends x and y to $*$. This means

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

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

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

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

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

Proposition 2.34. *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:[272]

$$\forall a, b \in A, \quad d(a, b) = \sup_{i \in I} d_i(a_i, b_i). \tag{2.31}$$

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 $\langle f_i \rangle : X \to A$

[272] For $a \in A$, let a_i be the ith coordinate of 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[273]

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

Note that a particular case of this construction for I being empty is the terminal object $\mathbf{1}$ from Proposition 2.33. Indeed, the empty Cartesian product is the singleton, and the empty supremum is the bottom element $\bot$. $\square$

In order to show that satisfaction of a quantitative equation is preserved by the product of L-spaces, we first prove a simple lemma.[274]

Lemma 2.35. *Let ϕ be a quantitative equation with context $\mathbf{X}$. If $f : \mathbf{A} \to \mathbf{B}$ is a nonexpansive map and $\mathbf{A} \vDash^{\hat{\imath}} \phi$ for a nonexpansive assignment $\hat{\imath} : \mathbf{X} \to \mathbf{A}$, then $\mathbf{B} \vDash^{f \circ \hat{\imath}} \phi$.*

Proof. There are two very similar cases. If ϕ is of the form $\mathbf{X} \vdash x = y$, we have[275]

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

If ϕ is of the form $\mathbf{X} \vdash x =_\varepsilon y$, we have[276]

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

Proposition 2.36. *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 $\hat{\imath} : \mathbf{X} \to \mathbf{A}$, the following equivalence holds:[277]

$$\left(\forall i \in I, \mathbf{A}_i \vDash^{\pi_i \circ \hat{\imath}} \phi \right) \iff \mathbf{A} \vDash^{\hat{\imath}} \phi. \tag{2.32}$$

The proposition follows because if $\mathbf{A}_i \vDash \phi$ for all $i \in I$, then the `L.H.S.` holds for any $\hat{\imath}$, hence the `R.H.S.` does too, and we conclude $\mathbf{A} \vDash \phi$. Let us prove (2.32).

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

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

($\Leftarrow$) Apply Lemma 2.35 for all π_i. $\square$

Corollary 2.37. *The category* **GMet** *has all products, and they are computed like in* L**Spa**.[278]

Unfortunately, this means that the notion of metric space originally defined in [Fré06], and incidentally what the majority of mathematicians calls a metric space, is not an instance of generalized metric space as we defined them. Since they only allow finite distances, some infinite products do not exist.[279] In general, if one wants to bound the distance above by some $B \in \mathsf{L}$, this can be done with the equation

[273] 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 inequality 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.

[274] It may remind you of Lemma 1.15 which states the same result for homomorphism and non-quantitative equations.

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

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

[277] When I is empty, the `L.H.S.` of (2.32) is vacuously true, and the `R.H.S.` is true since $\mathbf{A}$ is the terminal L-space which we showed satisfies all quantitative equations in Proposition 2.33.

[278] 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**.

[279] 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 Definition 0.1.

$x, y \vdash x =_B y$, 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 $x, y \vdash x =_1 y$.

Arguably, this is only a superficially negative result since it is already common in parts of the literature (e.g. [BvBR98, Law02, HST14]) to allow infinite distances because the resulting category of metric spaces has better properties (like having infinite products and coproducts).

Let us give two other conditions on $[0,\infty]$-spaces, arising in the definition of partial metrics [Mat94, Definition 3.1], which are not preserved under (finite) products.

Definition 2.38. A $[0,\infty]$-space (A, d) is called a **partial metric space** if it satisfies the following conditions :[280]

[280] There is some ambiguity in what $+$ and $-$ means when dealing with ∞ (the original paper [Mat94] supposes distances are finite), but it is irrelevant for us.

$$\forall a, b \in A, \quad a = b \iff d(a,a) = d(a,b) = d(b,b) \tag{2.33}$$
$$\forall a, b \in A, \quad d(a,a) \leq d(a,b) \tag{2.34}$$
$$\forall a, b \in A, \quad d(a,b) = d(b,a) \tag{2.35}$$
$$\forall a, b, c \in A, \quad d(a,c) \leq d(a,b) + d(b,c) - d(b,b) \tag{2.36}$$

These conditions look similar to what we were able to translate into equations before, but the first and last are problematic.[281]

[281] We can translate (2.34) into $x =_\varepsilon y \vdash x =_\varepsilon x$, and (2.35) is just symmetry which we can translate into $y =_\varepsilon x \vdash x =_\varepsilon y$.

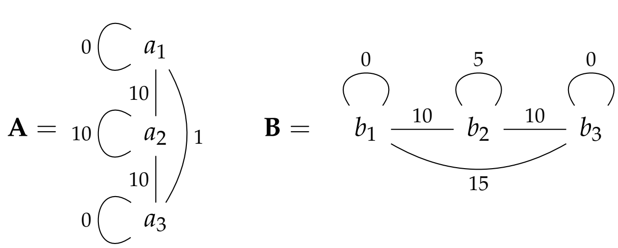
For (2.33), note that the forward implication is trivial, but for the converse, we would need to compare three distances at once inside the context, which seems impossible because the context only individually bounds distances by above. For (2.36), the problem comes from the minus operation on distances which will not interact well with upper bounds. Indeed, if we naively 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 huge (even ∞) and make the distance between x and z as close to 0 as we would like (provided we can take ε_1 and ε_2 finite).

These are just informal arguments, but thanks to Corollary 2.37, we can prove formally that these conditions are not expressible as (classes of) quantitative equations. Let **A** and **B** be the $[0,\infty]$-spaces pictured below (the distances are symmetric).[282]

[282] 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 $d_{\mathbf{B}}(b_2, b_3) = 10$.

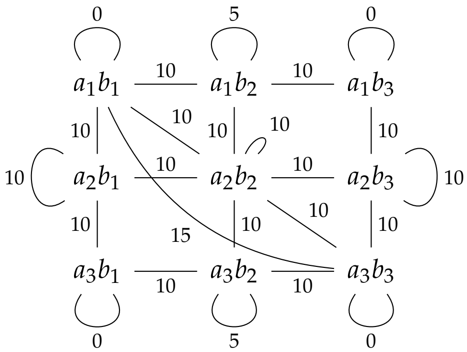


We can verify (by exhaustive checks) that **A** and **B** are partial metric spaces. If we take their product inside $[0,\infty]$**Spa**, we find the following $[0,\infty]$-space (some distances are omitted) which does not satisfy (2.33) nor (2.36).[283]

[283] For (2.33), the three points in the middle row $\{a_2b_1, a_2b_2, a_2b_3\}$ are all at distance 10 from each other and from themselves while not being equal. For (2.36), we have (on the diagonal)

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

but $15 > 10$.

$\mathbf{A} \times \mathbf{B} =$ 

We infer that there is no class $\hat{E}$ of quantitative equations such that $\mathbf{GMet}([0,\infty], \hat{E})$ is the full subcategory of $[0,\infty]\mathbf{Spa}$ containing all the partial metric spaces.[284]

[284] It is still possible that the category of partial metrics and nonexpansive maps is identified with some $\mathbf{GMet}(\mathsf{L}, \hat{E})$ for some cleverly picked L and $\hat{E}$. That would mean (infinite) products of partial metrics exist but they are not computed with supremums.

Coproducts

The case of coproducts in **GMet** is more delicate. While L**Spa** has coproducts, they do not always satisfy the equations satisfied by each of their components.

Proposition 2.39. *The category* **GMet** *has an initial object.*

Proof. The initial object $\varnothing$ in L**Spa** is the empty set with the only possible L-relation $\varnothing \times \varnothing \to \mathsf{L}$ (the empty function). The empty function $f : \varnothing \to X$ is always nonexpansive from $\varnothing$ to $\mathbf{X}$ because (2.3) is vacuously satisfied.

Just as for the terminal object, since **GMet** is a full subcategory of L**Spa**, it suffices to show $\varnothing$ is in **GMet** to conclude it is initial in this subcategory. We do this by showing $\varnothing$ satisfies absolutely all quantitative equations, and in particular those of $\hat{E}$. This is easily done because when $\mathbf{X}$ is not empty,[285] there are no assignments $\mathbf{X} \to \varnothing$, so $\varnothing$ vacuously satisfies $\mathbf{X} \vdash x = y$ and $\mathbf{X} \vdash x =_\varepsilon y$. $\square$

[285] The context of a quantitative equation cannot be empty because the variables, say x and y, must belong to the context.

Proposition 2.40. *The category* L**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:[286]

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

[286] 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 $\mathbf{A}$).

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 $a, b \in A_i$, $d(a,b) = d_i(a,b)$.[287] We show $\mathbf{A}$ with these coprojections is the coproduct $\coprod_{i \in I} \mathbf{A}_i$.

[287] Hence κ_i is even an isometric embedding.

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[288]

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

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

□

Because the distance between elements in different copies does not depend on the original spaces, it is easy to construct a quantitative equation that is not preserved by coproducts of L-spaces. For instance, even if all $\mathbf{A}_i$ satisfy $x,y \vdash x =_\varepsilon y$ for some fixed $\varepsilon \neq \top \in \mathsf{L}$,[289] the coproduct $\coprod_{i\in I}\mathbf{A}_i$ in L**Spa** does not satisfy it because some distances are $\top > \varepsilon$.

[289] i.e. there is an upper bound smaller than $\top$ on all distances in all $\mathbf{A}_i$.

Still, **GMet** always has coproducts as we will show in Corollary 3.51, but they are not computed like in L**Spa**, and they are not that easy to define.[290]

[290] In many cases like **Met** and **Poset**, they are computed like in L**Spa**.

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 L**Spa** is a bijection between the carriers. What is more, we show in Proposition 2.42 it must preserve distances on the nose, i.e. it is an isometry.

Definition 2.41 (Isometry)**.** A nonexpansive map $f : \mathbf{X} \to \mathbf{Y}$ is called an **isometry** if[291]

$$\forall x,x' \in X, \quad d_{\mathbf{Y}}(f(x),f(x')) = d_{\mathbf{X}}(x,x'). \tag{2.37}$$

[291] The inequality in (2.3) is replaced by an equation.

If furthermore f is injective, we call it an **isometric embedding.**[292] 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.

[292] This name is relatively rare because when dealing with metric spaces, the separation axiom implies that an isometry is automatically injective. This is also true for partial orders, where the name *order embedding* is common [DP02, Definition 1.34.(ii)].

Proposition 2.42. *In* **GMet***, isomorphisms are precisely the bijective isometries.*

Proof. We show a morphism $f : \mathbf{X} \to \mathbf{Y}$ 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 (2.3) twice, we find that for any $x,x' \in X$,[293]

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

[293] 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.

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

($\Leftarrow$) 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

$$d_{\mathbf{X}}(f^{-1}(y),f^{-1}(y')) = d_{\mathbf{X}}(f^{-1}f(x),f^{-1}f(x')) = d_{\mathbf{X}}(x,x') \overset{(2.37)}{=} d_{\mathbf{Y}}(f(x),f(x')) = d_{\mathbf{Y}}(y,y').$$

Hence f^{-1} is nonexpansive, it is even an isometry. □

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.[294]

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

Proposition 2.43. *Let $f : \mathbf{Y} \to \mathbf{Z}$ be an isometric embedding between* L*-spaces and ϕ a quantitative equation, then*

$$\mathbf{Z} \vDash \phi \implies \mathbf{Y} \vDash \phi. \tag{2.38}$$

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

$$\mathbf{Z} \vDash^{f\circ\hat{\iota}} \phi. \tag{2.39}$$

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

If $\phi = \mathbf{X} \vdash x = y$, then we have $\hat{\iota}(x) = \hat{\iota}(y)$ because we know $f\hat{\iota}(x) = f\hat{\iota}(x)$ by (2.39) and f is injective.

If $\phi = \mathbf{X} \vdash x =_\varepsilon y$, then we have $d_{\mathbf{Y}}(\hat{\iota}(x), \hat{\iota}(y)) = d_{\mathbf{Z}}(f\hat{\iota}(x), f\hat{\iota}(y)) \leq \varepsilon$, where the equation holds because f is an isometry and the inequality holds by (2.39). □

Corollary 2.44. *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.*[295]

[295] Both parts are immediate. The first follows from applying (2.38) to all ϕ in $\hat{E}$, the class of quantitative equations defining **GMet**. The second follows from the inclusion of a subspace being an isometric embedding.

Examples 2.45. Corollary 2.44 can be useful to identify some properties of L-spaces that cannot be modelled with quantitative equations. Here are a few 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 $\mathbf{GMet}(\mathsf{B}, \hat{E})$ for some $\hat{E}$? 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 class $\hat{E}$ defines total relations (i.e. $\mathbf{X} \vDash \hat{E}$ 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 2.44.[296]

   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[297] 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 metric from Example 2.14 is complete but the subspace $\mathbb{Q}$ is not. Thus, completeness cannot be defined via quantitative equations.[298]

[296] 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 $\mathbf{TotRel} \cong \mathbf{GMet}(\mathsf{L}, \hat{E})$.

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

[298] Still with the caveat that the full subcategory of complete metric spaces might still be isomorphic to some $\mathbf{GMet}(\mathsf{L}, \hat{E})$.

With this characterization of isomorphisms, we can also show the forgetful functor $U : \mathbf{GMet} \to \mathbf{Set}$ is an isofibration which concretely means that if you have a bijection $f : X \to Y$ and a generalized metric $d_{\mathbf{Y}}$ on Y, then you can construct a generalized metric $d_{\mathbf{X}}$ on X such that $f : \mathbf{X} \to \mathbf{Y}$ is an isomorphism. Indeed, if you let $d_{\mathbf{X}}(x, x') = d_{\mathbf{Y}}(f(x), f(x'))$, then f is automatically a bijective isometry.[299]

[299] Clearly, it is the unique distance on X that works, and we know that $\mathbf{X}$ belongs to $\mathbf{GMet}$ thanks to Corollary 2.44.

Definition 2.46 (Isofibration)**.** A functor $P : \mathbf{C} \to \mathbf{D}$ is called an **isofibration**[300] if for any isomorphism $f : X \to PY$ in $\mathbf{D}$, there is an isomorphism $g : X' \to Y$ such that $Pg = f$, in particular $PX' = X$.

[300] This term seems to have been coined by Lack and Paoli in [Lac07, §3.1] or [LP08, §6].

Proposition 2.47. *The forgetful functor* $U : \mathbf{GMet} \to \mathbf{Set}$ *is an isofibration.*

We wonder now how to complete the conceptual diagram below.

$$\text{isomorphism in } \mathbf{GMet} \longleftrightarrow \text{bijective isometries}$$
$$\text{??? in } \mathbf{GMet} \longleftrightarrow \text{isometric embeddings}$$

Since isometric embeddings correspond to subspaces, one might think that they are the monomorphisms in **GMet**. Unfortunately, they are way more restrained.[301] Any nonexpansive map that is injective is a monomorphism. To prove this, we rely on the existence of a space Ⱥ that informally *can pick elements*.

[301] They are the split monomorphisms, essentially by Footnote 293.

Proposition 2.48. *There is a generalized metric space* Ⱥ *on the set* $\{*\}$ *such that for any other space* $\mathbf{X}$*, any function* $f : \{*\} \to X$ *is a nonexpansive map* $\text{Ⱥ} \to \mathbf{X}$.[302]

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

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

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

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

Recall that $\mathbf{1}$ is a generalized metric space on the same set $\{*\}$, but with $d_{\mathbf{1}}(*, *) = \perp$. 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 $d_{\mathbf{X}}(x, x) = \perp$ for all $x \in X$, which is not always the case.[303]

[303] 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 $d_{\text{Ⱥ}}(*, *) = \varepsilon$ yields a space that lives in **GMet** but is not too restricted.

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

$$d_{\text{Ⱥ}}(*, *) = \sup \left\{ \varepsilon \in \mathsf{L} \mid (\{*\}, d_\varepsilon) \vDash \hat{E} \right\}.$$

It remains to check that any function $f : \{*\} \to X$ is nonexpansive from Ⱥ to $\mathbf{X} \in \mathbf{GMet}$. Consider the image of f seen as a subspace of $\mathbf{X}$. By Corollary 2.44, it belongs to **GMet** and hence satisfies $\hat{E}$. Moreover, it is clearly isomorphic to the L-space $(\{*\}, d_\varepsilon)$ with $\varepsilon = d_{\mathbf{X}}(f(*), f(*))$, which means that L-space satisfies $\hat{E}$ as well (by Corollary 2.44 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**.[304] □

[304] Use Lemma 2.35.

Proposition 2.49. *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 2.48, 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 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 find that $f \circ g = f \circ h$ also as functions. Since Uf is monic (i.e. injective), Ug and Uh must be equal, and since U is faithful, we obtain $g = h$. □

It remains to give a categorical characterization of isometric embeddings. This will rely on a well-known[305] abstract notion that we define here for completeness.

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

Definition 2.50 (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** (relative to F) 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.

$$\begin{tikzcd}
X \arrow[d, dashed, "\widehat{u}"'] \arrow[dr, "g"] & & & FX \arrow[d, "u"'] \arrow[dr, "Fg"] & \\
A \arrow[r, "f"'] & B \arrow[r, phantom, ""] & {} \arrow[from=1-2, to=1-3, "F", start anchor={[xshift=-2ex, yshift=-3ex]}, end anchor={[yshift=-3ex]}] & FA \arrow[r, "Ff"'] & FB
\end{tikzcd}$$

Example 2.51 (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 $u : \mathbf{X} \to \mathbf{A}$.[306]

[306] 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 $\mathbf{A}$.[307] In [AHS06, Definition 8.6], f is also called an *initial morphism*.

[307] 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 2.52. *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 2.49, 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 $\mathbf{X} \to \mathbf{A}$.[308] This is true because

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

where the equation follows from f being an isometry, and the inequality 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 $\mathbf{B}$ containing only the image of a and a' (its carrier is $\{f(a), f(a')\}$), and $u : X \to A$ be the function sending $f(a)$ to a and $f(a')$ to a'.[309] Notice that $f \circ u$ is the inclusion of $\mathbf{X}$ in $\mathbf{B}$ which is nonexpansive. Because f is cartesian, u must then be nonexpansive from $\mathbf{X}$ to $\mathbf{A}$ which implies

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

We conclude that f is an isometry. $\square$

[308] We use the second definition of cartesian in Example 2.51.

[309] We use the injectivity of f here.

Corollary 2.53. *If the composition $\mathbf{A} \xrightarrow{f} \mathbf{B} \xrightarrow{g} \mathbf{C}$ is an isometric embedding, then f is an isometric embedding.*[310]

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 $a, a' \in X$,[311]

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

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

[310] With the characterization of Proposition 2.52, this abstractly follows from [AHS06, Proposition 8.9]. We give the concrete proof anyways.

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

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

Proposition 2.54. *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 $\mathbf{A}$.[312]

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 L**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 **Set**, we know $Ug = Ug'$. Since U is faithful, we conclude that $g = g'$.[313] $\square$

[312] 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 $d_{\mathbf{A}_a}(a, *_a) = 0$.

[313] This direction works in an arbitrary **GMet**, that is, surjections are epic in any **GMet**.

The standard example to show that Proposition 2.54 does not generalize to an arbitrary **GMet** is the inclusion of $\mathbb{Q}$ into $\mathbb{R}$ with the Euclidean metric inside **Met**.

It is not surjective, but it is epic because any nonexpansive function from $\mathbb{R}$ is determined by its image on the rationals.[314]

[314] For any $r \in \mathbb{R}$, you can always find $q_n \in \mathbb{Q}$ such that $d(q_n, r) \leq \frac{1}{n}$, hence $d_{\mathbf{A}}(f(q_n), f(r)) \leq \frac{1}{n}$ for any nonexpansive $f : (\mathbb{R}, d) \to \mathbf{A}$. We infer that $f(r)$ is determined by the value of $f(q_n)$ for all n.

Proposition 2.55. *Let $f : \mathbf{A} \to \mathbf{B}$ be a split epimorphism between* L*-spaces and ϕ a quantitative equation, then*

$$\mathbf{A} \vDash \phi \implies \mathbf{B} \vDash \phi. \tag{2.40}$$

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.[315] Any nonexpansive assignment $\hat{\imath} : \mathbf{X} \to \mathbf{B}$ yields an assignment $g \circ \hat{\imath} : \mathbf{X} \to \mathbf{A}$. By hypothesis, we know that $\mathbf{A}$ satisfies ϕ for this particular assignment, namely,

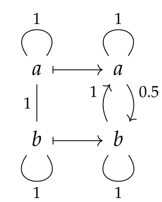
$$\mathbf{A} \vDash^{g \circ \hat{\imath}} \phi. \tag{2.41}$$

Now, we can apply Lemma 2.35 with $f : \mathbf{A} \to \mathbf{B}$ to obtain $\mathbf{B} \vDash^{f \circ g \circ \hat{\imath}} \phi$, and since $f \circ g = \mathrm{id}_{\mathbf{B}}$, we conclude $\mathbf{B} \vDash^{\hat{\imath}} \phi$. ☐

[315] Note that we already argued in Footnote 293 that the right inverse implies g is an isometric embedding. Then we could conclude by Corollary 2.44. The proof given here is essentially the same.

Remark 2.56. 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,[316] 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.

[316] Here is a graphical depiction:



Proposition 2.55 is basically a dual of Proposition 2.43 because isometric embeddings are split monomorphisms, so we do not get additional examples of properties that cannot be expressed with quantitative equations.[317]

[317] In theory, duality may help in some settings, but I find isometric embeddings are easier to grasp.

Discrete Spaces

The forgetful functor $U : \mathbf{GMet} \to \mathbf{Set}$ has a left adjoint. Its concrete description is too involved, so we will prove this later in Corollary 3.49, but for the special case of $\mathsf{L}\mathbf{Spa}$, we can prove it now.

Proposition 2.57. *The forgetful functor $U : \mathsf{L}\mathbf{Spa} \to \mathbf{Set}$ has a left adjoint.*

Proof. For any set X, we define the **discrete space** $\mathbf{X}_\top$ to be the set X equipped with the L-relation $d_\top : X \times X \to \mathsf{L}$ sending any pair to $\top$.

For any L-space $\mathbf{A}$ and function $f : X \to A$, the function f is nonexpansive from $\mathbf{X}_\top$ to $\mathbf{A}$, thus $\mathbf{X}_\top$ is the free object on X (relative to U).

We conclude there is a functor $F : \mathbf{Set} \to \mathsf{L}\mathbf{Spa}$ sending X to $\mathbf{X}_\top$ that is left adjoint to $U : \mathsf{L}\mathbf{Spa} \to \mathbf{Set}$.[318] ☐

[318] This follows from an abstract categorical argument, see e.g., [Awo10, Proposition 9.4].

3 Universal Quantitative Algebra

Saxophone concerto E minor OP 88

Deluxe

For a comprehensive introduction to the concepts and themes explored in this chapter, please refer to §0.3. Here, we only give a brief overview.

It is time to combine what we learned about universal algebra in Chapter 1 and about generalized metric spaces in Chapter 2 to develop universal quantitative algebra. We follow an outline similar to that of Chapter 1 for definitions, results, and proofs, and we give some examples (reusing those of the previous chapters) throughout this chapter.

Outline: In §3.1, we define quantitative algebras and quantitative equations over a signature, and we explain how to construct the free quantitative algebras. In §3.2, we give the rules for quantitative equational logic to derive quantitative equations from other quantitative equations, and we show it is sound and complete. In §3.3, we define presentations for monads on generalized metric spaces, and we give some examples.[319] In §3.4, we show that any monad lifting of a **Set** monad with an algebraic presentation to **GMet** can also be presented.

[319] Notice the parallel with the outline of Chapter 1.

In the sequel and unless otherwise stated, Σ is an arbitrary signature and **GMet** is an arbitrary category of generalized metric spaces defined by a class $\hat{\mathbf{E}}_{\mathbf{GMet}}$ of quantitative equations.[320]

[320] Those defined in Definition 2.22.

3.1 Quantitative Algebras

Definition 3.1 (Quantitative algebra)**.** A **quantitative Σ-algebra** (or just quantitative algebra)[321] is a set A equipped with a Σ-algebra structure $(A, [\![-]\!]_A) \in \mathbf{Alg}(\Sigma)$ and a generalized metric space structure $(A, d_{\mathbf{A}}) \in \mathbf{GMet}$. We will switch between using the single symbol $\hat{\mathbb{A}}$ or the triple $(A, [\![-]\!]_A, d_{\mathbf{A}})$ when referring to a quantitative algebra, we will also write $\mathbb{A}$ for the **underlying** Σ-algebra, $\mathbf{A}$ for the underlying space, and A for the underlying set.

[321] We sometimes write simply algebra, with the `knowldege` link going to this definition.

A **homomorphism** from $\hat{\mathbb{A}}$ to $\hat{\mathbb{B}}$ is a function $h : A \to B$ between the underlying sets of $\hat{\mathbb{A}}$ and $\hat{\mathbb{B}}$ that is both a homomorphism $h : \mathbb{A} \to \mathbb{B}$ and a nonexpansive function $h : \mathbf{A} \to \mathbf{B}$. We sometimes emphasize and call h a nonexpansive homomorphism.[322] 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 quantitative algebras and morphisms are nonexpansive homomorphisms.

[322] We will not distinguish between a nonexpansive homomorphism $h : \hat{\mathbb{A}} \to \hat{\mathbb{B}}$ and its underlying homomorphism or nonexpansive function or function. We may write Uh with U being the appropriate forgetful functor when necessary.

We denote it by $\mathbf{QAlg}(\Sigma)$.

This category is concrete over $\mathbf{Set}$, $\mathbf{Alg}(\Sigma)$, $\mathbf{GMet}$ with forgetful functors:

- $U : \mathbf{QAlg}(\Sigma) \to \mathbf{Set}$ sends a quantitative algebra $\hat{\mathbb{A}}$ to its underlying set A and a nonexpansive homomorphism to the underlying function between carriers.

- $U : \mathbf{QAlg}(\Sigma) \to \mathbf{Alg}(\Sigma)$ sends $\hat{\mathbb{A}}$ to its underlying algebra $\mathbb{A}$ and a nonexpansive homomorphism to the underlying homomorphism.

- $U : \mathbf{QAlg}(\Sigma) \to \mathbf{GMet}$ sends $\hat{\mathbb{A}}$ to its underlying space $\mathbf{A}$ and a nonexpansive homomorphism to the underlying nonexpansive function.

One can quickly check that the following diagram commutes, and that it yields an alternative definition of $\mathbf{QAlg}(\Sigma)$ as a pullback of categories.[323] We can also mention there is another forgetful functor $U : \mathbf{QAlg}(\Sigma) \to \mathsf{L}\mathbf{Spa}$ obtained by composing $U : \mathbf{QAlg}(\Sigma) \to \mathbf{GMet}$ with the inclusion $\mathbf{GMet} \to \mathsf{L}\mathbf{Spa}$.

[323] We do not spend time making this precise, but it post-rigorously makes the case for universal quantitative algebra as a straightforward combination of universal algebra and generalized metric spaces.

$$\begin{tikzcd}
\mathbf{QAlg}(\Sigma) \arrow[r, "U"] \arrow[d, "U"'] \arrow[dr, "U"] \arrow[dr, phantom, "\lrcorner", very near start] & \mathbf{GMet} \arrow[d, "U"] \\
\mathbf{Alg}(\Sigma) \arrow[r, "U"'] & \mathbf{Set}
\end{tikzcd}$$

Example 3.2. Since a quantitative algebra is just an algebra and a generalized metric space on the same set, we can find simple examples by combining pieces we have already seen.

1. In Example 1.4, we saw that an algebra for the signature $\Sigma = \{\mathsf{p} : 0\}$ is just a pair (X, x) comprising a set X with a distinguished point $x \in X$. In Example 2.14, we discussed the $\mathbb{N}_\infty$-space (H, d) where H is the set of humans and d is the collaboration distance. We can therefore consider the quantitative Σ-algebras $(H, \text{Paul Erdös}, d)$, which is the set of all humans with Paulo Erdös as a distinguished point and the collaboration distance.[324]

2. In Example 1.4, we saw the $\{\mathsf{f} : 1\}$-algebra $\mathbb{Z}$ where f is interpreted as adding 1. On top of that, we consider the B-relation corresponding to the partial order $\leq$ on $\mathbb{Z}$: $d_\leq : \mathbb{Z} \times \mathbb{Z} \to \mathsf{B}$ that sends (n, m) to $\bot$ if and only if $n \leq m$. We get a quantitative algebra $(\mathbb{Z}, - + 1, d_\leq)$.[325]

3. In Example 2.14, we saw that $\mathbb{R}$ equipped with the Euclidean distance d is a metric space, i.e. an object of $\mathbf{GMet} = \mathbf{Met}$. The addition of real numbers is the most natural interpretation of $\Sigma = \{+ : 2\}$, thus we get a quantitative algebra $(\mathbb{R}, +, d)$.

[324] Note that $\mathbf{GMet}$ is instantiated as $\mathbb{N}_\infty\mathbf{Spa}$, i.e. $\mathsf{L} = \mathbb{N}_\infty$ and $\hat{E}_{\mathbf{GMet}} = \emptyset$.

[325] This time, $\mathbf{GMet}$ is instantiated as $\mathbf{Poset}$ with $\mathsf{L} = \mathsf{B}$ and $\hat{E}_{\mathbf{GMet}} = \hat{E}_{\mathbf{Poset}}$ as defined after Definition 2.30.

Remark 3.3. Already here, we covered three examples that are not possible with the original (and predominant in the literature) definition of quantitative algebras [MPP16, Definition 3.1]. The first two are not possible because the base category is not $\mathbf{Met}$. The third is not possible even if it deals with metric spaces.

Indeed, as already noted in [Adá22, Remark 3.1.(2)], the addition of real numbers is not a nonexpansive function $(\mathbb{R}, d) \times (\mathbb{R}, d) \to (\mathbb{R}, d)$, where $\times$ denotes the categorical product because,[326] recalling Corollary 2.37, we have

$$(d \times d)((1,1),(2,2)) = \sup\{d(1,2), d(1,2)\} = 1 < 2 = d(2,4) = d(1+1, 2+2).$$

[326] In [MPP16], the interpretation of an n-ary operation symbol is required to be a nonexpansive map from the n-wise product of the carrier to the carrier.

Here are two more compelling examples from the original paper [MPP16].

Example 3.4 (Hausdorff)**.** In Example 2.16, we defined the Hausdorff distance $d^{\uparrow}$ on $\mathcal{P}_{\text{ne}} X$ that depends on an L-relation $d : X \times X \to \mathsf{L}$. In Example 1.66, we described a $\Sigma_{\mathbf{S}}$-algebra structure on $\mathcal{P}_{\text{ne}} X$ (interpreting $\oplus$ as union). Combining these, we get a quantitative $\Sigma_{\mathbf{S}}$-algebra $(\mathcal{P}_{\text{ne}} X, \cup, d^{\uparrow})$ for any L-space (X, d).

If we know that (X, d) satisfies some quantitative equations in $\hat{E}_{\mathbf{GMet}}$, we can sometimes prove that $(\mathcal{P}_{\text{ne}} X, d^{\uparrow})$ does too. For instance, picking $\mathsf{L} = [0,1]$ or $\mathsf{L} = [0,\infty]$, $\mathbf{GMet} = \mathbf{Met}$, and $\hat{E}_{\mathbf{GMet}} = \hat{E}_{\mathbf{Met}}$, one can show that if (X, d) belongs to $\mathbf{Met}$, then so does $(\mathcal{P}_{\text{ne}} X, d^{\uparrow})$, and we still get a quantitative $\Sigma_{\mathbf{S}}$-algebra $(\mathcal{P}_{\text{ne}} X, \cup, d^{\uparrow})$, now over $\mathbf{Met}$.[327]

[327] This is the quantitative algebra denoted by $\Pi[M]$ in [MPP16, Theorem 9.2].

Example 3.5 (Kantorovich)**.** Given a L-relation $d : X \times X \to [0,1]$, we define the **Kantorovich distance** d_{K} on $\mathcal{D}X$ as follows:[328] for all $\varphi, \psi \in \mathcal{D}X$,

$$d_{\mathrm{K}}(\varphi, \psi) = \inf \left\{ \sum_{(x,x')} \tau(x,x') d(x,x') \mid \tau \in \mathcal{D}(X \times X), \mathcal{D}\pi_1(\tau) = \varphi, \mathcal{D}\pi_2(\tau) = \psi \right\}.$$

[328] This lifting of a distance on X to a distance on $\mathcal{D}X$ is well-known in optimal transport theory [Vil09]. You can find a well-written concise description of d_{K} in [BBKK18, §2.1] in the case $\mathsf{L} = [0,\infty]$ where it is denoted $d^{\downarrow\mathcal{D}}$. They also give a dual description as we did for the Hausdorff distance in Example 2.16, but the strong duality result ($d^{\downarrow\mathcal{D}} = d^{\uparrow\mathcal{D}}$) does not hold in general.

The distributions τ above range over **couplings** of φ and ψ, i.e. distributions over $X \times X$ whose marginals are φ and ψ. Thus, what d_{K} does, in words, is computing the average distance according to all couplings, and then taking the smallest one.

In Example 1.67, we gave a $\Sigma_{\mathbf{CA}}$-algebra structure on $\mathcal{D}X$ (interpreting $+_p$ as convex combination). Combining the algebra and the $[0,1]$-space, we get a quantitative $\Sigma_{\mathbf{CA}}$-algebra $(\mathcal{D}X, [\![-]\!]_{\mathcal{D}X}, d_{\mathrm{K}})$. Once again, we can prove that if (X, d) is a metric space, then so is $(\mathcal{D}X, d_{\mathrm{K}})$, and we obtain a quantitative algebra $(\mathcal{D}X, [\![-]\!]_{\mathcal{D}X}, d_{\mathrm{K}})$ over $\mathbf{Met}$.[329]

[329] This is the quantitative algebra denoted by $\Pi[M]$ in [MPP16, Theorem 10.4].

Unlike the first examples, the interpretations in $(\mathcal{P}_{\text{ne}} X, \cup, d^{\uparrow})$ and $(\mathcal{D}X, [\![-]\!]_{\mathcal{D}X}, d_{\mathrm{K}})$ are nonexpansive with respect to the product distance. Concretely,

$$\forall S, S', T, T' \in \mathcal{P}_{\text{ne}} X, \qquad d^{\uparrow}(S \cup S', T \cup T') \leq \max\left\{ d^{\uparrow}(S,T), d^{\uparrow}(S',T') \right\} \tag{3.1}$$

$$\forall \varphi, \varphi', \psi, \psi' \in \mathcal{D}X, \qquad d_{\mathrm{K}}(p\varphi + \overline{p}\varphi', p\psi + \overline{p}\psi') \leq \max\left\{ d_{\mathrm{K}}(\varphi,\psi), d_{\mathrm{K}}(\varphi',\psi') \right\}. \tag{3.2}$$

The initial motivation to remove this requirement and arrive at Definition 3.1[330] came from a variant of the Kantorovich distance called the **Łukaszyk–Karmowski** (ŁK for short) distance [Łuk04, Eq. (21)] which sends $\varphi, \psi \in \mathcal{D}X$ to

[330] Which imposes no further relation between the Σ-algebra and the L-space other than being on the same set.

$$d_{\text{ŁK}}(\varphi, \psi) = \sum_{(x,x')} \varphi(x)\psi(x')d(x,x'). \tag{3.3}$$

In words, instead of looking at many different couplings to find the best one, we only look at the independent coupling $\tau(x,x') = \varphi(x)\psi(x')$.[331] In particular, it coincides

[331] The ŁK distance is easier to compute than the Kantorovich distance since there is no optimization. It is the reason why it was considered in [CKPR21] for an application to reinforcement learning.

with the Kantorovich distance on Dirac distributions since the independent coupling of δ_x and δ_y is the only coupling, we obtain

$$d_{\mathrm{K}}(\delta_x,\delta_y) = d_{\mathrm{ŁK}}(\delta_x,\delta_y) = d(x,y).$$

We can that convex combination is not nonexpansive with respect to the product of the ŁK distance, namely, there exists a $[0,1]$-space (X,d), distributions $\varphi,\varphi',\psi,\psi' \in \mathcal{D}X$, and $p \in (0,1)$ such that

$$d_{\mathrm{ŁK}}(p\varphi + \overline{p}\varphi', p\psi + \overline{p}\psi') > \sup\left\{d_{\mathrm{ŁK}}(\varphi,\psi), d_{\mathrm{ŁK}}(\varphi',\psi')\right\}.$$

Take $X = \{x,y\}$ with $d(x,y) = d(y,x) = 1$ and the self-distances being 0,[332] then for any $p \in (0,1)$,

$$\begin{aligned} d_{\mathrm{ŁK}}(p\delta_x + \overline{p}\delta_y, p\delta_x + \overline{p}\delta_y) &= p^2 d(x,x) + p\overline{p}d(x,y) + \overline{p}pd(y,x) + \overline{p}^2 d(y,y) \\ &= 2p\overline{p} \\ &> 0 \\ &= \sup\{0,0\} \\ &= \sup\left\{d_{\mathrm{ŁK}}(\delta_x,\delta_x), d_{\mathrm{ŁK}}(\delta_y,\delta_y)\right\}. \end{aligned}$$

Therefore, $(\mathcal{D}X, [\![-]\!]_{\mathcal{D}X}, d_{\mathrm{ŁK}})$ is always a quantitative algebra in the sense of Definition 3.1, but not always in the sense of [MPP16, Definition 3.1].[333]

[332] We gave another example in [MSV22, Lemma 5.3].

[333] In fact, even if d is a metric , $d_{\mathrm{ŁK}}$ is not a metric (by the example above, self-distances are not always 0, so it does not satisfy $x \vdash x =_0 x$). That is another reason why [MPP16] does not apply.

Quantitative Equations

Now, in order to get back the expressiveness of the original framework, we need a way to impose this property of nonexpansiveness with respect to the product distance, and we also need a way to impose other properties like the fact that $\oplus$ should be interpreted as a commutative operation. We achieve both things at once with the following definition.

Definition 3.6 (Quantitative Equation)**.** A **quantitative equation** (over Σ and L) is a tuple comprising an L-space $\mathbf{X}$ called the **context**,[334] two terms $s,t \in \mathcal{T}_\Sigma X$ and optionally a quantity $\varepsilon \in \mathsf{L}$. We write these as $\mathbf{X} \vdash s = t$ when no ε is given or $\mathbf{X} \vdash s =_\varepsilon t$ when it is given.

An quantitative algebra $\hat{\mathbb{A}}$ **satisfies** a quantitative equation[335]

- $\mathbf{X} \vdash s = t$ if for any nonexpansive assignment $\hat{\iota} : \mathbf{X} \to \mathbf{A}$, $[\![s]\!]_A^{\hat{\iota}} = [\![t]\!]_A^{\hat{\iota}}$.
- $\mathbf{X} \vdash s =_\varepsilon t$ if for any nonexpansive assignment $\hat{\iota} : \mathbf{X} \to \mathbf{A}$, $d_{\mathbf{A}}([\![s]\!]_A^{\hat{\iota}}, [\![t]\!]_A^{\hat{\iota}}) \leq \varepsilon$.

We use ϕ and ψ to refer to a quantitative equation, and we sometimes call them simply equations with the `knowldege` link going here. We write $\hat{\mathbb{A}} \vDash \phi$ when $\hat{\mathbb{A}}$ satisfies ϕ,[336] and we also write $\hat{\mathbb{A}} \vDash^{\hat{\iota}} \phi$ when the equality $[\![s]\!]_A^{\hat{\iota}} = [\![t]\!]_A^{\hat{\iota}}$ or the bound $d_{\mathbf{A}}([\![s]\!]_A^{\hat{\iota}}, [\![t]\!]_A^{\hat{\iota}}) \leq \varepsilon$ holds for a particular assignment $\hat{\iota} : \mathbf{X} \to \mathbf{A}$ (and not necessarily for all assignments).

[334] Note that even with algebras in **GMet**, the context is in **LSpa**. This differs slightly from [FMS21].

[335] Formally, we would need to write $[\![-]\!]_A^{U\hat{\iota}}$ instead of $[\![-]\!]_A^{\hat{\iota}}$ because $U\hat{\iota} : X \to A$ is the assignment we use to interpret the terms.

[336] As usual, satisfaction generalizes to classes of quantitative equations, i.e. if $\hat{E}$ is a classes of quantitative equations, $\hat{\mathbb{A}} \vDash \hat{E}$ means $\hat{\mathbb{A}} \vDash \phi$ for all $\phi \in \hat{E}$.

Our overloading of the terminology *quantitative equation* (recall Definition 2.22) is practically harmless because a quantitative equation from Chapter 2 $\mathbf{X} \vdash x = y$ (or $\mathbf{X} \vdash x =_\varepsilon y$) can be seen as the new kind of quantitative equation by viewing x and y as terms via the embedding η^Σ_X. Formally, since $[\![\eta^\Sigma_X(x)]\!]^{\hat{\iota}}_A = \hat{\iota}(x)$ for any $x \in X$ and $\hat{\iota} : \mathbf{X} \to \mathbf{A}$,[337]

$$\begin{aligned} \mathbf{A} \vDash \mathbf{X} \vdash x = y &\iff \hat{\mathbb{A}} \vDash \mathbf{X} \vdash \eta^\Sigma_X(x) = \eta^\Sigma_X(y) \\ \mathbf{A} \vDash \mathbf{X} \vdash x =_\varepsilon y &\iff \hat{\mathbb{A}} \vDash \mathbf{X} \vdash \eta^\Sigma_X(x) =_\varepsilon \eta^\Sigma_X(y). \end{aligned} \tag{3.4}$$

[337] Later on, we will seldom distinguish between x and $\eta^\Sigma_X(x)$ and write the former for simplicity.

In particular, since we assumed the underlying space of any $\hat{\mathbb{A}} \in \mathbf{QAlg}(\Sigma)$ to be a generalized metric space, we can say that $\hat{\mathbb{A}} \vDash \phi$ for any $\phi \in \hat{E}_{\mathbf{GMet}}$.[338] Another consequence is that over the empty signature $\Sigma = \emptyset$, the quantitative equations from Definition 2.22 and Definition 3.6 are the same.

[338] We implicitly see the equations in $\hat{E}_{\mathbf{GMet}}$ as the new kind of equations from Definition 3.6.

Furthermore, the new quantitative equations also generalize the equations of universal algebra (Definition 1.11). Indeed, given an equation $X \vdash s = t$, we construct the quantitative equation $\mathbf{X}_\top \vdash s = t$ where the new context is the discrete space on the old context. We show that

$$\mathbb{A} \vDash X \vdash s = t \iff \hat{\mathbb{A}} \vDash \mathbf{X}_\top \vdash s = t. \tag{3.5}$$

By Proposition 2.57, any assignment $\iota : X \to A$ is nonexpansive from $\mathbf{X}_\top$ to $\mathbf{A}$. Any nonexpansive assignment $\hat{\iota} : \mathbf{X}_\top \to \mathbf{A}$ also yields an assignment $X \to A$ by applying the forgetful functor U since the carrier of $\mathbf{X}_\top$ is X. Therefore, the interpretations of s and t coincide under all assignments if and only if they coincide under all nonexpansive assignments.

Remark 3.7. The name quantitative equation is already used in, e.g., [MPP16, MPP17, Adá22, ADV23], and it essentially refers to our quantitative equations with a quantity and a discrete context. We believe Definition 3.6 is a more accurate analog to the equations in equational logic, hence we propose to call those quantitative equations.

Let us get to more interesting examples now.[339]

[339] More examples are in the papers we cited in the introduction when we talked about universal algebra on partial orders and on metric spaces. In particular, there is a long list in [AFMS21, Example 3.19].

Example 3.8 (Almost commutativity)**.** Let $+ : 2 \in \Sigma$ be a binary operation symbol. As shown above, to ensure $+$ is interpreted as a commutative operation in a quantitative algebra, we can use the quantitative equation $\mathbf{X}_\top \vdash x + y = y + x$ where $X = \{x, y\}$. In fact, using the same syntactic sugar as we did in Chapter 2 to avoid explicitly describing all the context, we can write $x, y \vdash x + y = y + x$.[340]

[340] Whenever we will write $x_1, \dots, x_n \vdash s = t$, we will mean $\mathbf{X}_\top \vdash s = t$ where $X = \{x_1, \dots, x_n\}$, and similarly for $=_\varepsilon$.

Since the context can be any L-space, we can now add some nuance to the commutativity property. For instance, we can guarantee that $+$ is commutative only between elements that are close to each other with $x =_\varepsilon y \vdash x + y = y + x$ where $\varepsilon \in \mathsf{L}$ is fixed.[341] Unrolling the syntactic sugar, the context is the L-space containing two points x and y with $d_{\mathbf{X}}(x, y) = \varepsilon$ and all other distances being $\top$. Therefore, a nonexpansive assignment $\hat{\iota} : \mathbf{X} \to \mathbf{A}$ is a choice of two elements $\hat{\iota}(x)$ and $\hat{\iota}(y)$ with $d_{\mathbf{A}}(\hat{\iota}(x), \hat{\iota}(y)) \leq \varepsilon$ and no other constraint. We conclude that $\hat{\mathbb{A}}$ satisfies $x =_\varepsilon y \vdash x + y = y + x$ if and only if $[\![+]\!]_A(a, b) = [\![+]\!]_A(b, a)$ whenever $d_{\mathbf{A}}(a, b) \leq \varepsilon$.

[341] This example comes from [Adá22, Example 2.8.(4)].

Another possible variant on commutativity is $x =_\bot x, y =_\bot y \vdash x + y = y + x$. This means $+$ is guaranteed to be commutative only on elements which have a self-distance of $\bot$. For instance, in distributions with the ŁK distance, $d_{ŁK}(\varphi, \varphi) = 0$ only when the elements in the support of φ are all at distance 0 from each other. In particular, when d is a metric, $d_{ŁK}(\varphi, \varphi) = 0$ if and only if φ is a Dirac distribution. So that quantitative equation would ensure commutativity only on Dirac distributions.

Remark 3.9. Note that our syntactic sugar now allow terms that are not variables in the conclusion but still not in the premises. This is in contrast with the quantitative inferences of [MPP16] as they allow arbitrary terms in the premises. Thus, when the signature is not empty, our quantitative equations cannot correspond their quantitative inferences. The authors had already identified the restriction to variables was valuable, and sometimes necessary, they call the restricted judgments basic quantitative inferences.[342] Following [MSV23, Lemma 8.4], one could prove that our quantitative equations are equivalent to quantitative inferences whose premises only contain variables.

[342] Basic quantitative inferences are further restricted to have a finite set of premises.

Example 3.10 (Nonexpansiveness)**.** We can translate (3.1) and (3.2) into the following (family of) quantitative equations.

$$\forall \varepsilon, \varepsilon' \in \mathsf{L}, \quad x =_\varepsilon y, x' =_{\varepsilon'} y' \vdash x \oplus x' =_{\max\{\varepsilon,\varepsilon'\}} y \oplus y' \tag{3.6}$$

$$\forall \varepsilon, \varepsilon' \in \mathsf{L}, \quad x =_\varepsilon y, x' =_{\varepsilon'} y' \vdash x +_p x' =_{\max\{\varepsilon,\varepsilon'\}} y +_p y' \tag{3.7}$$

The quantitative algebra from Example 3.4 satisfies (3.6), and the one from Example 3.5 satisfies (3.7), but the variant with the ŁK distance does not satisfy (3.7).

In general, if we want an n-ary operation symbol $\mathsf{op} \in \Sigma$ to be interpreted as a nonexpansive map $\mathbf{A}^n \to \mathbf{A}$, we can impose the equations[343]

[343] This is an axiom in the logic of [MPP16]. It is not in our formulation of quantitative equational logic.

$$\forall \{\varepsilon_i\}_{i \in I} \subseteq \mathsf{L}, \quad \{x_i =_{\varepsilon_i} y_i \mid 1 \leq i \leq n\} \vdash \mathsf{op}(x_1, \ldots, x_n) =_{\max_i \varepsilon_i} \mathsf{op}(y_1, \ldots, y_n). \tag{3.8}$$

Example 3.11 (L-nonexpansiveness)**.**[344] In most papers on quantitative algebras this property is called "nonexpansiveness of the operations". In [MSV22], we remarked this can be ambiguous because one could consider a different distance on n-tuples of inputs than the product distance. We then presented quantitative algebras for *lifted signature* which can deal with more general operations.

[344] c.f. [AFMS21, Examples 3.19.(2) and 3.19.(3)].

In a lifted signature, each operation symbol $\mathsf{op} : n \in \Sigma$ comes with an assignment $(A, d) \mapsto (A^n, L_{\mathsf{op}}(d))$ (on generalized metric spaces) which specifies the distance $L_{\mathsf{op}}(d)$ on n-tuples that needs to be considered. We say that the interpretation $[\![\mathsf{op}]\!]_A$ is L_{op}-nonexpansive when it is a nonexpansive map $[\![\mathsf{op}]\!]_A : (A^n, L_{\mathsf{op}}(d)) \to (A, d)$.[345] We can also express L_{op}-nonexpansiveness with a family of quantitative equations like we did in Example 3.10:[346]

[345] See [MSV22, Definitions 3.4 and 3.6].

[346] This is the L-NE rule of [MSV22, Definition 3.11], but it has been written more cleanly with quantitative equations with contexts.

$$\forall \mathbf{X} \in \mathbf{GMet}, \forall x, y \in X^n, \quad \mathbf{X} \vdash \mathsf{op}(x_1, \ldots, x_n) =_{L_{\mathsf{op}}(d_{\mathbf{X}})(x,y)} \mathsf{op}(y_1, \ldots, y_n). \tag{3.9}$$

If an algebra $\hat{\mathbb{A}}$ satisfies these equations, then in particular, for all $a, b \in A^n$, it satisfies $\mathbf{A} \vdash \mathsf{op}(a_1, \ldots, a_n) =_{L_{\mathsf{op}}(d_{\mathbf{A}})(a,b)} \mathsf{op}(b_1, \ldots, b_n)$ under the assignment $\mathrm{id}_A : \mathbf{A} \to \mathbf{A}$. This means

$$d_{\mathbf{A}}([\![\mathsf{op}]\!]_A(a_1, \ldots, a_n), [\![\mathsf{op}]\!]_A(b_1, \ldots, b_n)) \leq L_{\mathsf{op}}(d_{\mathbf{A}})(a, b),$$

so we conclude that $[\![\mathsf{op}]\!]_A : (A^n, L_{\mathsf{op}}(d_{\mathbf{A}})) \to \mathbf{A}$ is nonexpansive.

Now, we still have to show that L_{op}-nonexpansiveness is the only consequence of (3.9). This requires an assumption on L_{op} that morally says the distance between tuples x and y in $(X^n, L_{\mathsf{op}}(d_{\mathbf{X}}))$ depends only on the distances between the coordinates $x_1, \dots, x_n$ and $y_1, \dots, y_n$ in $\mathbf{X}$.[347] We refer to [MSV22] for more details, in particular Definitions 3.1 and 3.2 give the condition on L_{op}.[348]

As a particular case, one can take $L_{\mathsf{op}}(d)$ to be the product distance and recover the original nonexpansiveness of Example 3.10. Another interesting instance is taking $L_{\mathsf{op}}(d)$ to be the discrete distance (in case $\mathbf{GMet} = \mathsf{L}\mathbf{Spa}$, $\forall x, y \in X^n, L_{\mathsf{op}}(x,y) = \top$), then (3.9) becomes trivial as we will see in Lemma 3.26. Intuitively, it is because any function from the discrete space on A^n to $\mathbf{A}$ is nonexpansive.

[347] This is the case for nonexpansiveness with respect to the product distance. In fact, the only distances that matter there are the pairwise $d_{\mathbf{X}}(x_i, y_i)$ for all i. For L_{op}-nonexpansiveness, the other distances like $d_{\mathbf{X}}(x_1, x_1)$ or $d_{\mathbf{X}}(y_3, x_1)$ may be important, but never $d_{\mathbf{X}}(x, z)$ for some fresh z.

[348] Briefly, we need L_{op} to be a functor that preserves isometric embeddings.

Example 3.12 (Convexity)**.** The quantitative algebra $(\mathcal{D}X, [\![-]\!]_{\mathcal{D}X}, d_{\mathrm{K}})$ satisfies another family of quantitative equations that is stronger than (3.7):[349]

$$\forall \varepsilon, \varepsilon' \in \mathsf{L}, \quad x =_{\varepsilon} y, x' =_{\varepsilon'} y' \vdash x +_p x' =_{p\varepsilon + \overline{p}\varepsilon'} y +_p y'. \tag{3.10}$$

This property of $[\![+_p]\!]_{\mathcal{D}X}$ is called convexity in, e.g., [MV20, Definition 30].

[349] Instead of taking the maximum between ε and ε', we take their convex combination, and since the former is always larger than the latter, (3.10) is stronger than (3.7).

As a sanity check for our definitions, we can verify that homomorphisms preserve the satisfaction of quantitative equations.[350]

[350] Just like we did in Lemma 1.15 for **Set** and Lemma 2.35 for $\mathsf{L}\mathbf{Spa}$. In fact, the proofs are very similar.

Lemma 3.13. *Let ϕ be an equation with context $\mathbf{X}$. If $h : \hat{\mathbb{A}} \to \hat{\mathbb{B}}$ is a homomorphism and $\hat{\mathbb{A}} \vDash^{\hat{\iota}} \phi$ for an assignment $\hat{\iota} : \mathbf{X} \to \mathbf{A}$, then $\hat{\mathbb{B}} \vDash^{h \circ \hat{\iota}} \phi$.*

Proof. We have two very similar cases. Let ϕ be the equation $\mathbf{X} \vdash s = t$, we have

$$\begin{aligned}
\hat{\mathbb{A}} \vDash^{\hat{\iota}} \phi &\iff [\![s]\!]_A^{\hat{\iota}} = [\![t]\!]_A^{\hat{\iota}} && \text{definition of } \vDash \\
&\implies h([\![s]\!]_A^{\hat{\iota}}) = h([\![t]\!]_A^{\hat{\iota}}) \\
&\implies [\![s]\!]_B^{h\circ\hat{\iota}} = [\![t]\!]_B^{h\circ\hat{\iota}} && \text{by (1.11)} \\
&\iff \hat{\mathbb{B}} \vDash^{h\circ\hat{\iota}} \phi. && \text{definition of } \vDash
\end{aligned}$$

Let ϕ be the equation $\mathbf{X} \vdash s =_{\varepsilon} t$, we have

$$\begin{aligned}
\hat{\mathbb{A}} \vDash^{\hat{\iota}} \phi &\iff d_{\mathbf{A}}([\![s]\!]_A^{\hat{\iota}}, [\![t]\!]_A^{\hat{\iota}}) \leq \varepsilon && \text{definition of } \vDash \\
&\implies d_{\mathbf{A}}(h([\![s]\!]_A^{\hat{\iota}}), h([\![t]\!]_A^{\hat{\iota}})) \leq \varepsilon \\
&\implies d_{\mathbf{A}}([\![s]\!]_B^{h\circ\hat{\iota}}, [\![t]\!]_B^{h\circ\hat{\iota}}) \leq \varepsilon && \text{by (1.11)} \\
&\iff \hat{\mathbb{B}} \vDash^{h\circ\hat{\iota}} \phi. && \text{definition of } \vDash
\end{aligned}$$ □

Definition 3.14 (Quantitative variety)**.** Given a class $\hat{E}$ of quantitative equations, a $(\Sigma, \hat{E})$**-algebra** is a quantitative Σ-algebra that satisfies $\hat{E}$. We define $\mathbf{QAlg}(\Sigma, \hat{E})$, the category of $(\Sigma, \hat{E})$-algebras, to be the full subcategory of $\mathbf{QAlg}(\Sigma)$ containing only those algebras that satisfy $\hat{E}$. A **quantitative variety** is a category equal to $\mathbf{QAlg}(\Sigma, \hat{E})$ for some class of quantitative equations $\hat{E}$.[351]

[351] We will sometimes simply say variety with the `knowldege` link going to this definition.

There are many forgetful functors obtained by composing the forgetful functors from $\mathbf{QAlg}(\Sigma)$ with the inclusion functor $\mathbf{QAlg}(\Sigma, \hat{E}) \to \mathbf{QAlg}(\Sigma)$:

- $U : \mathbf{QAlg}(\Sigma, \hat{E}) \to \mathbf{Set} = \mathbf{QAlg}(\Sigma, \hat{E}) \to \mathbf{QAlg}(\Sigma) \xrightarrow{U} \mathbf{Set}$
- $U : \mathbf{QAlg}(\Sigma, \hat{E}) \to \mathbf{Alg}(\Sigma) = \mathbf{QAlg}(\Sigma, \hat{E}) \to \mathbf{QAlg}(\Sigma) \xrightarrow{U} \mathbf{Alg}(\Sigma)$
- $U : \mathbf{QAlg}(\Sigma, \hat{E}) \to \mathbf{GMet} = \mathbf{QAlg}(\Sigma, \hat{E}) \to \mathbf{QAlg}(\Sigma) \xrightarrow{U} \mathbf{GMet}$
- $U : \mathbf{QAlg}(\Sigma, \hat{E}) \to \mathsf{L}\mathbf{Spa} = \mathbf{QAlg}(\Sigma, \hat{E}) \to \mathbf{QAlg}(\Sigma) \xrightarrow{U} \mathsf{L}\mathbf{Spa}$

Remark 3.15. Compared to the usage of the term variety in the literature (e.g. [MPP17, Adá22, ADV23]), our quantitative varieties are more general, even when $\mathbf{GMet} = \mathbf{Met}$. First, we do not constrain our operations to be interpreted as nonexpansive maps from the product as the other authors do. Second, we do not restrict the size of the context of the equations in $\hat{E}$ as is done in loc. cit.[352]

[352] Their restrictions are subtler than just putting an upper bound on the cardinality of the underlying set of the context.

Examples 3.16. 1. With $\Sigma = \{\mathsf{p} : 0\}$, we now have a lot more varieties than we had in Example 1.20. Even restricting to a discrete context, we have the following quantitative equations where ε ranges over L:[353]

[353] The first row comes from the classical case, and the second row replaces equality with equality up to ε ($=_\varepsilon$). The only difference being that $\mathsf{p} =_\varepsilon x$ and $x =_\varepsilon \mathsf{p}$ are not equivalent, so we need two distinct equations.

$$\begin{array}{lllll} \vdash \mathsf{p} = \mathsf{p} & x \vdash x = x & x \vdash \mathsf{p} = x & & x, y \vdash x = y \\ \vdash \mathsf{p} =_\varepsilon \mathsf{p} & x \vdash x =_\varepsilon x & x \vdash \mathsf{p} =_\varepsilon x & x \vdash x =_\varepsilon \mathsf{p} & x, y \vdash x =_\varepsilon y \end{array}$$

The meaning of the first row does not change from Example 1.20, and the meaning of the second row can be inferred by replacing equality between terms with distance between terms. For example, $\vdash \mathsf{p} =_\varepsilon \mathsf{p}$ says that the self-distance of the interpretation of the constant p is at most ε. Classifying the quantitative varieties for this signature would require a lot more work than for the classical varieties.[354]

[354] Although I think it is feasible, tedious but feasible.

2. When $\Sigma = \emptyset$, we mentioned that the quantitative equations are those of Chapter 2, so $\mathbf{QAlg}(\emptyset, \hat{E})$ is the subcategory of L-spaces that satisfy $\hat{E}$. In particular, the category $\mathbf{GMet}$ is a quantitative variety as it equals $\mathbf{QAlg}(\emptyset, \hat{E}_{\mathbf{GMet}})$.

3. If $\hat{E}$ contains the equations in $E_{\mathbf{CA}}$ and the equations in (3.10), then $\mathbf{QAlg}(\Sigma_{\mathbf{CA}}, \hat{E})$ is the category of convex algebras equipped with a convex metric [MV20, Definition 30] and nonexpansive homomorphisms.

Definition 3.17 (Quantitative algebraic theory)**.** Given a class $\hat{E}$ of quantitative equations over Σ and L, the **quantitative algebraic theory** generated by $\hat{E}$, denoted by $\mathfrak{QTh}(\hat{E})$, is the class of quantitative equations that are satisfied in all $(\Sigma, \hat{E})$-algebras:[355]

$$\mathfrak{QTh}(\hat{E}) = \left\{\phi \mid \forall \hat{\mathbb{A}} \in \mathbf{QAlg}(\Sigma, \hat{E}), \hat{\mathbb{A}} \vDash \phi\right\}.$$

[355] Again $\mathfrak{QTh}(\hat{E})$ is never a set (recall Definition 1.21).

Equivalently, $\mathfrak{QTh}(\hat{E})$ contains the equations that are semantically entailed by $\hat{E}$,[356] namely $\phi \in \mathfrak{QTh}(\hat{E})$ if and only if

$$\forall \hat{\mathbb{A}} \in \mathbf{QAlg}(\Sigma), \quad \hat{\mathbb{A}} \vDash \hat{E} \implies \hat{\mathbb{A}} \vDash \phi. \tag{3.11}$$

[356] As in the non-quantitative case, $\mathfrak{QTh}(\hat{E})$ contains all of $\hat{E}$ but also many more equations like $x \vdash x = x$ or $x =_\varepsilon y \vdash x =_\varepsilon y$. Furthermore, $\mathfrak{QTh}(\hat{E})$ contains all the quantitative equations in $\hat{E}_{\mathbf{GMet}}$ because the underlying spaces of algebras in $\mathbf{QAlg}(\Sigma, \hat{E})$ belong to $\mathbf{GMet}$.

We will see in §3.2 how to find which quantitative equations are entailed by others.

We call a class of quantitative equations a quantitative algebraic theory if it is generated by some class $\hat{E}$.

We will see twice[357] that the algebraic reasoning we are used to from Chapter 1 is embedded in quantitative algebraic reasoning. In particular, Example 1.22 which showed some equations which belong to the algebraic theory of commutative monoids can be read *unchanged* to find quantitative equations that belong to the quantitative algebraic theory of commutative monoids. These are only about equality ($=$), so let us give another example.

[357] In Examples 3.56 and 3.57.

Example 3.18. We mentioned in Example 3.12 that the equations for convexity (3.10) are *stronger* than the equations for nonexpansiveness with respect to the product distance (3.7). Formally what this means is that if $\hat{E}$ contains (3.10), then the interpretation of $+_p$ in a $(\Sigma_{\mathbf{CA}}, \hat{E})$-algebra $\hat{\mathbb{A}}$ will be a nonexpansive map $\mathbf{A} \times \mathbf{A} \to \mathbf{A}$, hence $\hat{\mathbb{A}}$ will satisfy (3.7). Concisely, the equations of (3.7) belong to $\mathfrak{QTh}(\hat{E})$.

Free Quantitative Algebras

We turn to the construction of free algebras, and we start with a simple example.

Example 3.19 (Free metric)**.** We already have some intuitions about terms and equations from Example 1.23, thus we consider an empty signature in order to focus on the new contexts and quantities. For $\hat{E}$, let us take the set of equations defining a metric space (with $\mathsf{L} = [0,1]$),[358] so that $\mathbf{QAlg}(\varnothing, \hat{E}) = \mathbf{Met}$.

[358] As a reminder, $\hat{E}$ contains

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

Now we wonder, given an L-space $\mathbf{X}$, what is the free metric space on it? Rehashing Definition 1.38, we want to find a metric space $F\mathbf{X}$ and a nonexpansive map $\eta : \mathbf{X} \to F\mathbf{X}$ such that any nonexpansive map from $\mathbf{X}$ to a metric space $\mathbf{A}$ factors through η uniquely. Of course, if $\mathbf{X}$ is already a metric space, then taking $F\mathbf{X} = \mathbf{X}$ and $\eta = \mathrm{id}_{\mathbf{X}}$ works. Otherwise, we can look at what prevents $d_{\mathbf{X}}$ from being a metric.

For instance, if $\mathbf{X}$ does not satisfy $\vdash x =_0 x$, it means there is some $x \in X$ such that $d_{\mathbf{X}}(x,x) > 0$. Inside $F\mathbf{X}$, we know that the distance between $\eta(x)$ and $\eta(x)$ must be 0. Note that if $\mathbf{A}$ is a metric space and $f : \mathbf{X} \to \mathbf{A}$ is nonexpansive, we know that $d_{\mathbf{A}}(f(x), f(x)) = 0$ too, so sending $\eta(x)$ to $f(x)$ will not be a problem.

For a second example, suppose $d_{\mathbf{X}}$ is not symmetric, without loss of generality $d_{\mathbf{X}}(x,y) < d_{\mathbf{X}}(y,x)$ for some $x, y \in \mathbf{X}$. We know that $d_{F\mathbf{X}}(\eta(x), \eta(y)) = d_{F\mathbf{X}}(\eta(y), \eta(x))$, but what value should it be? To ensure that η is nonexpansive, this value must be at most $d_{\mathbf{X}}(x,y)$, but why not smaller? If this lack of symmetry is the only thing preventing $d_{\mathbf{X}}$ from being a metric (i.e. defining d' everywhere like $d_{\mathbf{X}}$ except $d'(x,y) = d'(y,x)$ yields a metric), we cannot make $d_{F\mathbf{X}}(x,y)$ smaller, because the identity function id_X would be a nonexpansive map $\mathbf{X} \to (X, d')$ that does not factor through η (since $d'(x,y) > d_{F\mathbf{X}}(\eta(x), \eta(y))$). In fact, you can check that $F\mathbf{X} = (X, d')$ with $\eta = \mathrm{id}_X$ is the free metric space on $\mathbf{X}$ because our definition of d' fixed the only problem with $d_{\mathbf{X}}$.

In general, for any $x, y \in X$, we want $d_{F\mathbf{X}}(\eta(x), \eta(y))$ to be as large as possible while guaranteeing that $d_{F\mathbf{X}}$ is a metric and η is nonexpansive, but it is not always that simple. The complexity comes from the possible interactions between different equations in $\hat{E}$. Say you have $d_{\mathbf{X}}(x,z) > d_{\mathbf{X}}(x,y) + d_{\mathbf{X}}(y,z)$ so the triangle inequality does not hold, hence you try to fix this by lowering $d_{F\mathbf{X}}(\eta x, \eta z)$ down exactly to

$d_{F\mathbf{X}}(\eta x, \eta y) + d_{F\mathbf{X}}(\eta y, \eta z)$.[359] Then, to ensure symmetry, you need to lower $d_{F\mathbf{X}}(z, x)$ down to that same value, but after that you may need to lower $d_{F\mathbf{X}}(x, y)$ so that it is not bigger than the new value of $d_{F\mathbf{X}}(y, z) + d_{F\mathbf{X}}(z, x)$. In the end, you can end up back with $d_{F\mathbf{X}}(x, z) > d_{F\mathbf{X}}(x, y) + d_{F\mathbf{X}}(y, z)$, so you have to do another round of fixes.

Intuitively, $F\mathbf{X}$ is the space you obtain by iterating this process (possibly for infinitely many steps) and looking at the limit. We will give a rigorous description in the case of a more general signature,[360] but we want to point out now that this process does not deal only with distances, it can also force some equations. For example, if $d_{\mathbf{X}}(x, y) = 0$ with $x \neq y$ at the start, you will end up with $\eta(x) = \eta(y)$ inside $F\mathbf{X}$.

[359] Let us not write η each time for better readability, this is a bit informal as we will see that η is not necessarily injective.

[360] This is the construction of free quantitative algebras that starts in the next paragraph.

Fix a class $\hat{E}$ of quantitative equations over Σ and L. For any generalized metric space $\mathbf{X}$, we can define a binary relation $\equiv_{\hat{E}}$ and an L-relation $d_{\hat{E}}$ on Σ-terms as follows:[361] for any $s, t \in \mathcal{T}_\Sigma X$,

$$s \equiv_{\hat{E}} t \iff \mathbf{X} \vdash s = t \in \mathfrak{QTh}(\hat{E}) \text{ and } d_{\hat{E}}(s, t) = \inf\{\varepsilon \mid \mathbf{X} \vdash s =_\varepsilon t \in \mathfrak{QTh}(\hat{E})\}. \quad (3.12)$$

The definition of $\equiv_{\hat{E}}$ is completely analogous to what we did in the non-quantitative case (1.21). The definition of $d_{\hat{E}}$ is new but it also looks like how we defined an L-relation from an L-structure in Proposition 2.20. In fact, we can also prove a counterpart to (2.8), giving us an equivalent definition of $d_{\hat{E}}$: for any $s, t \in \mathcal{T}_\Sigma X$ and $\varepsilon \in \mathsf{L}$,[362]

$$d_{\hat{E}}(s, t) \leq \varepsilon \iff \mathbf{X} \vdash s =_\varepsilon t \in \mathfrak{QTh}(\hat{E}). \quad (3.13)$$

[361] The notation for $\equiv_{\hat{E}}$ and $d_{\hat{E}}$ should really depend on the space $\mathbf{X}$, but we prefer to omit this for better readability.

[362] In words, $d_{\hat{E}}$ assigns a distance below ε to s and t if and only if their interpretations in each $(\Sigma, \hat{E})$-algebras are always at a distance below ε.

Proof of (3.13). ($\Leftarrow$) holds directly by definition of infimum. For ($\Rightarrow$), we need to show that any $(\Sigma, \hat{E})$-algebra satisfies $\mathbf{X} \vdash s =_\varepsilon t$. Let $\hat{\mathbb{A}} \in \mathbf{QAlg}(\Sigma, \hat{E})$ and $\hat{\imath} : \mathbf{X} \to \mathbf{A}$ be a nonexpansive assignment. We know that for every δ such that $\mathbf{X} \vdash s =_\delta t \in \mathfrak{QTh}(\hat{E})$, $d_{\mathbf{A}}(\llbracket s \rrbracket_A^{\hat{\imath}}, \llbracket t \rrbracket_A^{\hat{\imath}}) \leq \delta$, thus

$$d_{\mathbf{A}}(\llbracket s \rrbracket_A^{\hat{\imath}}, \llbracket t \rrbracket_A^{\hat{\imath}}) \leq \inf\{\delta \mid \mathbf{X} \vdash s =_\delta t \in \mathfrak{QTh}(\hat{E})\} = d_{\hat{E}}(s, t) \leq \varepsilon.$$

We conclude that $\hat{\mathbb{A}} \vDash^{\hat{\imath}} \mathbf{X} \vdash s =_\varepsilon t$, and we are done since $\hat{\mathbb{A}}$ and $\hat{\imath}$ were arbitrary. $\square$

When we were not dealing with distances, we only had to prove that the relation $\equiv_E$ defined between terms was a congruence (Lemma 1.24), and then we were able to construct the term algebra by quotienting the set of terms and interpreting the operation symbols syntactically. Here we have to prove a bit more, namely that $d_{\hat{E}}$ is invariant under $\equiv_{\hat{E}}$ so the L-relation restricts to the quotient, and that the resulting L-space is a generalized metric space.

Let us decompose this in several small lemmas. We also collect here some more lemmas that look similar, many of which will be part of the proof of soundness when we introduce quantitative equational logic.[363] Let $\mathbf{X} \in \mathsf{L}\mathbf{Spa}$ and $\hat{\mathbb{A}} \in \mathbf{QAlg}(\Sigma)$ be universally quantified in all these lemmas.

First, Lemmas 3.20–3.23 say that $\equiv_{\hat{E}}$ is an equivalence relation and a congruence.[364]

[363] We were less explicit back then, but that is what happened with Lemma 1.24 and soundness of equational logic.

[364] The proofs are exactly the same as for Lemma 1.24 because $\equiv_{\hat{E}}$ does not involve distances.

Lemma 3.20. *For any $t \in \mathcal{T}_\Sigma X$, $\hat{\mathbb{A}}$ satisfies $\mathbf{X} \vdash t = t$.*

Proof. Obviously, $[\![t]\!]_A^{\hat{\imath}} = [\![t]\!]_A^{\hat{\imath}}$ holds for all $\hat{\imath} : \mathbf{X} \to \mathbf{A}$. □

Lemma 3.21. *For any $s, t \in \mathcal{T}_\Sigma X$, if $\hat{\mathbb{A}}$ satisfies $\mathbf{X} \vdash s = t$, then $\hat{\mathbb{A}}$ satisfies $\mathbf{X} \vdash t = s$.*

Proof. If $[\![s]\!]_A^{\hat{\imath}} = [\![t]\!]_A^{\hat{\imath}}$ holds for all $\hat{\imath}$, then $[\![t]\!]_A^{\hat{\imath}} = [\![s]\!]_A^{\hat{\imath}}$ holds too. □

Lemma 3.22. *For any $s, t, u \in \mathcal{T}_\Sigma X$, if $\hat{\mathbb{A}}$ satisfies $\mathbf{X} \vdash s = t$ and $\mathbf{X} \vdash t = u$, then $\hat{\mathbb{A}}$ satisfies $\mathbf{X} \vdash s = u$.*

Proof. If $[\![s]\!]_A^{\hat{\imath}} = [\![t]\!]_A^{\hat{\imath}}$ and $[\![t]\!]_A^{\hat{\imath}} = [\![u]\!]_A^{\hat{\imath}}$ holds for all $\hat{\imath}$, then $[\![s]\!]_A^{\hat{\imath}} = [\![u]\!]_A^{\hat{\imath}}$ holds too. □

Lemma 3.23. *For any $\mathsf{op} : n \in \Sigma$, $s_1, \dots, s_n, t_1, \dots, t_n \in \mathcal{T}_\Sigma X$, if $\hat{\mathbb{A}}$ satisfies $\mathbf{X} \vdash s_i = t_i$ for all $1 \le i \le n$, then $\hat{\mathbb{A}}$ satisfies $\mathbf{X} \vdash \mathsf{op}(s_1, \dots, s_n) = \mathsf{op}(t_1, \dots, t_n)$.*

Proof. For any assignment $\hat{\imath} : \mathbf{X} \to \mathbf{A}$, we have $[\![s_i]\!]_A^{\hat{\imath}} = [\![t_i]\!]_A^{\hat{\imath}}$ for all i. Hence,

$$\begin{aligned} [\![\mathsf{op}(s_1, \dots, s_n)]\!]_A^{\hat{\imath}} &= [\![\mathsf{op}]\!]_A([\![s_1]\!]_A^{\hat{\imath}}, \dots, [\![s_n]\!]_A^{\hat{\imath}}) && \text{by (1.8)} \\ &= [\![\mathsf{op}]\!]_A([\![t_1]\!]_A^{\hat{\imath}}, \dots, [\![t_n]\!]_A^{\hat{\imath}}) && \forall i, [\![s_i]\!]_A^{\hat{\imath}} = [\![t_i]\!]_A^{\hat{\imath}} \\ &= [\![\mathsf{op}(s_1, \dots, s_n)]\!]_A^{\hat{\imath}}. && \text{by (1.8)} \end{aligned}$$ □

Lemmas 3.24 and 3.25 mean that $d_{\hat{E}}$ is well-defined on equivalence classes of $\equiv_{\hat{E}}$, namely, $d_{\hat{E}}(s, t) = d_{\hat{E}}(s', t')$ whenever $s \equiv_{\hat{E}} s'$ and $t \equiv_{\hat{E}} t'$.[365]

[365] By Lemmas 3.21 and 3.24, if $t \equiv_{\hat{E}} t'$, then

$$\mathbf{X} \vdash s =_\varepsilon t \iff \mathbf{X} \vdash s =_\varepsilon t.$$

By Lemmas 3.21 and 3.25, if $s \equiv_{\hat{E}} s'$, then

$$\mathbf{X} \vdash s =_\varepsilon t' \iff \mathbf{X} \vdash s' =_\varepsilon t'.$$

Combining these with (3.13), we get

$$d_{\hat{E}}(s, t) \le \varepsilon \iff d_{\hat{E}}(s', t') \le \varepsilon,$$

for all $\varepsilon \in \mathsf{L}$, and we conclude $d_{\hat{E}}(s, t) = d_{\hat{E}}(s', t')$.

Lemma 3.24. *For any $s, t, t' \in \mathcal{T}_\Sigma X$ and $\varepsilon \in \mathsf{L}$, if $\hat{\mathbb{A}}$ satisfies $\mathbf{X} \vdash s =_\varepsilon t$ and $\mathbf{X} \vdash t = t'$, then $\hat{\mathbb{A}}$ satisfies $\mathbf{X} \vdash s =_\varepsilon t'$.*

Proof. For any $\hat{\imath} : \mathbf{X} \to \mathbf{A}$, we have $d_{\mathbf{A}}([\![s]\!]_A^{\hat{\imath}}, [\![t]\!]_A^{\hat{\imath}}) \le \varepsilon$ and $[\![t]\!]_A^{\hat{\imath}} = [\![t']\!]_A^{\hat{\imath}}$, thus

$$d_{\mathbf{A}}([\![s]\!]_A^{\hat{\imath}}, [\![t']\!]_A^{\hat{\imath}}) = d_{\mathbf{A}}([\![s]\!]_A^{\hat{\imath}}, [\![t]\!]_A^{\hat{\imath}}) \le \varepsilon.$$ □

Lemma 3.25. *For any $s, s', t \in \mathcal{T}_\Sigma X$ and $\varepsilon \in \mathsf{L}$, if $\hat{\mathbb{A}}$ satisfies $\mathbf{X} \vdash s =_\varepsilon t$ and $\mathbf{X} \vdash s = s'$, then $\hat{\mathbb{A}}$ satisfies $\mathbf{X} \vdash s' =_\varepsilon t$.*

Proof. Symmetric argument to the previous proof. □

Lemmas 3.26–3.29 will correspond to other rules in quantitative equational logic, and they will be explained in more details in §3.2.

Lemma 3.26. *For any $s, t \in \mathcal{T}_\Sigma X$, $\hat{\mathbb{A}}$ satisfies $\mathbf{X} \vdash s =_\top t$.*

Proof. By definition of $\top$ (the supremum of all L), for any $\hat{\imath}$, $d_{\mathbf{A}}([\![s]\!]_A^{\hat{\imath}}, [\![t]\!]_A^{\hat{\imath}}) \le \top$. □

Lemma 3.27. *For any $x, x' \in X$, if $d_{\mathbf{X}}(x, x') = \varepsilon$, then $\hat{\mathbb{A}}$ satisfies $\mathbf{X} \vdash x =_\varepsilon x'$.*

Proof. For any nonexpansive $\hat{\imath} : \mathbf{X} \to \mathbf{A}$, we have[366]

[366] The equation holds by definition of $[\![-]\!]_A^{\hat{\imath}}$ on variables, and the inequality holds by definition of nonexpansiveness.

$$d_{\mathbf{A}}([\![x]\!]_A^{\hat{\imath}}, [\![x']\!]_A^{\hat{\imath}}) = d_{\mathbf{A}}(\hat{\imath}(x), \hat{\imath}(x')) \le d_{\mathbf{X}}(x, x') = \varepsilon.$$ □

Lemma 3.28. *For any $s, t \in \mathcal{T}_\Sigma X$ and $\varepsilon, \varepsilon' \in \mathsf{L}$, if $\hat{\mathbb{A}}$ satisfies $\mathbf{X} \vdash s =_\varepsilon t$ and $\varepsilon \le \varepsilon'$, then $\hat{\mathbb{A}}$ satisfies $\mathbf{X} \vdash s =_{\varepsilon'} t$.*[367]

[367] In words, if the interpretations of s and t are at distance at most ε, then they are also at distance at most ε' when $\varepsilon \le \varepsilon'$.

Proof. For any $\hat{\imath} : \mathbf{X} \to \mathbf{A}$, we have $d_{\mathbf{A}}(\llbracket s \rrbracket_A^{\hat{\imath}}, \llbracket t \rrbracket_A^{\hat{\imath}}) \leq \varepsilon \leq \varepsilon'$. □

Lemma 3.29. *For any $s, t \in \mathcal{T}_\Sigma X$ and $\{\varepsilon_i\}_{i \in I} \subseteq \mathsf{L}$, if $\hat{\mathbb{A}}$ satisfies $\mathbf{X} \vdash s =_{\varepsilon_i} t$ for all $i \in I$, then $\hat{\mathbb{A}}$ satisfies $\mathbf{X} \vdash s =_\varepsilon t$ with $\varepsilon = \inf_{i \in I} \varepsilon_i$.*

Proof. For any $\hat{\imath}$ and for all $i \in I$, we have $d_{\mathbf{A}}(\llbracket s \rrbracket_A^{\hat{\imath}}, \llbracket t \rrbracket_A^{\hat{\imath}}) \leq \varepsilon_i$ by hypothesis. By definition of infimum, this means $d_{\mathbf{A}}(\llbracket s \rrbracket_A^{\hat{\imath}}, \llbracket t \rrbracket_A^{\hat{\imath}}) \leq \inf_{i \in I} \varepsilon_i = \varepsilon$. □

This shall take care of all except two rules in quantitative equational logic which we will get to in no time. The following result is a generalization of Lemma 2.28, and it morally says that $\mathcal{T}_\Sigma f$ is well-defined and nonexpansive when f is nonexpansive.

Lemma 3.30. *Let $f : \mathbf{X} \to \mathbf{Y}$ be a nonexpansive map. If $\mathbf{A}$ satisfies $\mathbf{X} \vdash s = t$ (resp. $\mathbf{X} \vdash s =_\varepsilon t$), then $\mathbf{A}$ satisfies $\mathbf{Y} \vdash \mathcal{T}_\Sigma f(s) = \mathcal{T}_\Sigma f(t)$ (resp. $\mathbf{Y} \vdash \mathcal{T}_\Sigma f(s) =_\varepsilon \mathcal{T}_\Sigma f(t)$).*[368]

[368] Note that when s and t are variables, we get back Lemma 2.28.

Proof. Any nonexpansive assignment $\hat{\imath} : \mathbf{Y} \to \mathbf{A}$, yields a nonexpansive assignment $\hat{\imath} \circ f : \mathbf{X} \to \mathbf{A}$. Moreover, by functoriality of $\mathcal{T}_\Sigma$, we have

$$\llbracket - \rrbracket_A^{\hat{\imath} \circ f} \stackrel{(1.9)}{=} \llbracket - \rrbracket_A \circ \mathcal{T}_\Sigma(\hat{\imath} \circ f) = \llbracket - \rrbracket_A \circ \mathcal{T}_\Sigma \hat{\imath} \circ \mathcal{T}_\Sigma f = \llbracket \mathcal{T}_\Sigma f(-) \rrbracket_A^{\hat{\imath}}.$$

By hypothesis, we have

$$\mathbf{A} \vDash^{\hat{\imath} \circ f} \mathbf{X} \vdash s = t \qquad (\text{resp. } \mathbf{A} \vDash^{\hat{\imath} \circ f} \mathbf{X} \vdash s =_\varepsilon t),$$

which means

$$\begin{gathered} \llbracket \mathcal{T}_\Sigma f(s) \rrbracket_A^{\hat{\imath}} = \llbracket s \rrbracket_A^{\hat{\imath} \circ f} = \llbracket t \rrbracket_A^{\hat{\imath} \circ f} = \llbracket \mathcal{T}_\Sigma f(t) \rrbracket_A^{\hat{\imath}} \\ \text{resp. } d_{\mathbf{A}}(\llbracket \mathcal{T}_\Sigma f(s) \rrbracket_A^{\hat{\imath}}, \llbracket \mathcal{T}_\Sigma f(t) \rrbracket_A^{\hat{\imath}}) = d_{\mathbf{A}}(\llbracket s \rrbracket_A^{\hat{\imath} \circ f}, \llbracket t \rrbracket_A^{\hat{\imath} \circ f}) \leq \varepsilon. \end{gathered}$$

Thus, we conclude

$$\mathbf{A} \vDash^{\hat{\imath}} \mathbf{Y} \vdash \mathcal{T}_\Sigma f(s) = \mathcal{T}_\Sigma f(t) \qquad (\text{resp. } \mathbf{A} \vDash^{\hat{\imath}} \mathbf{Y} \vdash \mathcal{T}_\Sigma f(s) =_\varepsilon \mathcal{T}_\Sigma f(t)).$$ □

Let us end our list of small results with Lemmas 3.31–3.33 which are for later.

Lemma 3.31. *For any $s, t \in \mathcal{T}_\Sigma X$ if $\hat{\mathbb{A}}$ satisfies $\mathbf{X}_\top \vdash s = t$, then $\hat{\mathbb{A}}$ satisfies $\mathbf{X} \vdash s = t$, and for any $\varepsilon \in \mathsf{L}$, if $\hat{\mathbb{A}}$ satisfies $\mathbf{X}_\top \vdash s =_\varepsilon t$, then $\hat{\mathbb{A}}$ satisfies $\mathbf{X} \vdash s =_\varepsilon t$.*[369]

[369] In words, if $\hat{\mathbb{A}}$ satisfies an equation where the context is the discrete space on X, then $\hat{\mathbb{A}}$ satisfies that same equation with the context replaced by any other L-space on X. This is also a special case of Lemma 3.30 where $f : \mathbf{X}_\top \to \mathbf{X}$ is the identity map.

Proof. For any nonexpansive assignment $\hat{\imath} : \mathbf{X} \to \mathbf{A}$, you can pre-compose it with $\mathrm{id}_X : \mathbf{X}_\top \to \mathbf{X}$ (which is nonexpansive) without changing the interpretation of terms: $\llbracket s \rrbracket_A^{\hat{\imath}} = \llbracket s \rrbracket_A^{\hat{\imath} \circ \mathrm{id}_X}$. By hypothesis, we know that $\hat{\mathbb{A}}$ satisfies $s = t$ (resp. $s =_\varepsilon t$) under the nonexpansive assignment $\hat{\imath} \circ \mathrm{id}_X : \mathbf{X}_\top \to \mathbf{A}$, and we conclude $\hat{\mathbb{A}}$ also satisfies $s = t$ (resp. $s =_\varepsilon t$) under the assignment $\hat{\imath}$. □

Lemma 3.32. *For any $s, t \in \mathcal{T}_\Sigma X$, if $\mathbb{A}$ satisfies $X \vdash s = t$, then $\hat{\mathbb{A}}$ satisfies $\mathbf{X} \vdash s = t$.*[370]

[370] In words, if the underlying (not quantitative) algebra satisfies an equation, then so does the quantitative algebra where the context can be endowed with any L-relation.

Proof. Any nonexpansive assignment $\hat{\imath} : \mathbf{X} \to \mathbf{A}$ is in particular an assignment $\hat{\imath} : X \to A$, thus $\llbracket s \rrbracket_A^{\hat{\imath}} = \llbracket t \rrbracket_A^{\hat{\imath}}$ hold by hypothesis that $\mathbb{A}$ satisfies $X \vdash s = t$. □

Lemma 3.33. *For any $s, t \in \mathcal{T}_\Sigma X$, if $\hat{\mathbb{A}}$ satisfies $\mathbf{X}_\top \vdash s = t$, then $\mathbb{A}$ satisfies $X \vdash s = t$.*[371]

[371] Combining Lemmas 3.32 and 3.33, we find

$$\mathbb{A} \vDash X \vdash s = t \iff \hat{\mathbb{A}} \vDash \mathbf{X}_\top \vdash s = t. \tag{3.14}$$

This can be useful when comparing equational logic and quantitative equational logic in Example 3.57.

Proof. This follows by definition of the discrete space $\mathbf{X}_\top$. Indeed, any assignment $\iota : X \to A$ is the underlying function of a nonexpansive assignment $\hat{\iota} : \mathbf{X} \to \mathbf{A}$, and since $\hat{\mathbb{A}}$ satisfies $s = t$ under $\hat{\iota}$ by hypothesis, $\mathbb{A}$ satisfies $s = t$ under ι. □

We can now get back to the equality $\equiv_{\hat{E}}$ and distance $d_{\hat{E}}$ between terms, and define the underlying space of the quantitative term algebra.

Since $\equiv_{\hat{E}}$ is an equivalence relation for any $\mathbf{X}$, we can consider the set $\mathcal{T}_\Sigma X/{\equiv_{\hat{E}}}$ of **terms modulo** $\hat{E}$.[372] We denote with $[-]_{\hat{E}} : \mathcal{T}_\Sigma X \to \mathcal{T}_\Sigma X/{\equiv_{\hat{E}}}$ the canonical quotient map, and by Lemmas 3.24 and 3.25, we can define an L-relation on terms modulo $\hat{E}$ by factoring $d_{\hat{E}}$ through $[-]_{\hat{E}}$. We obtain the L-relation $d_{\hat{E}}$ as the unique function making the triangle below commute.[373]

$$\begin{tikzcd}
\mathcal{T}_\Sigma X \times \mathcal{T}_\Sigma X \arrow[r, "d_{\hat{E}}"] \arrow[d, "{[-]_{\hat{E}} \times [-]_{\hat{E}}}"'] & \mathsf{L} \\
\mathcal{T}_\Sigma X/{\equiv_{\hat{E}}} \times \mathcal{T}_\Sigma X/{\equiv_{\hat{E}}} \arrow[ur, dashed, "d_{\hat{E}}"']
\end{tikzcd} \tag{3.15}$$

We write $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$ for the resulting L-space $(\mathcal{T}_\Sigma X/{\equiv_{\hat{E}}}, d_{\hat{E}})$. We still have an alternative definition analog to (3.13) for the new L-relation d_E.[374]

$$d_{\hat{E}}([s]_{\hat{E}}, [t]_{\hat{E}}) \leq \varepsilon \iff \mathbf{X} \vdash s =_\varepsilon t \in \mathfrak{QTh}(\hat{E}). \tag{3.16}$$

This will be the carrier of the term algebra on $\mathbf{X}$, so we need to prove that $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$ belongs to **GMet**. We rely on the following generalization of Lemma 1.36. It essentially states that satisfaction of quantitative equations is preserved by substitutions that are nonexpansive. This result will also take care of the last two rules of quantitative equational logic.

Lemma 3.34. *Let* $\mathbf{Y}$ *be an* L*-space and* $\sigma : Y \to \mathcal{T}_\Sigma X$ *be an assignment such that*[375]

$$\forall y, y' \in Y, \quad \mathbf{X} \vdash \sigma(y) =_{d_{\mathbf{Y}}(y,y')} \sigma(y') \in \mathfrak{QTh}(\hat{E}), \tag{3.17}$$

and $\hat{\mathbb{A}}$ *a* $(\Sigma, \hat{E})$*-algebra. If* $\hat{\mathbb{A}}$ *satisfies* $\mathbf{Y} \vdash s = t$ *(resp.* $\mathbf{Y} \vdash s =_\varepsilon t$*), then it also satisfies* $\mathbf{X} \vdash \sigma^*(s) = \sigma^*(t)$ *(resp.* $\mathbf{X} \vdash \sigma^*(s) =_\varepsilon \sigma^*(t)$*).*

Proof. Let $\hat{\iota} : \mathbf{X} \to \mathbf{A}$ be a nonexpansive assignment, we need to show $[\![\sigma^*(s)]\!]_A^{\hat{\iota}} = [\![\sigma^*(t)]\!]_A^{\hat{\iota}}$ (resp. $d_{\mathbf{A}}([\![\sigma^*(s)]\!]_A^{\hat{\iota}}, [\![\sigma^*(t)]\!]_A^{\hat{\iota}}) \leq \varepsilon$). Just like in Lemma 1.36, we define the assignment $\hat{\iota}_\sigma : Y \to A$ that sends $y \in Y$ to $[\![\sigma(y)]\!]_A^{\hat{\iota}}$, and we had already proven $[\![-]\!]_A^{\hat{\iota}_\sigma} = [\![\sigma^*(-)]\!]_A^{\hat{\iota}}$. Now, it is enough to show $\hat{\iota}_\sigma$ is nonexpansive $\mathbf{Y} \to \mathbf{A}$[376] and the lemma will follow because by hypothesis, $[\![s]\!]_A^{\hat{\iota}_\sigma} = [\![t]\!]_A^{\hat{\iota}_\sigma}$ (reps. $d_{\mathbf{A}}([\![s]\!]_A^{\hat{\iota}_\sigma}, [\![t]\!]_A^{\hat{\iota}_\sigma}) \leq \varepsilon$).

For any $y, y' \in Y$, we have

$$d_{\mathbf{A}}(\hat{\iota}_\sigma(y), \hat{\iota}_\sigma(y')) = d_{\mathbf{A}}([\![\sigma(y)]\!]_A^{\hat{\iota}}, [\![\sigma(y')]\!]_A^{\hat{\iota}}) \leq d_{\mathbf{Y}}(y, y'),$$

where the equation holds by definition of $\hat{\iota}_\sigma$, and the inequality holds because $\hat{\mathbb{A}}$ belongs to $\mathbf{QAlg}(\Sigma, \hat{E})$ and hence satisfies $\mathbf{X} \vdash \sigma(y) =_{d_{\mathbf{Y}}(y,y')} \sigma(y') \in \mathfrak{QTh}(\hat{E})$ (in particular under the nonexpansive assignment $\hat{\iota}$). Hence $\hat{\iota}_\sigma$ is nonexpansive. □

Lemma 3.35. *For any* L*-space* $\mathbf{X}$ *and any quantitative equation* $\phi \in \hat{E}_{\mathbf{GMet}}$, $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \vDash \phi$.

[372] Keep in mind that for different L-relations on X, we may get different equivalence relations on $\mathcal{T}_\Sigma X$.

[373] We used the same symbol, because the first $d_{\hat{E}}$ was only used to define this new $d_{\hat{E}}$.

[374] In particular, the quotient map is nonexpansive:

$$[-]_{\hat{E}} : (\mathcal{T}_\Sigma X, d_{\hat{E}}) \to \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}.$$

[375] By combining (3.17) with (3.13) we find that σ is a nonexpansive map $\mathbf{Y} \to (\mathcal{T}_\Sigma X, d_{\hat{E}})$, and any such nonexpansive map satisfies (3.17). We explicitly write (3.17) to better emulate the corresponding rules in quantitative equational logic.

[376] Something we did not have to do in the non-quantitative case.

Proof. We mentioned in Footnote 356 that $\phi \in \mathfrak{QTh}(\hat{E})$ because the carriers of $(\Sigma, \hat{E})$-algebras are generalized metric spaces, so any $(\Sigma, \hat{E})$-algebra $\hat{\mathbb{A}}$ satisfies it.

Let $\hat{\imath} : \mathbf{Y} \to \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$ is a nonexpansive assignment. By the axiom of choice,[377] there is a function $\sigma : Y \to \mathcal{T}_\Sigma X$ satisfying $[\sigma(y)]_{\hat{E}} = \hat{\imath}(y)$ for all $y \in Y$. This assignment satisfies (3.17) because for all $y, y' \in Y$, (3.16) yields

$$d_{\hat{E}}([\sigma(y)]_{\hat{E}}, [\sigma(y')]_{\hat{E}}) \leq d_{\mathbf{Y}}(y, y') \overset{(3.16)}{\iff} \mathbf{X} \vdash \sigma(y) =_{d_{\mathbf{Y}}(y,y')} \sigma(y') \in \mathfrak{QTh}(\hat{E}),$$

and the L.H.S. holds because $\hat{\imath}$ is nonexpansive.

Therefore, if ϕ has the shape $\mathbf{Y} \vdash y = y'$ (resp. $\mathbf{Y} \vdash y =_\varepsilon y'$), by Lemma 3.34, all $(\Sigma, \hat{E})$-algebras satisfy $\mathbf{X} \vdash \sigma(y) = \sigma(y')$ (resp. $\mathbf{X} \vdash \sigma(y) =_\varepsilon \sigma(y')$). By definition of $\equiv_{\hat{E}}$ (resp. by definition of $d_{\hat{E}}$ (3.16)), we have

$$\hat{\imath}(y) = [\sigma(y)]_{\hat{E}} = [\sigma(y')]_{\hat{E}} = \hat{\imath}(y') \quad (\text{resp. } d_{\hat{E}}(\hat{\imath}(y), \hat{\imath}(y')) = d_{\hat{E}}([\sigma(y)]_{\hat{E}}, [\sigma(y')]_{\hat{E}}) \leq \varepsilon\ ),$$

which means $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$ satisfies ϕ under $\hat{\imath}$. Since $\hat{\imath}$ and ϕ were arbitrary, we conclude $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$ satisfies all of $\hat{E}_{\mathbf{GMet}}$, i.e. it is a generalized metric space. □

[377] Choice implies the quotient map $[-]_{\hat{E}}$ has a right inverse $r : \mathcal{T}_\Sigma X/{\equiv_{\hat{E}}} \to \mathcal{T}_\Sigma X$, and we set $\sigma = r \circ \hat{\imath}$.

As for **Set**, we obtain a functor $\widehat{\mathcal{T}}_{\Sigma,\hat{E}} : \mathbf{GMet} \to \mathbf{GMet}$[378] by setting $\widehat{\mathcal{T}}_{\Sigma,\hat{E}} f$ equal to the unique function making (3.18) commute. Concretely, we have $\widehat{\mathcal{T}}_{\Sigma,E} f([t]_{\hat{E}}) = [\mathcal{T}_\Sigma f(t)]_{\hat{E}}$ which is well-defined by one part of Lemma 3.30.

[378] In fact, we defined a functor $\mathbf{LSpa} \to \mathbf{GMet}$, but we are interested in its restriction to **GMet**.

$$\begin{tikzcd} \mathcal{T}_\Sigma X \arrow[r, "{[-]_{\hat{E}}}"] \arrow[d, "\mathcal{T}_\Sigma f"'] & \mathcal{T}_\Sigma X/{\equiv_{\hat{E}}} \arrow[d, dashed, "\widehat{\mathcal{T}}_{\Sigma,\hat{E}} f"] \\ \mathcal{T}_\Sigma Y \arrow[r, "{[-]_{\hat{E}}}"'] & \mathcal{T}_\Sigma Y/{\equiv_{\hat{E}}} \end{tikzcd} \tag{3.18}$$

Although we do have to check that $\widehat{\mathcal{T}}_{\Sigma,\hat{E}} f$ is nonexpansive whenever f is, and we use the other part of Lemma 3.30.

Lemma 3.36. *If $f : \mathbf{X} \to \mathbf{Y}$ is nonexpansive, then so is $\widehat{\mathcal{T}}_{\Sigma,\hat{E}} f : \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \to \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{Y}$.*

Proof. For any $s, t \in \mathcal{T}_\Sigma X$, we have

$$\begin{aligned} d_{\hat{E}}([s]_{\hat{E}}, [t]_{\hat{E}}) \leq \varepsilon &\iff \mathbf{X} \vdash s =_\varepsilon t \in \mathfrak{QTh}(\hat{E}) && \text{by (3.16)} \\ &\implies \mathbf{X} \vdash \mathcal{T}_\Sigma f(s) =_\varepsilon \mathcal{T}_\Sigma f(t) \in \mathfrak{QTh}(\hat{E}) && \text{Lemma 3.30} \\ &\iff d_{\hat{E}}([\mathcal{T}_\Sigma f(s)]_{\hat{E}}, [\mathcal{T}_\Sigma f(t)]_{\hat{E}}) \leq \varepsilon && \text{by (3.16)} \\ &\iff d_{\hat{E}}(\widehat{\mathcal{T}}_{\Sigma,\hat{E}} f[s]_{\hat{E}}, \widehat{\mathcal{T}}_{\Sigma,\hat{E}} f[t]_{\hat{E}}) \leq \varepsilon. && \text{by (3.18)} \end{aligned}$$

Therefore, $d_{\hat{E}}(\widehat{\mathcal{T}}_{\Sigma,\hat{E}} f[s]_{\hat{E}}, \widehat{\mathcal{T}}_{\Sigma,\hat{E}} f[t]_{\hat{E}}) \leq d_{\hat{E}}([s]_{\hat{E}}, [t]_{\hat{E}})$. □

We may now define the interpretation of operation symbols syntactically to obtain the quantitative term algebra.

Definition 3.37 (Quantitative term algebra, semantically)**.** The **quantitative term algebra** for $(\Sigma, \hat{E})$ on $\mathbf{X}$ is the quantitative Σ-algebra whose underlying space is $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$ and whose interpretation of $\mathsf{op} : n \in \Sigma$ is defined by[379]

$$[\![\mathsf{op}]\!]_{\widehat{\mathbb{T}}\mathbf{X}}([t_1]_{\hat{E}}, \ldots, [t_n]_{\hat{E}}) = [\mathsf{op}(t_1, \ldots, t_n)]_{\hat{E}}. \tag{3.19}$$

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

[379] This is well-defined by Lemma 3.23.

This should feel very familiar to what we had done in Definition 1.25.[380] In particular, we still have that $[-]_{\hat{E}}$ is a homomorphism from $\mathcal{T}_\Sigma X$ to the underlying algebra of $\widehat{\mathbb{T}}\mathbf{X}$,[381]namely, (3.20) commutes (recall Footnote 74).

$$\begin{tikzcd} \mathcal{T}_\Sigma\mathcal{T}_\Sigma X \arrow[r, "\mathcal{T}_\Sigma[-]_{\hat{E}}"] \arrow[d, "\mu^\Sigma_X"'] & \mathcal{T}_\Sigma\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[d, "[\![-]\!]_{\widehat{\mathbb{T}}\mathbf{X}}"] \\ \mathcal{T}_\Sigma X \arrow[r, "{[-]_{\hat{E}}}"'] & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \end{tikzcd} \tag{3.20}$$

While (3.20) is a diagram in **Set**, we write $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$ instead of the underlying set $\mathcal{T}_\Sigma X/\equiv_{\hat{E}}$ for better readability. We will keep this habit.

Your intuition for $[\![-]\!]_{\widehat{\mathbb{T}}\mathbf{X}}$ (the interpretation of arbitrary terms) should be exactly the same as the one for $[\![-]\!]_{\mathbb{T}X}$ in classical universal algebra: it takes a term in $\mathcal{T}_\Sigma\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$, replaces the leaves with a representative term, and gives back the equivalence class of the resulting term. We can also use it to define an analog to flattening.[382] For any space $\mathbf{X}$, let $\widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{X}}$ be the unique function making (3.21) commute.

$$\begin{tikzcd} \mathcal{T}_\Sigma\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[rr, "[\![-]\!]_{\widehat{\mathbb{T}}\mathbf{X}}"] \arrow[dr, "{[-]_{\hat{E}}}"'] & & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \\ & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[ur, dashed, "\widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{X}}"'] & \end{tikzcd} \tag{3.21}$$

Let us show that $\widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{X}}$ is nonexpansive and natural.

Lemma 3.38. *For any space $\mathbf{X}$, $\widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{X}}$ is a nonexpansive map $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \to \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$.*

Proof. Let $[s]_{\hat{E}}, [t]_{\hat{E}} \in \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$ be such that $d_{\hat{E}}([s]_{\hat{E}}, [t]_{\hat{E}}) \leq \varepsilon$. By (3.16), this means

$$\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \vdash s =_\varepsilon t \in \mathfrak{QTh}(\hat{E}), \tag{3.22}$$

namely, the distance between interpretations of s and t is bounded above by ε in all $(\Sigma, \hat{E})$-algebras. We need to show $d_{\hat{E}}(\widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{X}}([s]_{\hat{E}}), \widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{X}}([t]_{\hat{E}})) \leq \varepsilon$, or using (3.21),

$$d_{\hat{E}}([\![s]\!]_{\widehat{\mathbb{T}}\mathbf{X}}, [\![t]\!]_{\widehat{\mathbb{T}}\mathbf{X}}) \leq \varepsilon. \tag{3.23}$$

We want to use (3.16) again to reduce that inequality to a bound on distances between interpretations, but that requires choosing representatives for $[\![s]\!]_{\widehat{\mathbb{T}}\mathbf{X}}, [\![t]\!]_{\widehat{\mathbb{T}}\mathbf{X}} \in \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$.

Instead of choosing them naively, let $s', t' \in \mathcal{T}_\Sigma\mathcal{T}_\Sigma X$ be such that $\mathcal{T}_\Sigma[-]_{\hat{E}}(s') = s$ and $\mathcal{T}_\Sigma[-]_{\hat{E}}(t') = t$. In words, s' and t' are the same as s and t where equivalence classes at the leaves are replaced representative terms.[383] Commutativity of (3.20) implies $[\mu^\Sigma_X(s')]_{\hat{E}} = [\![s]\!]_{\widehat{\mathbb{T}}\mathbf{X}}$ and similarly for t. We can now use (3.16) to infer that proving (3.23) is equivalent to proving

$$\mathbf{X} \vdash \mu^\Sigma_X(s') =_\varepsilon \mu^\Sigma_X(t') \in \mathfrak{QTh}(\hat{E}). \tag{3.24}$$

This means we need to show that, for all $\hat{\mathbb{A}} \in \mathbf{QAlg}(\Sigma, \hat{E})$ and $\hat{\iota} : \mathbf{X} \to \mathbf{A}$, $d_{\mathbf{A}}([\![\mu^\Sigma_X(s')]\!]^{\hat{\iota}}_A, [\![\mu^\Sigma_X(t')]\!]^{\hat{\iota}}_A) \leq \varepsilon$.

[380] In fact, we can make the connection more precise, $\mathbb{T}X$ is constructed by quotienting $\mathcal{T}_\Sigma X$ by the congruence $\equiv_E$ and (the underlying algebra of) $\widehat{\mathbb{T}}\mathbf{X}$ by quotienting $\mathcal{T}_\Sigma X$ by the congruence $\equiv_{\hat{E}}$ (see Remark 1.26).

[381] Put $h = [-]_{\hat{E}}$ in (1.2) to get (3.19)

[382] Just as we did in (1.27).

[383] Since s and t have finitely many leaves, we are only doing finitely many choices of representatives.

We already know by (3.22) that for all $\hat{\sigma} : \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \to \mathbf{A}$, $d_{\mathbf{A}}([\![s]\!]_A^{\hat{\sigma}}, [\![t]\!]_A^{\hat{\sigma}}) \leq \varepsilon$, so it suffices to find, for each $\hat{\imath} : \mathbf{X} \to \mathbf{A}$, a nonexpansive assignment $\hat{\sigma}_{\hat{\imath}} : \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \to \mathbf{A}$ such that

$$[\![\mu_X^\Sigma(s')]\!]_A^{\hat{\imath}} = [\![s]\!]_A^{\hat{\sigma}_{\hat{\imath}}} \text{ and } [\![\mu_X^\Sigma(t')]\!]_A^{\hat{\imath}} = [\![t]\!]_A^{\hat{\sigma}_{\hat{\imath}}}. \tag{3.25}$$

We define $\hat{\sigma}_{\hat{\imath}} : \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \to \mathbf{A}$ to be the unique function making (3.26) commute.[384]

$$\begin{tikzcd} \mathcal{T}_\Sigma X \arrow[r, "\mathcal{T}_\Sigma \hat{\imath}"] \arrow[d, "{[-]_{\hat{E}}}"'] & \mathcal{T}_\Sigma A \arrow[d, "{[\![-]\!]_A}"] \\ \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[r, dashed, "\hat{\sigma}_{\hat{\imath}}"'] & A \end{tikzcd} \tag{3.26}$$

[384] It exists because $\hat{\mathbb{A}}$ satisfies all the equations in $\mathfrak{QTh}(\hat{E})$ so if $s \equiv_{\hat{E}} t$ then

$$[\![\mathcal{T}_\Sigma \hat{\imath}(s)]\!]_A \overset{(1.9)}{=} [\![s]\!]_A^{\hat{\imath}} = [\![t]\!]_A^{\hat{\imath}} \overset{(1.9)}{=} [\![\mathcal{T}_\Sigma \hat{\imath}(t)]\!]_A.$$

First, $\hat{\sigma}_{\hat{\imath}}$ is a nonexpansive map $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \to \mathbf{A}$ because for any $[u]_{\hat{E}}, [v]_{\hat{E}} \in \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$,

$$d_{\mathbf{A}}(\hat{\sigma}_{\hat{\imath}}[u]_{\hat{E}}, \hat{\sigma}_{\hat{\imath}}[v]_{\hat{E}}) \overset{(3.26)}{=} d_{\mathbf{A}}([\![\mathcal{T}_\Sigma \hat{\imath}(u)]\!]_A, [\![\mathcal{T}_\Sigma \hat{\imath}(v)]\!]_A) \overset{(1.9)}{=} d_{\mathbf{A}}([\![u]\!]_A^{\hat{\imath}}, [\![v]\!]_A^{\hat{\imath}}) \leq d_{\hat{E}}([u]_{\hat{E}}, [v]_{\hat{E}}),$$

where the inequality holds by definition of $d_{\hat{E}}$ and because $\hat{\mathbb{A}}$ satisfies all the equations in $\mathfrak{QTh}(\hat{E})$.

Second, we can prove that

$$[\![-]\!]_A^{\hat{\imath}} \circ \mu_X^\Sigma = [\![-]\!]_A^{\hat{\sigma}_{\hat{\imath}}} \circ \mathcal{T}_\Sigma[-]_{\hat{E}}, \tag{3.27}$$

which implies (3.25) holds (by applying both sides of (3.27) to s' and t'). We pave the following diagram.

$$\begin{tikzcd} \mathcal{T}_\Sigma\mathcal{T}_\Sigma X \arrow[rrr, "\mathcal{T}_\Sigma{[-]_{\hat{E}}}"] \arrow[dr, "\mathcal{T}_\Sigma\mathcal{T}_\Sigma\hat{\imath}"] \arrow[ddd, "\mu_X^\Sigma"'] & & & \mathcal{T}_\Sigma\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[dl, "\mathcal{T}_\Sigma\hat{\sigma}_{\hat{\imath}}"'] \arrow[ddd, "{[\![-]\!]_A^{\hat{\sigma}_{\hat{\imath}}}}"] \\ & \mathcal{T}_\Sigma X \arrow[r, "\mathcal{T}_\Sigma{[\![-]\!]_A}"'] \arrow[ur, phantom, "\text{(a)}"] \arrow[d, phantom, "\text{(b)}"] & \mathcal{T}_\Sigma A \arrow[ddr, "{[\![-]\!]_A}"'] \arrow[r, phantom, "\text{(c)}"] & {} \\ & {} & & \\ \mathcal{T}_\Sigma X \arrow[rrr, "{[\![-]\!]_A^{\hat{\imath}}}"'] & & & A \end{tikzcd} \tag{3.28}$$

Showing (3.28) commutes:

(a) Apply $\mathcal{T}_\Sigma$ to (3.26).

(b) By (1.14).

(c) By (1.9).

□

Lemma 3.39. *The family of maps $\widehat{\mu}_{\mathbf{X}}^{\Sigma,\hat{E}} : \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \to \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$ is natural in $\mathbf{X}$.*[385]

[385] We will (for posterity) reproduce the proof we did for Proposition 1.29, but it is important to note that nothing changes except the notation which now has lots of little hats.

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

$$\begin{tikzcd} \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[r, "\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\mathcal{T}}_{\Sigma,\hat{E}}f"] \arrow[d, "\widehat{\mu}_{\mathbf{X}}^{\Sigma,\hat{E}}"'] & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{Y} \arrow[d, "\widehat{\mu}_{\mathbf{Y}}^{\Sigma,\hat{E}}"] \\ \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[r, "\widehat{\mathcal{T}}_{\Sigma,\hat{E}}f"'] & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{Y} \end{tikzcd} \tag{3.29}$$

We can pave the following diagram.

$$
\begin{tikzcd}
\mathcal{T}_\Sigma\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[rr, "{[-]_{\hat{E}}}"] \arrow[drr, dashed, "\mathcal{T}_\Sigma\widehat{\mathcal{T}}_{\Sigma,\hat{E}}f"'] \arrow[dd, "{[-]_{\hat{E}}}"'] \arrow[ddrr, "{[\![-]\!]_{\widehat{\mathbb{T}}\mathbf{X}}}"'] & & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[rr, "\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\mathcal{T}}_{\Sigma,\hat{E}}f"] \arrow[d, phantom, "\text{(a)}"] & & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{Y} \arrow[dd, "\hat{\mu}^{\Sigma,\hat{E}}_{\mathbf{Y}}"] \\
\text{(b)} & \text{(c)} & \mathcal{T}_\Sigma\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{Y} \arrow[urr, "{[-]_{\hat{E}}}"'] \arrow[drr, "{[\![-]\!]_{\widehat{\mathbb{T}}\mathbf{Y}}}"'] & \text{(d)} & \\
\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[rr, "\hat{\mu}^{\Sigma,\hat{E}}_{\mathbf{X}}"'] & & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[rr, "\widehat{\mathcal{T}}_{\Sigma,\hat{E}}f"'] & & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{Y}
\end{tikzcd}
$$

All of (a), (b) and (d) commute by definition. In more details, (a) is an instance of (3.18) with $\mathbf{X}$ replaced by $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$, $\mathbf{Y}$ by $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{Y}$ and f by $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}f$, and both (b) and (d) are instances of (3.21). 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 $\mathcal{T}_\Sigma[-]_{\hat{E}}$ is surjective (i.e. epic) to conclude that the front face must also commute.[386]

$$
\begin{tikzcd}
\mathcal{T}_\Sigma\mathcal{T}_\Sigma X \arrow[rr, "\mathcal{T}_\Sigma\mathcal{T}_\Sigma f"] \arrow[dr, "{\mathcal{T}_\Sigma[-]_{\hat{E}}}"] \arrow[dd, "\mu^\Sigma_X"'] & & \mathcal{T}_\Sigma\mathcal{T}_\Sigma Y \arrow[dr, "{\mathcal{T}_\Sigma[-]_{\hat{E}}}"] \arrow[dd, "\mu^\Sigma_Y"' near start] & \\
& \mathcal{T}_\Sigma\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[rr, "\mathcal{T}_\Sigma\widehat{\mathcal{T}}_{\Sigma,\hat{E}}f" near start] \arrow[dd, "{[\![-]\!]_{\widehat{\mathbb{T}}\mathbf{X}}}"' near start] & & \mathcal{T}_\Sigma\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{Y} \arrow[dd, "{[\![-]\!]_{\widehat{\mathbb{T}}\mathbf{Y}}}"] \\
\mathcal{T}_\Sigma X \arrow[rr, "\mathcal{T}_\Sigma f" near end] \arrow[dr, "{[-]_{\hat{E}}}"'] & & \mathcal{T}_\Sigma Y \arrow[dr, "{[-]_{\hat{E}}}"] & \\
& \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[rr, "\widehat{\mathcal{T}}_{\Sigma,\hat{E}}f"'] & & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{Y}
\end{tikzcd}
\tag{3.30}
$$

The first diagram we paved implies (1.28) commutes because $[-]_{\hat{E}}$ is surjective. $\square$

From the front face of the cube above, we find that for any $f : \mathbf{X} \to \mathbf{Y}$, $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}f$ is a homomorphism between the underlying algebras of $\widehat{\mathbb{T}}\mathbf{X}$ and $\widehat{\mathbb{T}}\mathbf{Y}$. We already showed $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}f$ is nonexpansive in Lemma 3.36, thus it is a homomorphism between the quantitative algebras $\widehat{\mathbb{T}}\mathbf{X}$ and $\widehat{\mathbb{T}}\mathbf{Y}$.

We now prove generalizations of results from Chapter 1 in order to show that $\widehat{\mathbb{T}}\mathbf{X}$ is not just a quantitative Σ-algebra but a $(\Sigma, \hat{E})$-algebra.

We can prove, analogously to Lemma 1.30, that for any $\hat{\mathbb{A}} \in \mathbf{QAlg}(\Sigma, \hat{E})$, $[\![-]\!]_A$ is a homomorphism between $\widehat{\mathbb{T}}\mathbf{A}$ and $\hat{\mathbb{A}}$.

Lemma 3.40. *For any $(\Sigma, \hat{E})$-algebra $\hat{\mathbb{A}}$, the square (3.31) commutes, and $[\![-]\!]_A$ is a nonexpansive map $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A} \to \mathbf{A}$.*[387]

[386] In more details, the left and right faces commute by (3.20), the bottom and top faces commute by (3.18), and the back face commutes by (1.7).

The function $\mathcal{T}_\Sigma[-]_{\hat{E}}$ is surjective (i.e. epic) because $[-]_{\hat{E}}$ 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 $\mathcal{T}_\Sigma[-]_{\hat{E}}$ 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 complete derivation (we write which face was used as justifications for each step).

$$
\begin{aligned}
&\widehat{\mathcal{T}}_{\Sigma,\hat{E}}f \circ [\![-]\!]_{\widehat{\mathbb{T}}\mathbf{X}} \circ \mathcal{T}_\Sigma[-]_{\hat{E}} & \\
&= \widehat{\mathcal{T}}_{\Sigma,\hat{E}}f \circ [-]_{\hat{E}} \circ \mu^\Sigma_X & \text{left} \\
&= [-]_{\hat{E}} \circ \mathcal{T}_\Sigma f \circ \mu^\Sigma_X & \text{bottom} \\
&= [-]_{\hat{E}} \circ \mu^\Sigma_Y \circ \mathcal{T}_\Sigma\mathcal{T}_\Sigma f & \text{back} \\
&= [\![-]\!]_{\widehat{\mathbb{T}}\mathbf{Y}} \circ \mathcal{T}_\Sigma[-]_{\hat{E}} \circ \mathcal{T}_\Sigma\mathcal{T}_\Sigma f & \text{right} \\
&= [\![-]\!]_{\widehat{\mathbb{T}}\mathbf{Y}} \circ \mathcal{T}_\Sigma\widehat{\mathcal{T}}_{\Sigma,\hat{E}}f \circ \mathcal{T}_\Sigma[-]_{\hat{E}} & \text{top}
\end{aligned}
$$

[387] We use the same convention as in (1.31) and write $[\![-]\!]_A$ for both maps $\mathcal{T}_\Sigma A \to A$ and $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A} \to A$. Recall the latter is well-defined because whenever $[s]_{\hat{E}} = [t]_{\hat{E}}$, $\hat{\mathbb{A}}$ must satisfy $\mathbf{A} \vdash s = t$, and in particular under the assignment $\mathrm{id}_A : \mathbf{A} \to \mathbf{A}$, this yields $[\![s]\!]_A = [\![t]\!]_A$.

$$
\begin{tikzcd}
\mathcal{T}_\Sigma \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A} \arrow[r, "\mathcal{T}_\Sigma \llbracket - \rrbracket_A"] \arrow[d, "\llbracket - \rrbracket_{\widehat{\mathbb{T}}\mathbf{A}}"'] & \mathcal{T}_\Sigma A \arrow[d, "\llbracket - \rrbracket_A"] \\
\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A} \arrow[r, "\llbracket - \rrbracket_A"'] & A
\end{tikzcd}
\tag{3.31}
$$

Proof. For the commutative square, we can reuse the proof of Lemma 1.30.

Consider the following diagram that we can view as a triangular prism whose front face is (3.31). Both triangles commute by Footnote 387, the square face at the back and on the left commutes by (3.20), and the square face at the back and on the right commutes by (1.13). With the same trick as in the proof of Lemma 3.39 using the surjectivity of $\mathcal{T}_\Sigma[-]_{\hat{E}}$, we conclude that the front face commutes.[388]

$$
\begin{tikzcd}
& \mathcal{T}_\Sigma \mathcal{T}_\Sigma A \arrow[dl, "\mathcal{T}_\Sigma[-]_{\hat{E}}"'] \arrow[dr, "\mathcal{T}_\Sigma \llbracket - \rrbracket_A"] \arrow[dd, "\mu^\Sigma_A"'] & \\
\mathcal{T}_\Sigma \mathcal{T}_{\Sigma,\hat{E}} A \arrow[rr, "\mathcal{T}_\Sigma \llbracket - \rrbracket_A"] \arrow[dd, "\llbracket - \rrbracket_{\widehat{\mathbb{T}}\mathbf{A}}"'] & & \mathcal{T}_\Sigma A \arrow[dd, "\llbracket - \rrbracket_A"] \\
& \mathcal{T}_\Sigma A \arrow[dl, "{[-]_{\hat{E}}}"'] \arrow[dr, "\llbracket - \rrbracket_A"] & \\
\mathcal{T}_{\Sigma,\hat{E}} A \arrow[rr, "\llbracket - \rrbracket_A"'] & & A
\end{tikzcd}
$$

[388] Here is the complete derivation.

$$
\begin{aligned}
&\llbracket - \rrbracket_A \circ \llbracket - \rrbracket_{\widehat{\mathbb{T}}\mathbf{A}} \circ \mathcal{T}_\Sigma[-]_{\hat{E}} \\
&= \llbracket - \rrbracket_A \circ [-]_{\hat{E}} \circ \mu^\Sigma_A && \text{left} \\
&= \llbracket - \rrbracket_A \circ \mu^\Sigma_A && \text{bottom} \\
&= \llbracket - \rrbracket_A \circ \mathcal{T}_\Sigma \llbracket - \rrbracket_A && \text{right} \\
&= \llbracket - \rrbracket_A \circ \mathcal{T}_\Sigma \llbracket - \rrbracket_A \circ \mathcal{T}_\Sigma[-]_{\hat{E}} && \text{top}
\end{aligned}
$$

Then, since $\mathcal{T}_\Sigma[-]_{\hat{E}}$ is epic, we conclude that $\llbracket - \rrbracket_A \circ \llbracket - \rrbracket_{\widehat{\mathbb{T}}\mathbf{A}} = \llbracket - \rrbracket_A \circ \mathcal{T}_\Sigma \llbracket - \rrbracket_A$.

For nonexpansiveness, if $d_{\hat{E}}([s]_{\hat{E}}, [t]_{\hat{E}}) \leq \varepsilon$, then by (3.16) $\mathbf{A} \vdash s =_\varepsilon t$ belongs to $\mathfrak{QTh}(\hat{E})$ which means $\hat{\mathbb{A}}$ must satisfy that equation, and in particular under the assignment $\mathrm{id}_A : \mathbf{A} \to \mathbf{A}$, this yields $d_{\mathbf{A}}(\llbracket s \rrbracket_A, \llbracket t \rrbracket_A) \leq \varepsilon$. $\square$

We can prove, analogously to Lemma 1.31, that for any $\mathbf{X}$, $\widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{X}}$ is a homomorphism from $\widehat{\mathbb{T}}\widehat{\mathbb{T}}\mathbf{X}$ to $\widehat{\mathbb{T}}\mathbf{X}$.

Lemma 3.41. *For any generalized metric space* $\mathbf{X}$*, the following square commutes, and* $\widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{X}}$ *is a nonexpansive map* $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \to \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$.

$$
\begin{tikzcd}
\mathcal{T}_\Sigma \widehat{\mathcal{T}}_{\Sigma,\hat{E}} \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[r, "\mathcal{T}_\Sigma \widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{X}}"] \arrow[d, "\llbracket - \rrbracket_{\widehat{\mathbb{T}}\widehat{\mathbb{T}}\mathbf{X}}"'] & \mathcal{T}_\Sigma \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[d, "\llbracket - \rrbracket_{\widehat{\mathbb{T}}\mathbf{X}}"] \\
\widehat{\mathcal{T}}_{\Sigma,\hat{E}} \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[r, "\widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{X}}"'] & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}
\end{tikzcd}
\tag{3.32}
$$

Proof. We already showed nonexpansiveness in Lemma 3.38. For the commutative square, we can reuse the argument of Lemma 1.31 and add the little hats.

We prove it exactly like Lemma 3.40 with the following diagram.[389]

$$\begin{tikzcd}
 & \mathcal{T}_\Sigma\mathcal{T}_\Sigma\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A} \arrow[dl, "\mathcal{T}_\Sigma[-]_{\hat{E}}"'] \arrow[dr, "\mathcal{T}_\Sigma[\![-]\!]_{\widehat{\mathbb{T}}\mathbf{A}}"] \arrow[dd, "\mu^\Sigma_{\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A}}"] & \\
\mathcal{T}_\Sigma\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A} \arrow[rr, "\mathcal{T}_\Sigma\widehat{\mu}^{\Sigma,\hat{E}}_A"] \arrow[ddd, "[\![-]\!]_{\widehat{\mathbb{T}}\widehat{\mathbb{T}}\mathbf{A}}"'] & & \mathcal{T}_\Sigma\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A} \arrow[ddd, "[\![-]\!]_{\widehat{\mathbb{T}}\mathbf{A}}"] \\
 & \mathcal{T}_\Sigma\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A} \arrow[ddl, "[-]_{\hat{E}}"'] \arrow[ddr, "[\![-]\!]_{\widehat{\mathbb{T}}\mathbf{A}}"] & \\
 & & \\
\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A} \arrow[rr, "\widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{A}}"'] & & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A}
\end{tikzcd}$$

□

Of course, paired with the flattening we also have a map $\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{A}}$ which sends elements $a \in A$ to the equivalence class containing a seen as a trivial term, namely,

$$\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{A}} = \mathbf{A} \xrightarrow{\eta^\Sigma_A} \mathcal{T}_\Sigma A \xrightarrow{[-]_{\hat{E}}} \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A}. \tag{3.33}$$

We need to show $\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{A}}$ is nonexpansive and natural in $\mathbf{A}$.

Lemma 3.42. *For any space $\mathbf{A}$, $\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{A}}$ is a nonexpansive map $\mathbf{A} \to \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A}$.*

Proof. This is a direct consequence of Lemma 3.27. For any $a, a' \in X$ and $\varepsilon \in \mathsf{L}$,

$$\begin{aligned}
d_{\mathbf{A}}(a,a') \leq \varepsilon &\implies \mathbf{A} \vdash a =_\varepsilon a' \in \mathfrak{QTh}(\hat{E}) && \text{by Lemma 3.27}\\
&\iff d_{\hat{E}}([a]_{\hat{E}}, [a']_{\hat{E}}) \leq \varepsilon. && \text{by (3.16)}
\end{aligned}$$

Therefore, $d_{\hat{E}}([a]_{\hat{E}}, [a']_{\hat{E}}) \leq d_{\mathbf{A}}(a,a')$. □

Lemma 3.43. *For any nonexpansive map $f : \mathbf{A} \to \mathbf{B}$, the following square commutes.*[390]

$$\begin{tikzcd}
\mathbf{A} \arrow[r, "\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{A}}"] \arrow[d, "f"'] & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A} \arrow[d, "\widehat{\mathcal{T}}_{\Sigma,\hat{E}}f"] \\
\mathbf{B} \arrow[r, "\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{B}}"'] & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{B}
\end{tikzcd} \tag{3.34}$$

Proof. We pave the following diagram (in **Set**, but that is enough since $U : \mathbf{GMet} \to \mathbf{Set}$ is faithful).

$$\begin{tikzcd}
\mathbf{A} \arrow[rrrr, "\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{A}}"] \arrow[ddrr, "\eta^\Sigma_A"] \arrow[dddddd, "f"'] & & & & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A} \arrow[dddddd, "\widehat{\mathcal{T}}_{\Sigma,\hat{E}}f"] \\
 & & \text{(a)} & & \\
 & & \mathcal{T}_\Sigma A \arrow[uurr, "{[-]_{\hat{E}}}"'] \arrow[dd, "\mathcal{T}_\Sigma f"] & & \\
 & \text{(b)} & & \text{(c)} & \\
 & & \mathcal{T}_\Sigma B \arrow[ddrr, "{[-]_{\hat{E}}}"] & & \\
 & & \text{(d)} & & \\
\mathbf{B} \arrow[uurr, "\eta^\Sigma_B"] \arrow[rrrr, "\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{B}}"'] & & & & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{B}
\end{tikzcd} \tag{3.35}$$

[389] The top and bottom faces commute by definition of $\widehat{\mu}^{\Sigma,\hat{E}}_A$ (3.21), the back-left face by (3.20), and the back-right face by (1.13).
Then, $\mathcal{T}_\Sigma[-]_{\hat{E}}$ is epic, so the following derivation suffices.

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

[390] Naturality of $\eta^{\Sigma,E}$ was easier in **Set** because it is the vertical composition of two natural transformations, η^Σ and $[-]_E$, which do not have counterparts in **GMet**.

Showing (3.35) commutes:

(a) Definition of $\widehat{\eta}^{\Sigma,\hat{E}}$ (3.33).

(b) Naturality of η^Σ (1.5).

(c) Definition of $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}f$ (3.18).

(d) Definition of $\widehat{\eta}^{\Sigma,\hat{E}}$ (3.33).

□

We also have the following technical lemma and its corollary analogous to Lemma 1.32 and Lemma 1.33.

Lemma 3.44. *For any generalized metric space* $\mathbf{X}$, $[\![-]\!]^{\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}}_{\widehat{\mathbb{T}}\mathbf{X}} = [-]_{\hat{E}}$.[391]

[391] The proof is identical to that of Lemma 1.32.

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

$$\begin{aligned}
[\![\eta^{\Sigma}_{X}(x)]\!]^{\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}}_{\widehat{\mathbb{T}}\mathbf{X}} &= [\![\mathcal{T}_{\Sigma}\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}(\eta^{\Sigma}_{X}(x))]\!]_{\widehat{\mathbb{T}}\mathbf{X}} && \text{by (1.9)}\\
&= [\![\mathcal{T}_{\Sigma}[-]_{\hat{E}}(\mathcal{T}_{\Sigma}\eta^{\Sigma}_{X}(\eta^{\Sigma}_{X}(x)))]\!]_{\widehat{\mathbb{T}}\mathbf{X}} && \text{Proposition 1.6}\\
&= [\![\mathcal{T}_{\Sigma}[-]_{\hat{E}}(\eta^{\Sigma}_{\mathcal{T}_{\Sigma}X}(\eta^{\Sigma}_{X}(x)))]\!]_{\widehat{\mathbb{T}}\mathbf{X}} && \text{by (1.5)}\\
&= [\![\eta^{\Sigma}_{\mathcal{T}_{\Sigma,\hat{E}}X}([\eta^{\Sigma}_{X}(x)]_{\hat{E}})]\!]_{\widehat{\mathbb{T}}\mathbf{X}} && \text{by (1.5)}\\
&= [\eta^{\Sigma}_{X}(x)]_{\hat{E}} && \text{by (1.26)}
\end{aligned}$$

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

$$\begin{aligned}
[\![t]\!]^{\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}}_{\widehat{\mathbb{T}}\mathbf{X}} &= [\![\mathcal{T}_{\Sigma}\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}(t)]\!]_{\widehat{\mathbb{T}}\mathbf{X}} && \text{by (1.9)}\\
&= [\![\mathcal{T}_{\Sigma}\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}(\mathsf{op}(t_1,\dots,t_n))]\!]_{\widehat{\mathbb{T}}\mathbf{X}} \\
&= [\![\mathsf{op}(\mathcal{T}_{\Sigma}\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}(t_1),\dots,\mathcal{T}_{\Sigma}\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}(t_n))]\!]_{\widehat{\mathbb{T}}\mathbf{X}} && \text{by (1.4)}\\
&= [\![\mathsf{op}]\!]_{\widehat{\mathbb{T}}\mathbf{X}}\left([\![\mathcal{T}_{\Sigma}\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}(t_1)]\!]_{\widehat{\mathbb{T}}\mathbf{X}},\cdots,[\![\mathcal{T}_{\Sigma}\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}(t_n)]\!]_{\widehat{\mathbb{T}}\mathbf{X}}\right) && \text{by (1.26)}\\
&= [\![\mathsf{op}]\!]_{\widehat{\mathbb{T}}\mathbf{X}}\left([t_1]_{\hat{E}},\cdots,[t_n]_{\hat{E}}\right) && \text{I.H.}\\
&= [\mathsf{op}(t_1,\dots,t_n)]_{\hat{E}}. && \text{by (3.19)} \qquad \square
\end{aligned}$$

We get that for any quantitative equation ϕ with context $\mathbf{X}$, ϕ belongs to $\mathfrak{QTh}(\hat{E})$ if and only if the algebra $\widehat{\mathbb{T}}_{\Sigma,\hat{E}}\mathbf{X}$ satisfies it under the assignment $\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}$.

Lemma 3.45. *Let* ϕ *be an equation with context* $\mathbf{X}$, $\phi \in \mathfrak{QTh}(\hat{E})$ *if and only if* $\widehat{\mathbb{T}}\mathbf{X} \vDash^{\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}} \phi$.[392]

[392] Once again, we are only adapting the argument from the proof of Lemma 1.33.

Proof. We have two cases to show.

- $\mathbf{X} \vdash s = t \in \mathfrak{QTh}(\hat{E})$ if and only if $\widehat{\mathbb{T}}\mathbf{X} \vDash^{\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}} \mathbf{X} \vdash s = t$, and
- $\mathbf{X} \vdash s =_{\varepsilon} t \in \mathfrak{QTh}(\hat{E})$ if and only if $\widehat{\mathbb{T}}\mathbf{X} \vDash^{\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}} \mathbf{X} \vdash s =_{\varepsilon} t$.

By Lemma 3.44,

$$[\![s]\!]^{\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}}_{\widehat{\mathbb{T}}X} = [s]_{\hat{E}} \text{ and } [\![t]\!]^{\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}}_{\widehat{\mathbb{T}}X} = [t]_{\hat{E}}, \tag{3.36}$$

then by using definitions, we have (as desired)

$$\begin{aligned}
\mathbf{X} \vdash s = t \in \mathfrak{QTh}(\hat{E}) &\overset{(3.12)}{\iff} [s]_{\hat{E}} = [t]_{\hat{E}} &&\overset{(3.36)}{\iff} [\![s]\!]^{\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}}_{\widehat{\mathbb{T}}X} = [\![t]\!]^{\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}}_{\widehat{\mathbb{T}}X}\\
\mathbf{X} \vdash s =_{\varepsilon} t \in \mathfrak{QTh}(\hat{E}) &\overset{(3.16)}{\iff} d_{\hat{E}}([s]_{\hat{E}},[t]_{\hat{E}}) \leq \varepsilon &&\overset{(3.36)}{\iff} d_{\hat{E}}([\![s]\!]^{\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}}_{\widehat{\mathbb{T}}X},[\![t]\!]^{\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}}_{\widehat{\mathbb{T}}X}) \leq \varepsilon. \qquad \square
\end{aligned}$$

The next result, analogous to Lemma 1.34, tells us that $\widehat{\eta}^{\Sigma,\hat{E}}$ and $\widehat{\mu}^{\Sigma,\hat{E}}$ interact together like the unit and multiplication of a monad.

Lemma 3.46. *The following diagram commutes.*[393]

$$\begin{tikzcd}
\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[r, "\widehat{\eta}^{\Sigma,\hat{E}}_{\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}}"] \arrow[dr, "\mathrm{id}_{\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}}"'] & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[d, "\widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{X}}"] & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[l, "\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}"'] \arrow[dl, "\mathrm{id}_{\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}}"] \\
& \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} &
\end{tikzcd}$$

[393] We reuse the proof of Lemma 1.34, although when using naturality of $[-]_{\hat{E}}$ in **Set**, we replace it by (3.18) which is not formally a naturality property (because $\mathcal{T}_\Sigma$ is not a functor on **GMet**).

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

$$\begin{tikzcd}
\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[r, "\eta^{\Sigma}_{\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}}"] \arrow[rr, bend left=50, "\widehat{\eta}^{\Sigma,\hat{E}}_{\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}}"] \arrow[drr, bend right, "\mathrm{id}_{\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}}"'] & \mathcal{T}_\Sigma\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[r, "{[-]_{\hat{E}}}"] \arrow[dr, "\llbracket - \rrbracket_{\mathbb{T}X}"'] \arrow[u, phantom, "\text{(a)}"] \arrow[d, phantom, "\text{(b)}"] & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[d, "\widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{X}}"] \arrow[dl, phantom, "\text{(c)}"] \\
& & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}
\end{tikzcd} \tag{3.37}$$

Showing (3.37) commutes:

(a) Definition of $\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}$ (3.33).

(b) Definition of $\llbracket - \rrbracket_{\mathbb{T}X}$ (1.26).

(c) Definition of $\widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{X}}$ (3.21).

For the triangle on the right, we show that $[-]_{\hat{E}} = \widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{X}} \circ \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}} \circ [-]_{\hat{E}}$ by paving (3.38), and we can conclude since $[-]_{\hat{E}}$ is epic that $\mathrm{id}_{\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}} = \widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{X}} \circ \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}$.

$$\begin{tikzcd}
\mathcal{T}_\Sigma X \arrow[r, "{[-]_{\hat{E}}}"] \arrow[drr, "\mathcal{T}_\Sigma\eta^\Sigma_X"'] \arrow[ddrr, "\mathrm{id}_{\mathcal{T}_\Sigma X}"'] & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[r, "\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\eta^\Sigma_X"] \arrow[rrr, bend left=30, "\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}"] \arrow[d, phantom, "\text{(b)}"] & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathcal{T}_\Sigma X \arrow[rr, "\widehat{\mathcal{T}}_{\Sigma,\hat{E}}{[-]_{\hat{E}}}"] \arrow[r, phantom, "\text{(a)}", near end, shift left=4] & & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[dd, "\widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{X}}"] \\
& \text{(d)} & \mathcal{T}_\Sigma\mathcal{T}_\Sigma X \arrow[u, "{[-]_{\hat{E}}}"] \arrow[r, "\mathcal{T}_\Sigma{[-]_{\hat{E}}}"'] \arrow[d, "\mu^\Sigma_X"] & \mathcal{T}_\Sigma\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[ur, "{[-]_{\hat{E}}}"'] \arrow[dr, "\llbracket - \rrbracket_{\mathbb{T}X}"] \arrow[u, phantom, "\text{(c)}"] \arrow[d, phantom, "\text{(e)}"] & \text{(f)} \\
& & \mathcal{T}_\Sigma X \arrow[rr, "{[-]_{\hat{E}}}"'] & & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}
\end{tikzcd} \tag{3.38}$$

Showing (3.38) commutes:

(a) Definition of $\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}$ and functoriality of $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}$.

(b) "Naturality" of $[-]_{\hat{E}}$ (3.18).

(c) By (3.18) again.

(d) Definition of μ^Σ_X (1.6).

(e) By (3.20).

(f) By (3.21).

□

Finally, we can show that $\widehat{\mathbb{T}}_{\Sigma,\hat{E}}\mathbf{X}$ is $(\Sigma, \hat{E})$-algebra (analogous to Proposition 1.37).

Proposition 3.47. *For any space* $\mathbf{A}$*, the term algebra* $\widehat{\mathbb{T}}_{\Sigma,\hat{E}}\mathbf{A}$ *satisfies all the equations in* $\hat{E}$*.*

Proof. Let $\phi \in \hat{E}$ be an equation with context $\mathbf{X}$ and $\hat{\imath} : \mathbf{X} \to \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A}$ be a nonexpansive assignment. We factor $\hat{\imath}$ into[394]

$$\hat{\imath} = X \xrightarrow{\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}} \widehat{\mathcal{T}}_{\Sigma,\hat{E}}X \xrightarrow{\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\hat{\imath}} \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A} \xrightarrow{\widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{A}}} \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A}.$$

[394] This factoring is correct because

$$\begin{aligned}
\hat{\imath} &= \mathrm{id}_{\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A}} \circ \hat{\imath} \\
&= \widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{A}} \circ \widehat{\eta}^{\Sigma,\hat{E}}_{\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A}} \circ \hat{\imath} && \text{Lemma 3.46} \\
&= \widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{A}} \circ \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\hat{\imath} \circ \widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}. && \text{naturality of } \widehat{\eta}^{\Sigma,\hat{E}}
\end{aligned}$$

Now, Lemma 3.45 says that ϕ is satisfied in $\widehat{\mathbb{T}}\mathbf{X}$ under the assignment $\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}$. We also know by Lemma 3.13 that homomorphisms preserve satisfaction, so we can apply it twice using the facts that $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\hat{\imath}$ and $\widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{A}}$ are homomorphisms (the former was shown after Lemma 3.39 and the latter in Lemma 3.41) to conclude that $\widehat{\mathbb{T}}\mathbf{A}$ satisfies ϕ under $\widehat{\mu}^{\Sigma,\hat{E}}_{\mathbf{A}} \circ \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\hat{\imath} \circ \widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}} = \hat{\imath}$. $\square$

We end this section just like we ended §1.1 by showing that $\widehat{\mathbb{T}}\mathbf{X}$ is the free $(\Sigma, \hat{E})$-algebra.[395]

[395] In both [MSV22] and [MSV23], we constructed the free algebra using quantitative equational logic.

Theorem 3.48. *For any space* $\mathbf{X}$*, the term algebra* $\widehat{\mathbb{T}}X$ *is the free* $(\Sigma, \hat{E})$*-algebra on* $\mathbf{X}$*.*

Proof. Note that the morphism witnessing freeness of $\widehat{\mathbb{T}}\mathbf{X}$ is $\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}} : \mathbf{X} \to \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$.[396]

[396] As expected, the proof goes exactly like for Proposition 1.40 except for dealing with nonexpansiveness at the end.

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

$$\begin{tikzcd}
\text{in } \mathbf{GMet} & & & \text{in } \mathbf{QAlg}(\Sigma,\hat{E}) \\
\mathbf{X} \arrow[r, "\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}"] \arrow[dr, "f"'] & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[d, dashed, "f^*"] & & \widehat{\mathbb{T}}\mathbf{X} \arrow[d, dashed, "f^*"] \\
& \mathbf{A} & & \hat{\mathbb{A}}
\end{tikzcd}
\qquad \xleftarrow{\;U\;}$$

First, f^* is a homomorphism because it is the composite of two homomorphisms $\widehat{\mathcal{T}}_{\Sigma,\hat{E}} f$ (by (3.30)) and $[\![-]\!]_A$ (by Lemma 3.40 since $\hat{\mathbb{A}}$ satisfies $\hat{E}$). Next, the triangle commutes by the following derivation.

$$\begin{aligned}
[\![-]\!]_A \circ \widehat{\mathcal{T}}_{\Sigma,\hat{E}} f \circ \widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}} &= [\![-]\!]_A \circ \widehat{\eta}^{\Sigma,\hat{E}}_A \circ f && \text{by (3.34)} \\
&= [\![-]\!]_A \circ [-]_{\hat{E}} \circ \eta^{\Sigma}_A \circ f && \text{definition of } \widehat{\eta}^{\Sigma,\hat{E}} \\
&= [\![-]\!]_A \circ \eta^{\Sigma}_A \circ f && \text{Footnote 387} \\
&= f && \text{definition of } [\![-]\!]_A \text{ (3.19)}
\end{aligned}$$

Finally, uniqueness follows from the inductive definition of $\widehat{\mathbb{T}}\mathbf{X}$ 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.[397]

[397] Formally, let $f, g : \widehat{\mathbb{T}}\mathbf{X} \to \hat{\mathbb{A}}$ be two homomorphisms such that for any $x \in X$, $f[x]_{\hat{E}} = g[x]_{\hat{E}}$, then, we can show that $f = g$. For any $t \in \mathcal{T}_{\Sigma}X$, we showed in Lemma 3.44 that $[t]_{\hat{E}} = [\![t]\!]^{\widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}}_{\widehat{\mathbb{T}}\mathbf{X}}$. Then using (1.11), we have

$$f[t]_{\hat{E}} = [\![t]\!]_A^{f \circ \widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}} = [\![t]\!]_A^{g \circ \widehat{\eta}^{\Sigma,\hat{E}}_{\mathbf{X}}} = g[t]_{\hat{E}},$$

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

It remains to show that $f^* : \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \to \mathbf{A}$ is nonexpansive. This follows by the following derivation, where we implicitly use nonexpansiveness of f in the second step, where f is used as a nonexpansive assignment.

$$\begin{aligned}
d_{\hat{E}}([s]_{\hat{E}}, [t]_{\hat{E}}) \leq \varepsilon &\iff \mathbf{X} \vdash s =_{\varepsilon} t \in \mathfrak{QTh}(\hat{E}) && \text{by (3.16)} \\
&\implies d_{\mathbf{A}}([\![s]\!]^f_A, [\![t]\!]^f_A) \leq \varepsilon && \hat{\mathbb{A}} \in \mathbf{QAlg}(\Sigma, \hat{E}) \\
&\iff d_{\mathbf{A}}([\![\mathcal{T}_{\Sigma} f(s)]\!]_A, [\![\mathcal{T}_{\Sigma} f(t)_A]\!]) && \text{by (1.9)} \\
&\iff d_{\mathbf{A}}([\![[\mathcal{T}_{\Sigma} f(s)]_{\hat{E}}]\!]_A, [\![[\mathcal{T}_{\Sigma} f(t)]_{\hat{E}}]\!]_A) && \text{Footnote 387} \\
&\iff d_{\mathbf{A}}([\![\widehat{\mathcal{T}}_{\Sigma,\hat{E}} f[s]_{\hat{E}}]\!]_A, [\![\widehat{\mathcal{T}}_{\Sigma,\hat{E}} f[t]_{\hat{E}}]\!]_A) && \text{by (3.18)} \\
&\iff d_{\mathbf{A}}(f^*[s]_{\hat{E}}, f^*[t]_{\hat{E}}) && \text{definition of } f^* \quad \square
\end{aligned}$$

Since we have a free $(\Sigma, \hat{E})$-algebra $\widehat{\mathbb{T}}\mathbf{X}$ for every generalized metric space $\mathbf{X}$, we get a left adjoint to $U : \mathbf{QAlg}(\Sigma, \hat{E}) \to \mathbf{GMet}$. This automatically yields a monad structure on $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}$ that we will study after developing quantitative equational logic. Before that, we make use of a special case of the adjunction above.

Corollary 3.49. *The forgetful functor $U : \mathbf{GMet} \to \mathbf{Set}$ has a left adjoint.*

Proof. The following adjoints compose to yield a left adjoint to $U : \mathbf{GMet} \to \mathbf{Set}$.[398]

$$\begin{tikzcd} \mathbf{GMet} \arrow[r, shift left] \arrow[r, phantom, "\top"] \arrow[rr, bend left, "U"] & \mathsf{L}\mathbf{Spa} \arrow[l, shift left] \arrow[r, shift left, "U"] \arrow[r, phantom, "\top"] & \mathbf{Set} \arrow[l, shift left] \end{tikzcd}$$ □

Example 3.50 (Discrete metric)**.** To make this more concrete, one can wonder what is the free metric space on a set X (with $\mathsf{L} = [0,1]$). According to the diagram above, we first need to construct the discrete space $\mathbf{X}_\top$ on X, then construct the free metric space on $\mathbf{X}_\top$. We know how to do the first step (Proposition 2.57), and the second step is also fairly easy to do.[399] The only thing that prevents $\mathbf{X}_\top$ from being a metric is reflexivity, i.e. $d_\top(x,x) = 1 \neq 0$. If we define $d_\mathbf{X}$ just like $d_\top$ except with $d_\mathbf{X}(x,x) = 0$, then it is a metric,[400] and $(X, d_\mathbf{X})$ is the free metric space over X.

Corollary 3.49 applies to any category $\mathbf{GMet}$, so we can always construct the discrete generalized metric on a set.

With the help of quantitative algebraic theories and free algebras, we can now define coproducts inside $\mathbf{GMet}$.

Corollary 3.51. *The category* $\mathbf{GMet}$ *has coproducts.*

Proof. We will only do the case of binary coproducts for exposition's sake, but the proof can be adapted to arbitrary families. For any generalized metric space $\mathbf{A}$, the quantitative algebraic theory of $\mathbf{A}$ is generated by the signature $\Sigma_\mathbf{A} = \{a : 0 \mid a \in A\}$ and the quantitative equations[401]

$$\hat{E}_\mathbf{A} = \left\{ \vdash a =_{d_\mathbf{A}(a,a')} a' \mid a, a' \in A \right\}.$$

A $(\Sigma_\mathbf{A}, \hat{E}_\mathbf{A})$-algebra $\hat{\mathbb{B}}$ is a generalized metric space $\mathbf{B}$ equipped with an interpretation $[\![a]\!]_B$ for every $a \in A$ such that $d_\mathbf{B}([\![a]\!]_B, [\![a']\!]_B) \leq d_\mathbf{A}(a,a')$ for every $a, a' \in A$. Equivalently, all the interpretations can be seen as a single nonexpansive map $[\![-]\!]_B : \mathbf{A} \to \mathbf{B}$. Therefore, $\mathbf{QAlg}(\Sigma_\mathbf{A}, \hat{E}_\mathbf{A})$ is the coslice category $\mathbf{A}/\mathbf{GMet}$.

Given another space $\mathbf{A}'$, if we combine the theories of $\mathbf{A}$ and $\mathbf{A}'$ with no additional equations, we get the category $\mathbf{QAlg}(\Sigma_\mathbf{A} + \Sigma_{\mathbf{A}'}, \hat{E}_\mathbf{A} + \hat{E}_{\mathbf{A}'})$ of spaces $\mathbf{B}$ equipped with two nonexpansive maps $[\![-]\!]_B : \mathbf{A} \to \mathbf{B}$ and $[\![-]\!]'_B : \mathbf{A}' \to B$. This category has an initial object, the free algebra on the initial generalized metric space from Proposition 2.39. Moreover, this category can be equivalently described as the comma category $[\mathbf{A}, \mathbf{A}'] \downarrow \mathrm{id}_{\mathbf{GMet}}$ where $[\mathbf{A}, \mathbf{A}'] : \mathbf{1} + \mathbf{1} \to \mathbf{GMet}$ is the constant functor sending the two objects in the domain to $\mathbf{A}$ and $\mathbf{A}'$ respectively.[402] The initial object of this category (we just showed it exists) is the coproduct $\mathbf{A} + \mathbf{A}'$ (by definition of coproducts and comma categories). □

[398] The adjunction between $\mathsf{L}\mathbf{Spa}$ and $\mathbf{Set}$ was described in Proposition 2.57. The adjunction between $\mathbf{GMet}$ and $\mathsf{L}\mathbf{Spa}$ is the one we just obtained via Theorem 3.48 that we instantiate with $\mathbf{GMet} = \mathbf{QAlg}(\emptyset, \hat{E}_{\mathbf{GMet}})$ (recall Example 3.16).

[399] Even though we said in Example 3.19 that the free metric space on an arbitrary $\mathbf{X}$ is harder to describe.

[400] Identity of indiscernibles and symmetry hold because $d_\mathbf{X}(x,y) = d_\mathbf{X}(y,x) = 1$ when $x \neq y$. The triangle inequality holds because

$$d_\mathbf{X}(x,z) = 1 \leq 1 + 1 = d_\mathbf{X}(x,y) + d_\mathbf{X}(y,z).$$

[401] Note that a and a' are seen as constants, not variables, so the context of these equations is the empty L-space.

[402] The category $\mathbf{1} + \mathbf{1}$ has two objects, their identity morphisms and that is it.

3.2 Quantitative Equational Logic

It is now time to introduce quantitative equational logic (QEL), which you can think of as both a generalization and an extension of equational logic. It is a generalization because it is parametrized by a complete lattice L, and when instantiating $\mathsf{L} = 1$, we get back equational logic as explained in Example 3.56. It is an extension because all the rules of equational logic are valid in QEL when replacing the contexts with discrete spaces as explained in Example 3.57. Figure 3.1 displays the inference rules of **quantitative equational logic**. The notion of **derivation** is straightforwardly adapted from Definition 1.41, the crucial difference is that proof trees can now be infinite.[403]

[403] This is necessary due to the rules SUB, SUBQ, and CONT, which can have infinitely many quantitative equations as hypotheses.

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

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

$$\frac{\sigma : Y \to \mathcal{T}_\Sigma X \qquad \mathbf{Y} \vdash s = t \qquad \forall y, y' \in Y,\ \mathbf{X} \vdash \sigma(y) =_{d_{\mathbf{Y}}(y,y')} \sigma(y')}{\mathbf{X} \vdash \sigma^*(s) = \sigma^*(t)}\ \textsc{Sub}$$

$$\frac{}{\mathbf{X} \vdash s =_\top t}\ \textsc{Top} \qquad \frac{d_{\mathbf{X}}(x, x') = \varepsilon}{\mathbf{X} \vdash x =_\varepsilon x'}\ \textsc{Vars} \qquad \frac{\mathbf{X} \vdash s =_\varepsilon t \qquad \varepsilon \leq \varepsilon'}{\mathbf{X} \vdash s =_{\varepsilon'} t}\ \textsc{Max}$$

$$\frac{\forall i, \mathbf{X} \vdash s =_{\varepsilon_i} t \qquad \varepsilon = \inf_i \varepsilon_i}{\mathbf{X} \vdash s =_\varepsilon t}\ \textsc{Cont} \qquad \frac{\phi \in \hat{E}_{\mathbf{GMet}}}{\phi}\ \textsc{GMet}$$

$$\frac{\mathbf{X} \vdash s = t \qquad \mathbf{X} \vdash s =_\varepsilon u}{\mathbf{X} \vdash t =_\varepsilon u}\ \textsc{CompL} \qquad \frac{\mathbf{X} \vdash s = t \qquad \mathbf{X} \vdash u =_\varepsilon s}{\mathbf{X} \vdash u =_\varepsilon t}\ \textsc{CompR}$$

$$\frac{\sigma : Y \to \mathcal{T}_\Sigma X \qquad \mathbf{Y} \vdash s =_\varepsilon t \qquad \forall y, y' \in Y, \mathbf{X} \vdash \sigma(y) =_{d_{\mathbf{Y}}(y,y')} \sigma(y')}{\mathbf{X} \vdash \sigma^*(s) =_\varepsilon \sigma^*(t)}\ \textsc{SubQ}$$

Figure 3.1: Rules of quantitative equational logic over the signature Σ and the complete lattice L, where $\mathbf{X}$ and $\mathbf{Y}$ can be any L-space, s, t, u, s_i and t_i can be any term in $\mathcal{T}_\Sigma X$, and ε, ε' and ε_i range over L. As indicated in the premises of the rules CONG, SUB and SUBQ, they can be instantiated for any n-ary operation symbol and for any function σ respectively.

Given any class of quantitative equations $\hat{E}$, we denote by $\mathfrak{QTh}'(\hat{E})$ the class of equations that can be proven from $\hat{E}$ in quantitative equational logic, in other words, $\phi \in \mathfrak{QTh}'(\hat{E})$ if and only if there is a derivation of ϕ in QEL with axioms $\hat{E}$.

Our goal now is to prove that $\mathfrak{QTh}'(\hat{E}) = \mathfrak{QTh}(\hat{E})$. We say that QEL is sound and complete for $(\Sigma, \hat{E})$-algebras. Less concisely, soundness means that whenever QEL proves an equation ϕ with axioms $\hat{E}$, ϕ is satisfied by all $(\Sigma, \hat{E})$-algebras, and completeness says that whenever an equation ϕ is satisfied by all $(\Sigma, \hat{E})$-algebras, there is a derivation of ϕ in QEL with axioms $\hat{E}$.

Just like for equational logic, all the rules in Figure 3.1 are sound for any fixed quantitative algebra meaning that if $\hat{\mathbb{A}}$ satisfies the equations on top of a rule, it must satisfy the conclusion of that rule. Let us explain the rules as we prove soundness.

The first four rules say that equality is an equivalence relation that is preserved by the operations, we showed they were sound in Lemmas 3.20–3.23. More formally, we can define (for any $\mathbf{X}$) a binary relation $\equiv'_{\hat{E}}$ on Σ-terms[404] that contains the pair (s,t) whenever $\mathbf{X} \vdash s = t$ can be proven in QEL (c.f. (3.12)): for any $s,t \in \mathcal{T}_\Sigma X$,

[404] Again, we omit the L-space $\mathbf{X}$ from the notation.

$$s \equiv'_{\hat{E}} t \iff \mathbf{X} \vdash s = t \in \mathfrak{QTh}'(\hat{E}). \tag{3.39}$$

Then, Refl, Symm, Trans, and Cong make $\equiv'_{\hat{E}}$ a congruence relation.

Lemma 3.52. *For any* L*-space* $\mathbf{X}$*, the relation* $\equiv'_{\hat{E}}$ *is reflexive, symmetric, transitive, and for any* $\mathsf{op} : n \in \Sigma$ *and* $s_1, \dots, s_n, t_1, \dots, t_n \in \mathcal{T}_\Sigma X$,[405]

[405] i.e. $\equiv'_{\hat{E}}$ is a congruence on the Σ-algebra $\mathcal{T}_\Sigma X$ defined in Remark 1.17.

$$\forall 1 \leq i \leq n, s_i \equiv'_{\hat{E}} t_i \implies \mathsf{op}(s_1, \dots, s_n) \equiv'_{\hat{E}} \mathsf{op}(t_1, \dots, t_n). \tag{3.40}$$

We denote with $\lbag - \rbag_{\hat{E}}$ the canonical quotient map $\mathcal{T}_\Sigma X \to \mathcal{T}_\Sigma X / {\equiv'_{\hat{E}}}$.

Skipping Sub for now, the Top rule says that $\top$ is an upper bound for all distances since it is the maximum element of L. We showed it is sound in Lemma 3.26.

The Vars rule is, in a sense, the quantitative version of Refl. It reflects the fact that assignments of variables are nonexpansive with respect to the distance in the context. Indeed, $\hat{\iota} : \mathbf{X} \to \mathbf{A}$ is nonexpansive precisely when, for all $x, x' \in X$,

$$d_{\mathbf{A}}(\hat{\iota}(x), \hat{\iota}(x')) = d_{\mathbf{A}}([\![x]\!]_A^{\hat{\iota}}, [\![x']\!]_A^{\hat{\iota}}) \leq d_{\mathbf{X}}(x, x').$$

How is this related to Refl? Letting $t = x \in X$, Refl says that for any assignment $\hat{\iota} : \mathbf{X} \to \mathbf{A}$, $\hat{\iota}(x) = \hat{\iota}(x)$. This seems trivial, but it hides a deeper fact that the assignment must be deterministic (a functional relation), as it cannot assign two different values to the same input.[406] So just like Refl imposes the constraint of determinism on assignments, Vars imposes nonexpansiveness. We showed Vars is sound in Lemma 3.27.

[406] A similar thing happens for Cong which says that the interpretations of operation are deterministic (both in equational logic and QEL). In [MPP16], the logic has a rule NExp which morally says that the interpretations of operations are nonexpansive too, i.e. NExp is to Cong what Vars is to Refl. We said more on our choice to omit NExp in §0.3.

The rules Max and Cont should remind you of the definition of L-structure (Definition 2.18). Very briefly, they ensure that equipping the set of terms over X with the relations $R_\varepsilon^{\mathbf{X}} \subseteq \mathcal{T}_\Sigma X \times \mathcal{T}_\Sigma X$ defined by

$$s \, R_\varepsilon^{\mathbf{X}} \, t \iff \mathbf{X} \vdash s =_\varepsilon t \in \mathfrak{QTh}'(\hat{E}), \tag{3.41}$$

yields an L-structure.[407] We showed they are sound in Lemmas 3.28 and 3.29. Note that Top is an instance of Cont with the empty index set (recall that $\top = \inf \emptyset$).

[407] Monotonicity and continuity hold by Max and Cont respectively. This is where the name Cont comes from, and this is why I prefer it over the other names in the literature.

The soundness of GMet is a consequence of (3.4) and the definition of quantitative algebra which requires the underlying space to satisfy all the equations in $\hat{E}_{\mathbf{GMet}}$.

CompL and CompR guarantee that the L-structure we just defined factors through the quotient $\mathcal{T}_\Sigma X / {\equiv'_{\hat{E}}}$.[408] We showed they are sound in Lemmas 3.24 and 3.25. In the presence of a symmetry axiom, only one of them would be sufficient.

[408] i.e. the following relation is well-defined:

$$\lbag s \rbag_{\hat{E}} \, R_\varepsilon^{\mathbf{X}} \, \lbag t \rbag_{\hat{E}} \iff \mathbf{X} \vdash s =_\varepsilon t \in \mathfrak{QTh}'(\hat{E}), \tag{3.42}$$

Finally, we get to the substitutions Sub and SubQ, they are the same except for replacing $=$ with $=_\varepsilon$. Recall that the substitution rule in equational logic is

$$\frac{\sigma : Y \to \mathcal{T}_\Sigma X \qquad Y \vdash s = t}{X \vdash \sigma^*(s) = \sigma^*(t)},$$

which morally means that variables in the context Y are universally quantified. In SUB and SUBQ, there is an additional condition on σ which arises because the variables in Y are *not* universally quantified, an assignment $Y \to A$ is considered in the definition of satisfaction only if it is nonexpansive from $\mathbf{Y}$ to $\mathbf{A}$.[409]

[409] Put differently, the variables are universally quantified subject to certain constraints on their distances relative to the context $\mathbf{Y}$.

We proved SUB and SUBQ are sound in Lemma 3.34, and we can compare with the proof of soundness of SUB in equational logic (Lemma 1.36) to find the same key argument: the interpretation of $\sigma^*(t)$ under some assignment $\hat{\imath}$ is equal to the interpretation of t under the assignment $\hat{\imath}_\sigma$ sending y to the interpretation of $\sigma(y)$ under $\hat{\imath}$. Since satisfaction for quantitative algebras only deals with nonexpansive assignments, we needed to check that $\hat{\imath}_\sigma$ is nonexpansive whenever $\hat{\imath}$ is, and this was true thanks to the conditions on σ. Let us give an illustrative example of why the extra conditions are necessary.

Example 3.53. We work over $\mathsf{L} = [0,1]$, $\mathbf{GMet} = \mathbf{Met}$, $\Sigma = \emptyset$, and $\hat{E} = \emptyset$. Let $\mathbf{Y} = \{y_0, y_1\}$ with $d_{\mathbf{Y}}(y_0, y_1) = d_{\mathbf{Y}}(y_1, y_0) = \frac{1}{2}$ and $\mathbf{X} = \{x_0, x_1\}$ with $d_{\mathbf{X}}(x_0, x_1) = d_{\mathbf{X}}(x_1, x_0) = 1$.[410] We consider the algebra $\hat{\mathbb{A}}$ whose underlying space is $\mathbf{A} = \mathbf{X}$ (since Σ is empty that is the only data required to define an algebra). It satisfies the equation $\mathbf{Y} \vdash y_0 = y_1$ because any nonexpansive assignment of $\mathbf{Y}$ into $\mathbf{A}$ must identify y_0 and y_1 (there are no distinct points with distance less than $\frac{1}{2}$).

[410] We can see both $\mathbf{Y}$ and $\mathbf{X}$ as subspace of $[0,1]$ with the Euclidean metric, where e.g. y_0 is embedded as 0 and y_1 as $\frac{1}{2}$, and x_0 is embedded as 0 and x_1 as 1.

Take the substitution $\sigma : Y \to \mathcal{T}_\Sigma X$ defined by $y_0 \mapsto x_0$ and $y_1 \mapsto x_1$, we can check $\hat{\mathbb{A}}$ does not satisfy $\mathbf{X} \vdash \sigma^*(y_0) = \sigma^*(y_1)$.[411] This means that σ cannot satisfy the extra conditions in SUB. Indeed, $\hat{\mathbb{A}}$ does not satisfy $\mathbf{X} \vdash \sigma(y_0) =_{\frac{1}{2}} \sigma(y_1)$ (take the assignment $\mathrm{id}_{\mathbf{X}}$ again).

[411] That equation is $\mathbf{X} \vdash x_0 = x_1$ and with the assignment $\mathrm{id}_{\mathbf{X}} : \mathbf{X} \to \mathbf{X} = \mathbf{A}$, we have
$$[\![x_0]\!]_A^{\mathrm{id}_{\mathbf{X}}} = x_0 \neq x_1 = [\![x_1]\!]_A^{\mathrm{id}_{\mathbf{X}}}.$$

Remark 3.54. The substitution rule in the original paper [MPP16, (Subst) in Definition 2.1] is
$$\frac{\{s_i =_{\varepsilon_i} t_i\} \vdash s =_\varepsilon t}{\{\sigma^*(s_i) =_{\varepsilon_i} \sigma^*(t_i)\} \vdash \sigma^*(s) =_\varepsilon \sigma^*(t)}.$$
This cannot easily be translated into our framework because it has to work with quantitative inferences that are not basic (Remark 3.9). Indeed, even if the top inference is basic (i.e. each s_i and t_i are variables), the bottom one will not basic be when σ sends these variables to complex terms. In a sense, we can say that our quantitative equational logic is closed under basic quantitative inferences,[412] while theirs is not.

[412] Recall that basic quantitative inferences correspond to quantitative equations.

This is an advantage of our presentation with respect to its comparison with equational logic. Indeed, non-basic quantitative inferences are a better analog for implications in implicational logic [Wec12, §3.3, Definition 1]. For example, you can model cancellative monoids, with something like $x + y =_0 x + z \vdash y =_0 z$, which are the canonical example of structures not captured by universal algebra.

By proving each rule is sound, we have shown that QEL is sound.

Theorem 3.55 (Soundness). *If $\phi \in \mathfrak{QTh}'(\hat{E})$, then $\phi \in \mathfrak{QTh}(\hat{E})$.*

Let us explain how to recover equational logic from quantitative equational logic in two different ways.

Example 3.56 (Recovering equational logic I)**.** In Example 2.19, we saw that $1\mathbf{Spa}$ is the category **Set**. Here we show that QEL over the complete lattice 1 with $\hat{E}_{\mathbf{GMet}} = \emptyset$ is the same thing as equational logic. First, what is a quantitative equation ϕ over 1? Since the context is a 1-space, it is just a set,[413] and furthermore, since 1 contains a single element (which we call $\top$ here, but it is equal to $\bot$) ϕ is either

$$X \vdash s = t \quad \text{or} \quad X \vdash s =_{\top} t.$$

Now, the second equation always belongs to $\mathfrak{QTh}'(\hat{E})$ for any $\hat{E}$ by Top. Therefore, the rules whose conclusions have an equation with a quantity (all but the first five) can be replaced by Top. The remaining rules are exactly those of equational logic except the substitution rule which has some additional constraints. The latter require proving only equations with quantities which we can always do with Top.

Thus, we can infer that for any $\hat{E}$, the equations without quantities in $\mathfrak{QTh}'(\hat{E})$ are exactly the equations in $\mathfrak{Th}'(E)$, where E contains the quantitative equations without quantities of $\hat{E}$ seen as equations.[414]

If we had naively generalized the original logic of [MPP16] by replacing $[0,\infty]$ with an arbitrary complete lattice, this instantiation to 1 would not have been equivalent to equational logic. Indeed, as we explained in §0.3, the judgments of [MPP16], called quantitative inferences, are more general than quantitative equations, and they can express properties which cannot be expressed with equations.[415]

[413] In other words, X and $\mathbf{X}$ are the same thing.

[414] i.e. $E = \{X \vdash s = t \mid \mathbf{X} \vdash s = t \in \hat{E}\}$

[415] The standard example of left-cancellability of a binary operation would be expressed with the quantitative inference

$$x \cdot y = x \cdot z \vdash y = z,$$

but it cannot be expressed with equations. Quantitative inferences are better quantitative versions of the implications in [Wec12, §3.3, Definition 1].

Example 3.57 (Recovering equational logic II)**.** There is a less trivial way to see that equational reasoning faithfully embeds into quantitative equational reasoning.

We are back to the general case of L being an arbitrary complete lattice and $\hat{E}_{\mathbf{GMet}}$ being possibly non-empty. Let E be a class of non-quantitative equations, and let $\hat{E}$ contain every equation in E seen as a quantitative equation with its context being the discrete space, i.e.

$$\hat{E} = \{\mathbf{X}_{\top} \vdash s = t \mid X \vdash s = t \in E\}. \tag{3.43}$$

Claim. If $X \vdash s = t \in \mathfrak{Th}'(E)$, then $\mathbf{X}_{\top} \vdash s = t \in \mathfrak{QTh}'(\hat{E})$.[416]

[416] Depending on the equations inside $\hat{E}_{\mathbf{GMet}}$, it is possible that $\mathfrak{QTh}'(\hat{E})$ contains more equations without quantities than $\mathfrak{Th}'(E)$. Nevertheless, we show that everything you can prove in equational logic can also be proven in QEL.

Proof 1. You can show by induction that a derivation of $X \vdash s = t$ in equational logic with axioms E can be transformed into a derivation of $\mathbf{X}_{\top} \vdash s = t$ in QEL with axioms $\hat{E}$. The base cases are handled by the definition of $\hat{E}$ and the rule Refl in QEL instantiated with the discrete spaces which perfectly emulates the rule Refl in equational logic.

For the inductive step, the rules Symm, Trans, and Cong in equational logic all have perfect counterparts in QEL. The substitution rule needs a bit more work. If the last rule in the derivation in equational logic is

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

then by induction hypothesis, there is a derivation of $\mathbf{Y}_{\top} \vdash s = t$ in QEL. We obtain the following derivation noting that for all $y, y' \in Y$, $d_\top(y, y') = \top$.

$$\dfrac{\sigma : Y \to \mathcal{T}_\Sigma X \qquad \dfrac{\text{I.H.}}{\mathbf{Y}_\top \vdash s = t} \qquad \dfrac{}{\forall y, y' \in Y,\ \mathbf{X}_\top \vdash \sigma(y) =_{d_\top(y,y')} \sigma(y')}\textsc{Top}}{\mathbf{X}_\top \vdash \sigma^*(s) = \sigma^*(t)}\textsc{Sub}$$

□

Proof 2. The proof above reasoning on derivations is useful to get familiar with QEL, but there is a faster *semantic* proof that relies on completeness. By soundness and completeness,[417] it is enough to prove that if $X \vdash s = t \in \mathfrak{Th}(E)$, then $\mathbf{X}_\top \vdash s = t \in \mathfrak{QTh}(\hat{E})$. This follows from the equivalence (3.14) (which was easy to prove):

$$\hat{\mathbb{A}} \vDash \hat{E} \overset{(3.14)}{\iff} \hat{\mathbb{A}} \vDash E \overset{(1.18)}{\implies} \mathbb{A} \vDash X \vdash s = t \overset{(3.14)}{\iff} \hat{\mathbb{A}} \vDash \mathbf{X}_\top \vdash s = t. \qquad \square$$

This second proof also points to a stronger version of the claim that we state as a lemma for future use.

Lemma 3.58. *Let E be a class of non-quantitative equations and $\hat{E}$ be defined as in* (3.43). *If $X \vdash s = t \in \mathfrak{Th}'(E)$, then, for any* L*-space* $\mathbf{X}$ *with carrier* X, $\mathbf{X} \vdash s = t \in \mathfrak{QTh}'(\hat{E})$.[418]

Let us get back to our goal of showing QEL is complete. We follow the proof sketch of completeness for equational logic.[419] We define a quantitative algebra exactly like $\widehat{\mathbb{T}}\mathbf{X}$ but using the equality relation and L-relation induced by $\mathfrak{QTh}'(\hat{E})$ instead of $\mathfrak{QTh}(\hat{E})$, and then we show it satisfies $\hat{E}$ which, by construction, will imply $\mathfrak{QTh}(\hat{E}) \subseteq \mathfrak{QTh}'(\hat{E})$.

Definition 3.59 (Quantitative term algebra, syntactically)**.** The *new* quantitative term algebra for $(\Sigma, \hat{E})$ on $\mathbf{X}$ is the quantitative Σ-algebra whose underlying space is $\mathcal{T}_\Sigma X/{\equiv'_{\hat{E}}}$ equipped with the L-relation corresponding to the L-structure defined in (3.42),[420] and whose interpretation of $\mathsf{op} : n \in \Sigma$ is defined by[421]

$$[\![\mathsf{op}]\!]_{\widehat{\mathbb{T}}'\mathbf{X}}(\wr t_1 \int_{\hat{E}}, \ldots, \wr t_n \int_{\hat{E}}) = \wr \mathsf{op}(t_1, \ldots, t_n) \int_{\hat{E}}. \tag{3.45}$$

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

We will prove this alternative definition of the term algebra coincides with $\widehat{\mathbb{T}}\mathbf{X}$. First, we have to show that $\widehat{\mathbb{T}}'\mathbf{X}$ belongs to $\mathbf{QAlg}(\Sigma, \hat{E})$ like we did for $\widehat{\mathbb{T}}X$ in Proposition 3.47, and we state a technical lemma before that.

Lemma 3.60. *Let $\iota : Y \to \mathcal{T}_\Sigma X/{\equiv'_E}$ be any assignment. For any function $\sigma : Y \to \mathcal{T}_\Sigma X$ satisfying $\wr \sigma(y) \int_{\hat{E}} = \iota(y)$ for all $y \in Y$, we have $[\![-]\!]^\iota_{\widehat{\mathbb{T}}'X} = \wr \sigma^*(-) \int_{\hat{E}}$.*[422]

Proposition 3.61. *For any space* $\mathbf{X}$, $\widehat{\mathbb{T}}'\mathbf{X}$ *satisfies all the equations in* $\hat{E}$.

Proof. Let $\mathbf{Y} \vdash s = t$ (resp. $\mathbf{Y} \vdash s =_\varepsilon t$) belong to $\hat{E}$ and $\hat{\iota} : \mathbf{Y} \to (\mathcal{T}_\Sigma X/{\equiv'_{\hat{E}}}, d'_{\hat{E}})$ be a nonexpansive assignment. By the axiom of choice,[423] there is a function $\sigma : Y \to \mathcal{T}_\Sigma X$ satisfying $\wr \sigma(y) \int_{\hat{E}} = \hat{\iota}(y)$ for all $y \in Y$. Thanks to Lemma 3.60, it is enough to show $\wr \sigma^*(s) \int_{\hat{E}} = \wr \sigma^*(t) \int_{\hat{E}}$ (resp. $d'_{\hat{E}}(\wr \sigma^*(s) \int_{\hat{E}}, \wr \sigma^*(t) \int_{\hat{E}}) \leq \varepsilon$).[424]

[417] Of both equational logic (Theorems 1.43 and 1.48) and QEL (Theorems 3.55 and 3.62).

[418] Follow the second proof above but instead of the second use of (3.14), use Lemma 3.32. (This requires assuming $\mathfrak{QTh}(\hat{E}) = \mathfrak{QTh}'(\hat{E})$ which we prove soon.)

[419] Our proof of completeness for the logic in [MSV22] seems more complex (in my opinion), but it morally follows the same sketch. It is obfuscated however by the fact that [MSV22] did not deal with contexts, instead we were using what we now call syntactic sugar to describe quantitative equations.

[420] Explicitly, it is the L-relation $'d'_{\hat{E}}$ that satisfies

$$d'_{\hat{E}}(\wr s \int_{\hat{E}}, \wr t \int_{\hat{E}}) \leq \varepsilon \iff \mathbf{X} \vdash s =_\varepsilon t \in \mathfrak{QTh}'(\hat{E}). \tag{3.44}$$

[421] This is well-defined (i.e. invariant under change of representative) by (3.40).

[422] The proof goes as in the classical case (Lemma 1.46). We do not even need to ask ι to be nonexpansive, but we will use the result with a nonexpansive assignment.

[423] Choice implies the quotient map $\wr - \int_{\hat{E}}$ has a right inverse $r : \mathcal{T}_\Sigma X/{\equiv'_{\hat{E}}} \to \mathcal{T}_\Sigma X$, and we set $\sigma = r \circ \hat{\iota}$.

[424] By Lemma 3.60, it implies

$$[\![s]\!]^{\hat{\iota}}_{\widehat{\mathbb{T}}'\mathbf{X}} = \wr \sigma^*(s) \int_{\hat{E}} = \wr \sigma^*(t) \int_{\hat{E}} = [\![t]\!]^{\hat{\iota}}_{\widehat{\mathbb{T}}'X},$$

resp. $d'_{\hat{E}}([\![s]\!]^{\hat{\iota}}_{\widehat{\mathbb{T}}'\mathbf{X}}, [\![t]\!]^{\hat{\iota}}_{\widehat{\mathbb{T}}'\mathbf{X}}) = d'_{\hat{E}}(\wr \sigma^*(s) \int_{\hat{E}}, \wr \sigma^*(t) \int_{\hat{E}}) \leq \varepsilon$

and since $\hat{\iota}$ was arbitrary, we conclude that $\widehat{\mathbb{T}}'\mathbf{X}$ satisfies $\mathbf{Y} \vdash s = t$ (resp. $\mathbf{Y} \vdash s =_\varepsilon t$).

Equivalently, by definition of $\lbag - \rbag_{\hat{E}}$ and $\mathfrak{QTh}'(\hat{E})$, we can just exhibit a derivation of $\mathbf{X} \vdash \sigma^*(s) = \sigma^*(t)$ (resp. $\mathbf{X} \vdash \sigma^*(s) =_\varepsilon \sigma^*(t)$) in QEL with axioms $\hat{E}$. That equation can be proven with the SUB (resp. SUBQ) rule instantiated with $\sigma : Y \to \mathcal{T}_\Sigma X$ and the equation $\mathbf{Y} \vdash s = t$ (resp. $\mathbf{Y} \vdash s =_\varepsilon$) which is an axiom, but we need derivations showing σ satisfies the side conditions of the substitution rules. This follows from nonexpansiveness of $\hat{\imath}$ because for any $y, y' \in Y$, we know that

$$d_{\hat{E}}(\lbag \sigma(y) \rbag_{\hat{E}}, \lbag \sigma(y) \rbag_{\hat{E}}) = d_{\hat{E}}(\hat{\imath}(y), \hat{\imath}(y')) \leq d_{\mathbf{Y}}(y, y'),$$

which means by (3.44) that $\mathbf{X} \vdash \sigma(y) =_{d_{\mathbf{Y}}(y,y')} \sigma(y)$ belongs to $\mathfrak{QTh}'(\hat{E})$. $\square$

Completeness of quantitative equational logic readily follows.

Theorem 3.62 (Completeness)**.** *If $\phi \in \mathfrak{QTh}(\hat{E})$, then $\phi \in \mathfrak{QTh}'(\hat{E})$.*

Proof. Let $\phi \in \mathfrak{QTh}(\hat{E})$ and $\mathbf{X}$ be its context. By Proposition 3.61 and definition of $\mathfrak{QTh}(\hat{E})$, we know that $\widehat{\mathbb{T}}'\mathbf{X} \vDash \phi$. In particular, $\widehat{\mathbb{T}}'\mathbf{X}$ satisfies ϕ under the assignment

$$\hat{\imath} = \mathbf{X} \xrightarrow{\eta^\Sigma_X} \mathcal{T}_\Sigma X \xrightarrow{\lbag - \rbag_{\hat{E}}} \mathcal{T}_\Sigma X/{\equiv'_{\hat{E}}},$$

which is nonexpansive by VARS.[425]

Moreover with $\sigma = \eta^\Sigma_X$, we can show σ satisfies the hypothesis of Lemma 3.60 and $\sigma^* = \mathrm{id}_{\mathcal{T}_\Sigma X}$,[426] thus we conclude

- if $\phi = \mathbf{X} \vdash s = t$: $\lbag s \rbag_{\hat{E}} = [\![s]\!]^{\hat{\imath}}_{\widehat{\mathbb{T}}'\mathbf{X}} = [\![t]\!]^{\hat{\imath}}_{\widehat{\mathbb{T}}'\mathbf{X}} = \lbag t \rbag_{\hat{E}}$, and
- if $\phi = \mathbf{X} \vdash s =_\varepsilon t$: $d'_{\hat{E}}(\lbag s \rbag_{\hat{E}}, \lbag t \rbag_{\hat{E}}) = d'_{\hat{E}}([\![s]\!]^{\hat{\imath}}_{\widehat{\mathbb{T}}'\mathbf{X}}, [\![t]\!]^{\hat{\imath}}_{\widehat{\mathbb{T}}'\mathbf{X}}) \leq \varepsilon$.

By definition of $\equiv'_{\hat{E}}$ (3.39) and $d'_{\hat{E}}$ (3.44), this implies $\mathbf{X} \vdash s = t$ (resp. $\mathbf{X} \vdash s =_\varepsilon t$) belongs to $\mathfrak{QTh}'(\hat{E})$. $\square$

Note that because $\widehat{\mathbb{T}}\mathbf{X}$ and $\widehat{\mathbb{T}}'\mathbf{X}$ were defined in the same way in terms of $\mathfrak{QTh}(\hat{E})$ and $\mathfrak{QTh}'(\hat{E})$ respectively, and since we have proven the latter to be equal, we obtain that $\widehat{\mathbb{T}}\mathbf{X}$ and $\widehat{\mathbb{T}}'\mathbf{X}$ are the same quantitative algebra. In the sequel, we will work with $\widehat{\mathbb{T}}\mathbf{X}$ mostly but we may use the facts that $s \equiv_{\hat{E}} t$ (resp. $d_{\hat{E}}(s,t) \leq \varepsilon$) if and only if there is a derivation of $\mathbf{X} \vdash s = t$ (resp. $\mathbf{X} \vdash s =_\varepsilon t$) in QEL.[427]

Remark 3.63. Mirroring Remark 1.49, we would like to say that the axiom of choice was not necessary in the proofs above. Unfortunately, this situation is more delicate, and I do not know for sure that we can avoid using choice (although I expect we can).

At first, you might think that since terms are still finite, we can still restrict the context to the free variables which is finite. Unfortunately, even if $x \in \mathrm{FV}\{s,t\}$ and $y \notin \mathrm{FV}\{s,t\}$, it is possible that the distance between x and y in the context is necessary to state the right property. Here is an example that we carry with $\mathbf{GMet} = [0,1]\mathbf{Spa}$, $\Sigma = \emptyset$, and $\hat{E}$ defining discrete metrics:[428]

$$\hat{E} = \{x =_\varepsilon y \vdash x = y \mid 1 \neq \varepsilon \in \mathsf{L}\} \cup \{x = y \vdash x =_0 y\}.$$

[425] Explicitly, VARS means $\mathbf{X} \vdash x =_{d_{\mathbf{X}}(x,x')} x'$ belongs to $\mathfrak{QTh}'(\hat{E})$, hence, (3.44) implies

$$d'_{\hat{E}}(\lbag x \rbag_{\hat{E}}, \lbag x' \rbag_{\hat{E}}) \leq d_{\mathbf{X}}(x, x').$$

[426] We defined $\hat{\imath}$ precisely to have $\lbag \eta^\Sigma_X(x) \rbag_{\hat{E}} = \hat{\imath}(x)$. To show $\sigma^* = \eta^{\Sigma *}_X$ is the identity, use (1.35) and the fact that $\mu^\Sigma \cdot \eta^\Sigma \mathcal{T}_\Sigma = \mathbb{1}_{\mathcal{T}_\Sigma}$ (it holds by definition (1.6)).

[427] i.e. when proving that an equation holds in some theory $\mathfrak{QTh}(\hat{E})$, we can either use the rules of QEL or the several lemmas from §3.1 which are morally the semantic counterparts to the inference rules.

[428] When $d_{\mathbf{A}}(a,b)$ is not 1, it must be that $a = b$ by the first set of equations, by the second set, it must be that $d_{\mathbf{A}}(a,b) = 0$. Under such constraints $\mathbf{A}$ must be the discrete metric on A that we described in Example 3.50, so $\mathbf{QAlg}(\emptyset, \hat{E})$ is the category of discrete metrics.

Let $\mathbf{X} = \{x, z\}$ and $\mathbf{Y} = \{x, y, z\}$ with the following distances ($\mathbf{X}$ is a subspace of $\mathbf{Y}$):

$$\begin{tikzcd} x \arrow[loop above, no head, "0"] \arrow[r, no head, "\frac{1}{2}"] & y \arrow[loop above, no head, "0"] \arrow[r, no head, "\frac{1}{2}"] & z \arrow[loop above, no head, "0"] \end{tikzcd}$$

The equation $\mathbf{Y} \vdash x = z$ belongs to $\mathfrak{QTh}(\hat{E})$. Indeed, if $\mathbf{A} \vDash \hat{E}$, then $d_{\mathbf{A}}(a, b) \leq \frac{1}{2}$ implies $a = b$, so any nonexpansive assignment $\hat{\iota} : \mathbf{Y} \to \mathbf{A}$ must identify x and y, and y and z, hence $\hat{\iota}(x) = \hat{\iota}(z)$. However, the equation $\mathbf{X} \vdash x = z$ is not in $\mathfrak{QTh}(\hat{E})$ because you can have $d_{\mathbf{A}}(\hat{\iota}(x), \hat{\iota}(z)) \leq 1$ without $\hat{\iota}(x) = \hat{\iota}(z)$.

This shows that some variables in the context which are not used in the terms of the equation (in this instance y) might still be important. One may still wonder whether it is possible to restrict the contexts to be finite or countable.[429] I do not know if that is true, but I expect that countable contexts are enough and that finite contexts are not.

[429] i.e. for any equation ϕ, is there an equation ψ with finite (or countable) context such that

$$\hat{\mathbb{A}} \vDash \phi \iff \hat{\mathbb{A}} \vDash \psi.$$

In summary, while there can be an analog to the derivable Add rule in equational logic, the obvious counterpart to the Del rule is not even sound.

Let us highlight one last feature of quantitative equational logic: the rule GMet defining what kind of generalized metric spaces are considered is independent of all the other rules.[430] As a consequence, and we give more details in [MSV23, §8], you can choose to work over $\mathsf{L}\mathbf{Spa}$ all the time and add the equations in $\hat{E}_{\mathbf{GMet}}$ as axioms in $\hat{E}$ anytime you wish to restrict to algebras whose carriers are generalized metric spaces. Written a bit ambiguously,[431]

$$\mathbf{QAlg}(\Sigma, \hat{E}) = \mathbf{QAlg}(\Sigma, \hat{E} \cup \hat{E}_{\mathbf{GMet}}) \quad \text{and} \quad \mathfrak{QTh}(\hat{E}) = \mathfrak{QTh}(\hat{E} \cup \hat{E}_{\mathbf{GMet}}). \tag{3.46}$$

[430] Although it was less explicit because only **Met** was considered, this was already a feature of the logic in [MPP16].

[431] What we really mean is that on the left, **QAlg** and $\mathfrak{QTh}$ are the operators we described with the parameter **GMet** built in, and on the right, they are the same operators instantiated with $\mathsf{L}\mathbf{Spa}$ instead.

3.3 Quantitative Algebraic Presentations

In order to obtain a more categorical understanding of quantitative algebras, a first step is to show that the functor $\widehat{\mathcal{T}}_{\Sigma,\hat{E}} : \mathbf{GMet} \to \mathbf{GMet}$ we constructed is a monad.

Proposition 3.64. *The functor $\widehat{\mathcal{T}}_{\Sigma,\hat{E}} : \mathbf{GMet} \to \mathbf{GMet}$ defines a monad on $\mathbf{GMet}$ with unit $\widehat{\eta}^{\Sigma,\hat{E}}$ and multiplication $\widehat{\mu}^{\Sigma,\hat{E}}$. We call it the **term monad** for $(\Sigma, \hat{E})$.*

Proof. A first proof uses a standard result of category theory. Since we showed that $\widehat{\mathbb{T}}_{\Sigma,\hat{E}}\mathbf{A}$ is the free $(\Sigma, \hat{E})$-algebra on $\mathbf{A}$ for every space $\mathbf{A}$ (Theorem 3.48), we obtain a monad sending $\mathbf{A}$ to the underlying space of $\widehat{\mathbb{T}}_{\Sigma,\hat{E}}\mathbf{A}$, i.e. $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A}$.[432]

[432] The unit is automatically $\widehat{\eta}^{\Sigma,\hat{E}}$, but some computations are needed to show the multiplication is $\widehat{\mu}^{\Sigma,\hat{E}}$.

One could also follow the proof we gave for **Set** and explicitly show that $\widehat{\eta}^{\Sigma,\hat{E}}$ and $\widehat{\mu}^{\Sigma,\hat{E}}$ obey the laws for the unit and multiplication (most of the work having been done earlier in this chapter). $\square$

What is arguably more important is that quantitative $(\Sigma, \hat{E})$-algebras on a space $\mathbf{A}$ correspond to $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}$-algebras on $\mathbf{A}$.[433] We construct an isomorphism between $\mathbf{QAlg}(\Sigma, \hat{E})$ and $\mathbf{EM}(\widehat{\mathcal{T}}_{\Sigma,\hat{E}})$ using the isomorphism $P : \mathbf{Alg}(\Sigma) \cong \mathbf{EM}(\mathcal{T}_{\Sigma}) : P^{-1}$ that we defined in Proposition 1.58,[434] the forgetful functor $U : \mathbf{QAlg}(\Sigma, \hat{E}) \to \mathbf{Alg}(\Sigma)$ that sends $\hat{\mathbb{A}}$ to the underlying algebra $\mathbb{A}$, and the functor $\mathbf{EM}(\widehat{\mathcal{T}}_{\Sigma,\hat{E}}) \to \mathbf{EM}(\mathcal{T}_{\Sigma})$ we define below.

[433] i.e. $U : \mathbf{QAlg}(\Sigma, \hat{E}) \to \mathbf{GMet}$ is monadic.

[434] Take the statement of Proposition 1.58 with $E = \emptyset$.

Lemma 3.65. *For any $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}$-algebra (A,α), the map $U\alpha \circ [-]_{\hat{E}} : \mathcal{T}_\Sigma A \to A$ is a $\mathcal{T}_\Sigma$-algebra. Furthermore, this defines a functor $U^{[-]_{\hat{E}}} : \mathbf{EM}(\widehat{\mathcal{T}}_{\Sigma,\hat{E}}) \to \mathbf{EM}(\mathcal{T}_\Sigma)$.*

Proof. Apply Proposition 1.70 after checking that $(U, [-]_{\hat{E}})$ is monad functor from $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}$ to $\mathcal{T}_\Sigma$.[435] □

[435] The appropriate diagrams (1.58) and (1.59) commute by (3.33) and a combination of (3.20) and (3.21).

Theorem 3.66. *There is an isomorphism $\mathbf{QAlg}(\Sigma,\hat{E}) \cong \mathbf{EM}(\widehat{\mathcal{T}}_{\Sigma,\hat{E}})$.*

Proof. In the diagram below, we already have the functors drawn with solid arrows, and we want to construct $\widehat{P}$ and $\widehat{P}^{-1}$ drawn with dashed arrows before proving they are inverses to each other.

$$\begin{tikzcd}
\mathbf{QAlg}(\Sigma,\hat{E}) \arrow[r, dashed, shift left, "\widehat{P}"] \arrow[d, "U"'] & \mathbf{EM}(\widehat{\mathcal{T}}_{\Sigma,\hat{E}}) \arrow[l, dashed, shift left, "\widehat{P}^{-1}"] \arrow[d, "U^{[-]_{\hat{E}}}"] \\
\mathbf{Alg}(\Sigma) \arrow[r, shift left, "P"] & \mathbf{EM}(\mathcal{T}_\Sigma) \arrow[l, shift left, "P^{-1}"]
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathbf{QAlg}(\Sigma,\hat{E}) \arrow[rr, dashed, shift left, "\widehat{P}"] \arrow[dr, "U"'] & & \mathbf{EM}(\widehat{\mathcal{T}}_{\Sigma,\hat{E}}) \arrow[ll, dashed, shift left, "\widehat{P}^{-1}"] \arrow[dl, "U^{\widehat{\mathcal{T}}_{\Sigma,\hat{E}}}"] \\
& \mathbf{GMet} &
\end{tikzcd}$$

A (meaningful) sidequest for us is to make the diagrams above commute, namely, the underlying $\mathcal{T}_\Sigma$-algebra of $\widehat{P}\hat{\mathbb{A}}$ should be $P\mathbb{A}$ and the underlying space of $\widehat{P}\hat{\mathbb{A}}$ should be the underlying space of $\hat{\mathbb{A}}$, and similarly for $\widehat{P}^{-1}$. It turns out this completely determines our functors, up to some quick checks. We will move between spaces and their underlying sets without indicating it by $U : \mathbf{GMet} \to \mathbf{Set}$.

Given $\hat{\mathbb{A}} \in \mathbf{QAlg}(\Sigma,\hat{E})$, we look at the underlying Σ-algebra $\mathbb{A}$, apply P to it to get $\alpha_{\mathbb{A}} : \mathcal{T}_\Sigma A \to A$ which sends a term t to its interpretation $[\![t]\!]_A$, and we need to check that it factors through $[-]_{\hat{E}}$ and a nonexpansive map $\widehat{\alpha}_{\hat{\mathbb{A}}}$ as in (3.47).

$$\begin{tikzcd}
\mathcal{T}_\Sigma A \arrow[rr, "\alpha_{\mathbb{A}}"] \arrow[dr, "{[-]_{\hat{E}}}"'] & & \mathbf{A} \\
& \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A} \arrow[ur, "\widehat{\alpha}_{\hat{\mathbb{A}}}"'] &
\end{tikzcd} \tag{3.47}$$

First, $\alpha_{\mathbb{A}}$ is well-defined on terms modulo $\hat{E}$ because if $s \equiv_{\hat{E}} t$, then $\hat{\mathbb{A}}$ satisfies $\mathbf{A} \vdash s = t \in \mathfrak{QTh}(\hat{E})$, and this in turn means (taking the assignment $\mathrm{id}_{\mathbf{A}} : \mathbf{A} \to \mathbf{A}$):

$$\alpha_{\mathbb{A}}(s) = [\![s]\!]_A = [\![s]\!]_A^{\mathrm{id}_{\mathbf{A}}} = [\![t]\!]_A^{\mathrm{id}_{\mathbf{A}}} = [\![t]\!]_A = \alpha_{\mathbb{A}}(t).$$

Next, the factor we obtain $\widehat{\alpha}_{\hat{\mathbb{A}}} : \mathcal{T}_\Sigma A/{\equiv_{\hat{E}}} \to A$ is nonexpansive from $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A}$ to $\mathbf{A}$. Indeed, if $d_{\hat{E}}([s]_{\hat{E}}, [t]_{\hat{E}}) \leq \varepsilon$, then $\hat{\mathbb{A}}$ satisfies $\mathbf{A} \vdash s =_\varepsilon t \in \mathfrak{QTh}(\hat{E})$, and this means:

$$d_{\mathbf{A}}(\widehat{\alpha}_{\hat{\mathbb{A}}}[s]_{\hat{E}}, \widehat{\alpha}_{\hat{\mathbb{A}}}[t]_{\hat{E}}) = d_{\mathbf{A}}(\alpha_{\mathbb{A}}(s), \alpha_{\mathbb{A}}(t)) = d_{\mathbf{A}}([\![s]\!]_A, [\![t]\!]_A) = d_{\mathbf{A}}([\![s]\!]_A^{\mathrm{id}_{\mathbf{A}}}, [\![t]\!]_A^{\mathrm{id}_{\mathbf{A}}}) \leq \varepsilon.$$

Finally, if $h : \hat{\mathbb{A}} \to \hat{\mathbb{B}}$ is a homomorphism, then by definition it is nonexpansive $\mathbf{A} \to \mathbf{B}$ and it commutes with $[\![-]\!]_A$ and $[\![-]\!]_B$. The latter means it commutes with $\alpha_{\mathbb{A}}$ and $\alpha_{\mathbb{B}}$, which in turn means it commutes with $\widehat{\alpha}_{\hat{\mathbb{A}}}$ and $\widehat{\alpha}_{\hat{\mathbb{B}}}$ because $[-]_{\hat{E}}$ is epic (see (3.48)). We obtain our functor $\widehat{P} : \mathbf{QAlg}(\Sigma,\hat{E}) \to \mathbf{EM}(\widehat{\mathcal{T}}_{\Sigma,\hat{E}})$.

$$\begin{tikzcd}
\mathcal{T}_\Sigma A \arrow[rr, "\mathcal{T}_\Sigma h"] \arrow[drr, "\alpha_{\mathbb{A}}"] \arrow[ddr, "{[-]_{\hat{E}}}"'] & & \mathcal{T}_\Sigma B \arrow[drr, "\alpha_{\mathbb{B}}"] \arrow[ddr, "{[-]_{\hat{E}}}"'] & & \\
& & \mathbf{A} \arrow[rr, "h"] & & \mathbf{B} \\
& \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A} \arrow[ur, "\widehat{\alpha}_{\hat{\mathbb{A}}}"'] \arrow[rr, "\widehat{\mathcal{T}}_{\Sigma,\hat{E}}h"'] & & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{B} \arrow[ur, "\widehat{\alpha}_{\hat{\mathbb{B}}}"'] &
\end{tikzcd} \tag{3.48}$$

The top face of the prism in (3.48) commutes because h is a homomorphism, the back face commutes by (3.18), and the side faces commute by (3.47). Thus, the bottom face commutes because $[-]_{\hat{E}}$ is epic.

Given a $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}$-algebra $\widehat{\alpha} : \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{A} \to \mathbf{A}$, we look at the $\mathcal{T}_\Sigma$-algebra

$$U^{[-]_{\hat{E}}}\widehat{\alpha} = U\widehat{\alpha} \circ [-]_{\hat{E}} : \mathcal{T}_\Sigma A \to A$$

obtained via Lemma 3.65, then we apply P^{-1} to get the Σ-algebra $(A, [\![-]\!]_{U^{[-]_{\hat{E}}}\widehat{\alpha}})$. Since $\mathbf{A} = (A, d_{\mathbf{A}})$ is a generalized metric space (because $\widehat{\alpha}$ belongs to $\mathbf{EM}(\widehat{\mathcal{T}}_{\Sigma,\hat{E}})$), we obtain a quantitative algebra $\hat{\mathbb{A}}_{\widehat{\alpha}} = (A, [\![-]\!]_{U^{[-]_{\hat{E}}}\widehat{\alpha}}, d_{\mathbf{A}})$, and we need to check it satisfies the equations in $\hat{E}$.

Recall from the proof of Proposition 1.58 that interpreting terms in $\hat{\mathbb{A}}_{\widehat{\alpha}}$ is the same thing as applying $U^{[-]_{\hat{E}}}\widehat{\alpha} = U\widehat{\alpha} \circ [-]_{\hat{E}}$. Therefore, given any L-space $\mathbf{X}$, nonexpansive assignment $\hat{\iota} : \mathbf{X} \to \mathbf{A}$, and $t \in \mathcal{T}_\Sigma X$, we have

$$[\![t]\!]^{\hat{\iota}}_{U^{[-]_{\hat{E}}}} \stackrel{(1.9)}{=} [\![\mathcal{T}_\Sigma \hat{\iota}(t)]\!]_{U^{[-]_{\hat{E}}}} = \widehat{\alpha}[\mathcal{T}_\Sigma \hat{\iota}(t)]_{\hat{E}}.$$

Now, if $\mathbf{X} \vdash s = t \in \hat{E}$, we also have $\mathbf{A} \vdash \mathcal{T}_\Sigma \hat{\iota}(s) = \mathcal{T}_\Sigma \hat{\iota}(t) \in \mathfrak{QTh}(\hat{E})$ by Lemma 3.30, which means

$$[\![s]\!]^{\hat{\iota}}_{U^{[-]_{\hat{E}}}} = \widehat{\alpha}[\mathcal{T}_\Sigma \hat{\iota}(s)]_{\hat{E}} = \widehat{\alpha}[\mathcal{T}_\Sigma \hat{\iota}(t)]_{\hat{E}} = [\![t]\!]^{\hat{\iota}}_{U^{[-]_{\hat{E}}}}.$$

Similarly for $\mathbf{X} \vdash s =_\varepsilon t \in \hat{E}$, Lemma 3.30 means $\mathbf{A} \vdash \mathcal{T}_\Sigma \hat{\iota}(s) =_\varepsilon \mathcal{T}_\Sigma \hat{\iota}(t) \in \mathfrak{QTh}(\hat{E})$, so[436]

$$d_{\mathbf{A}}([\![s]\!]^{\hat{\iota}}_{U^{[-]_{\hat{E}}}}, [\![t]\!]^{\hat{\iota}}_{U^{[-]_{\hat{E}}}}) = d_{\mathbf{A}}(\widehat{\alpha}[\mathcal{T}_\Sigma \hat{\iota}(s)]_{\hat{E}}, \widehat{\alpha}[\mathcal{T}_\Sigma \hat{\iota}(t)]_{\hat{E}}) \leq d_{\hat{E}}([\mathcal{T}_\Sigma \hat{\iota}(s)]_{\hat{E}}, [\mathcal{T}_\Sigma \hat{\iota}(t)]_{\hat{E}}) \leq \varepsilon.$$

[436] The first inequality holds by nonexpansiveness of $\widehat{\alpha}$ and the second by definition of $d_{\hat{E}}$ (3.16).

Finally, if $h : (\mathbf{A}, \widehat{\alpha}) \to (\mathbf{B}, \widehat{\beta})$ is $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}$-homomorphism, then by definition, it is nonexpansive $\mathbf{A} \to \mathbf{B}$, and by Lemma 3.65 it commutes with $U^{[-]_{\hat{E}}}\widehat{\alpha}$ and $U^{[-]_{\hat{E}}}\widehat{\beta}$ which means it is a homomorphism of the underlying algebras of $\hat{\mathbb{A}}_{\widehat{\alpha}}$ and $\hat{\mathbb{B}}_{\widehat{\beta}}$. We conclude it is also a homomorphism between the quantitative algebras $\hat{\mathbb{A}}_{\widehat{\alpha}}$ and $\hat{\mathbb{B}}_{\widehat{\beta}}$.[437] We obtain our functor $\widehat{P}^{-1} : \mathbf{EM}(\widehat{\mathcal{T}}_{\Sigma,\hat{E}}) \to \mathbf{QAlg}(\Sigma, \hat{E})$.

[437] Recall that homomorphisms between quantitative algebras are just nonexpansive homomorphisms.

The diagrams at the start of the proof commute by construction, and P and P^{-1} are inverses by Proposition 1.58. That is enough to conclude that $\widehat{P}$ and $\widehat{P}^{-1}$ are also inverses. Indeed, by commutativity of the triangle, $\widehat{P}$ and $\widehat{P}^{-1}$ preserve the underlying spaces, and if we fix a space $\mathbf{A}$, the forgetful functors U and $U^{[-]_{\hat{E}}}$ are injective.[438] Then, still with a fixed space $\mathbf{A}$, by commutativity of the square, we have

[438] For U, it is clear because it only forgets the L-relation. For $U^{[-]_{\hat{E}}}$, it is also not too hard to see, and it is because $U : \mathbf{GMet} \to \mathbf{Set}$ is faithful and $[-]_{\hat{E}}$ is epic.

$$U\widehat{P}^{-1}\widehat{P}\hat{\mathbb{A}} = P^{-1}U^{[-]_{\hat{E}}}\widehat{P}\hat{\mathbb{A}} = P^{-1}PU\hat{\mathbb{A}} = U\hat{\mathbb{A}}, \text{ and}$$
$$U^{[-]_{\hat{E}}}\widehat{P}\widehat{P}^{-1}\widehat{\alpha} = PU\widehat{P}^{-1}\widehat{\alpha} = PP^{-1}U^{[-]_{\hat{E}}}\widehat{\alpha} = U^{[-]_{\hat{E}}}\widehat{\alpha},$$

with which we can conclude by injectivity of U and $U^{[-]_{\hat{E}}}$. □

Remark 3.67. We followed the proof of [MSV22] which does not rely on monadicity theorems (c.f. Remark 1.59).[439] To show that $U : \mathbf{QAlg}(\Sigma, \hat{E}) \to \mathbf{GMet}$ is (strictly) monadic, it would be enough to check that the isomorphism we constructed above is the comparison functor.

[439] For a proof that does, see [MSV23, Theorems 6.3 and 8.10] where we showed strict monadicity for $[0,1]$-spaces first, then for generalized metric spaces using (3.46), and the cancellability of monadicity [Bou92, Proposition 5].

This motivates the following definition.

Definition 3.68 (**GMet** presentation)**.** Let M be a monad on **GMet**, a **quantitative algebraic presentation** of M is signature Σ and a class of quantitative equations $\hat{E}$ along with a monad isomorphism $\rho : \widehat{\mathcal{T}}_{\Sigma,\hat{E}} \cong M$. We also say M is presented by $(\Sigma, \hat{E})$. By Proposition 1.64 and Theorem 3.66, this is equivalent to having an isomorphism $\mathbf{EM}(\widehat{\mathcal{T}}_{\Sigma,\hat{E}}) \cong \mathbf{QAlg}(\Sigma, \hat{E})$ that commutes with the forgetful functors.

Example 3.69 (Hausdorff)**.** We saw in Example 1.66 that the monad $\mathcal{P}_{\mathrm{ne}}$ on **Set** is presented by the theory of semilattices. In this example,[440] we define the theory of quantitative semilattices and show it presents a monad which sends (X, d) to $\mathcal{P}_{\mathrm{ne}}X$ equipped with the Hausdorff distance $d^\uparrow$.

[440] We adapted it from [MPP16, §9.1].

A **quantitative semilattice** is a semilattice (i.e. a $(\Sigma_{\mathbf{S}}, E_{\mathbf{S}})$-algebra) equipped with an L-relation such that the interpretation of the semilattice operation is nonexpansive with respect to the product distance. Equivalently, it is a quantitative $\Sigma_{\mathbf{S}}$-algebra that satisfies $\hat{E}_{\mathbf{S}}$ which contains:[441]

$$\begin{aligned} x &\vdash x = x \oplus x \\ x, y &\vdash x \oplus y = y \oplus x \\ x, y, z &\vdash x \oplus (y \oplus z) = (x \oplus y) \oplus z \\ \forall \varepsilon, \varepsilon' \in \mathsf{L}, \quad x =_{\varepsilon} y, x' =_{\varepsilon'} y' &\vdash x \oplus x' =_{\max\{\varepsilon, \varepsilon'\}} y \oplus y' \end{aligned}$$

[441] The first three equations are those of $E_{\mathbf{S}}$ seen with the discrete context as in Example 3.57. The last row is (3.6) which enforces the nonexpansiveness property of $[\![\oplus]\!]$.

We can give an alternative description of the free quantitative semilattice.

Lemma 3.70. *The free quantitative semilattice on (X, d) is $\hat{\mathbb{P}}_{(X,d)} = (\mathcal{P}_{\mathrm{ne}} X, \cup, d^{\uparrow})$.*[442]

[442] This corresponds to [MPP16, Theorem 9.3].

Proof. We know from Example 1.66 that $(\mathcal{P}_{\mathrm{ne}} X, \cup)$ is the free semilattice and hence satisfies $E_{\mathbf{S}}$, thus by Lemma 3.32, $\hat{\mathbb{P}}_{(X,d)}$ satisfies the first three equations above. We already mentioned that $\hat{\mathbb{P}}_{(X,d)}$ satisfies (3.6) because it satisfies (3.1).[443] Thus, $\hat{\mathbb{P}}_{(X,d)}$ is a quantitative semilattice.

[443] We did not give a proof for (3.1).

Let $\hat{\mathbb{A}}$ be a quantitative semilattice and $f : (X, d) \to \mathbf{A}$ be a nonexpansive map. By Lemma 3.33, $\mathbb{A}$ is a semilattice, hence the universal property of the free semilattice gives a unique homomorphism of $(\Sigma_{\mathbf{S}}, E_{\mathbf{S}})$-algebras $f^* : (\mathcal{P}_{\mathrm{ne}} X, \cup) \to \mathbb{A}$ such that $f^*(\{x\}) = f(x)$ for all $x \in X$. It remains to show that f^* is a nonexpansive map $(\mathcal{P}_{\mathrm{ne}} X, d^{\uparrow}) \to \mathbf{A}$.[444]

[444] Actually, you also have to prove that $\eta : (X, d) \to (\mathcal{P}_{\mathrm{ne}} X, d^{\uparrow})$ sending x to $\{x\}$ is nonexpansive. This is easy to check.

Let $S, T \in \mathcal{P}_{\mathrm{ne}} X$, $C \in \mathcal{P}_{\mathrm{ne}}(X \times X)$ be a coupling for S and T, and suppose C is ordered with $C = \{c_1, \dots, c_n\}$. In particular, we have $S = \pi_1(c_1) \cup \cdots \cup \pi_1(c_n)$ and $T = \pi_2(c_1) \cup \cdots \cup \pi_2(c_n)$. Since f^* is a homomorphism of semilattices, this implies

$$\begin{aligned} f^*(S) &= f(\pi_1(c_1)) [\![\oplus]\!]_A \cdots [\![\oplus]\!]_A f(\pi_1(c_n)), \text{and} \\ f^*(T) &= f(\pi_2(c_1)) [\![\oplus]\!]_A \cdots [\![\oplus]\!]_A f(\pi_2(c_n)). \end{aligned}$$

Now, we can use the fact that $\hat{\mathbb{A}}$ satisfies the equations in (3.6) n times in the first step of the following derivation.

$$\begin{aligned} d_{\mathbf{A}}(f^*(S), f^*(T)) &\leq \max_{1 \leq i \leq n} d_{\mathbf{A}}(f(\pi_1(c_i)), f(\pi_2(c_i))) && \text{by (3.6)} \\ &\leq \max_{1 \leq i \leq n} d(\pi_1(c_i), \pi_2(c_i)) && f \text{ nonexpansive} \\ &\leq d^{\downarrow}(S, T) && \text{definition of } d^{\downarrow} \\ &= d^{\uparrow}(S, T) && \text{Lemma 2.17} \end{aligned}$$

We conclude that f^* is a homomorphism between the quantitative algebras $\hat{\mathbb{P}}_{(X,d)}$ and $\hat{\mathbb{A}}$. The uniqueness follows from it being unique as a homomorphism of semilattices and the faithfulness of $U : \mathbf{QAlg}(\Sigma_{\mathbf{S}}, \hat{E}_{\mathbf{S}}) \to \mathbf{Alg}(\Sigma_{\mathbf{S}})$. □

Since $\widehat{\mathbb{T}}(X, d)$ is also the free quantitative semilattice on (X, d) by Theorem 3.48 and free objects are unique by Proposition 1.39, there is an isomorphism of quantitative algebras $\rho_{(X,d)} : \widehat{\mathbb{T}}(X, d) \cong \hat{\mathbb{P}}_{(X,d)}$. After some abstract categorical arguments we do not reproduce, one finds that ρ is a monad isomorphism $\widehat{\mathcal{T}}_{\Sigma_{\mathbf{S}},\hat{E}_{\mathbf{S}}} \cong \mathcal{P}_{\mathrm{ne}}^{\uparrow}$, where $\mathcal{P}_{\mathrm{ne}}^{\uparrow} : \mathbf{GMet} \to \mathbf{GMet}$ sends (X,d) to $(\mathcal{P}_{\mathrm{ne}}X, d^{\uparrow})$ and its unit and multiplication act just like those of $\mathcal{P}_{\mathrm{ne}}$.[445]

[445] This monad is famous independently of quantitative algebras, variations of it were studied in, e.g., [ACT10, §4], [Tho12, §4], [BBKK18, Example 8.3], and [DFM23, §6].

The second example of presentation is from [MPP16, §10.1].

Example 3.71 (Kantorovich)**.** We saw in Example 1.67 that the monad $\mathcal{D}$ on **Set** is presented by the theory of convex algebras. Let $\mathsf{L} = [0,\infty]$ and $\mathbf{GMet} = \mathbf{Met}$. The theory of **quantitative convex algebras** is generated by $\hat{E}_{\mathbf{CA}}$ which contains the equations of $E_{\mathbf{CA}}$ seen as quantitative equations (as explained in Example 3.57) and the quantitative equations for convexity (3.10).[446]

[446] As a reminder, $\hat{E}_{\mathbf{CA}}$ contains

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

Let $(\mathcal{D}X, [\![-]\!]_{\mathcal{D}X})$ be the free convex algebra, where $+_p$ is interpreted as convex combination of distributions (1.57). Thanks to Lemma 3.32, we know that for any metric d on X, we can equip $\mathcal{D}X$ with the Kantorovich distance d_{K} and obtain a quantitative algebra $(\mathcal{D}X, [\![-]\!]_{\mathcal{D}X}, d_{\mathrm{K}})$ that satisfies the equations of convex algebras (seen with a discrete context). Moreover, with Example 3.12 we can infer that $(\mathcal{D}X, [\![-]\!]_{\mathcal{D}X}, d_{\mathrm{K}})$ is a quantitative convex algebra (i.e. it also satisfies (3.10)). In [MPP16, Theorem 10.5], the authors show that, along with the map $\eta_X^{\mathcal{D}} : (X,d) \to (\mathcal{D}X, d_{\mathrm{K}})$ sending x to the Dirac distribution on x, it is the free quantitative convex algebra on (X,d).

We can conclude that $(\Sigma_{\mathbf{CA}}, \hat{E}_{\mathbf{CA}})$ presents a monad $\mathcal{D}_{\mathrm{K}} : \mathbf{Met} \to \mathbf{Met}$ which sends (X,d) to $(\mathcal{D}X, d_{\mathrm{K}})$ and whose unit and multiplication act just like those of the **Set** monad $\mathcal{D}$.[447]

[447] This monad is famous independently of quantitative algebras, variations of it were studied in, e.g., [vB05, §5], [MMM12], [BBKK18, Example 8.4], and [FP19].

Here is one last example.

Example 3.72 (Maybe)**.** We saw in Example 1.62 that the maybe monad on **Set** is presented by the theory of $\Sigma = \{\mathsf{p} : 0\}$ with no equations. Let us generalize this to the maybe monad on **GMet**.[448] We saw in Corollary 3.51 that $\mathbf{QAlg}(\Sigma, \hat{E}_{\mathbf{1}}) \cong \mathbf{1}/\mathbf{GMet}$, where $\hat{E}_{\mathbf{1}}$ contains the single equation $\vdash \mathsf{p} =_\varepsilon \mathsf{p}$ with ε being the self-distance of the unique element in $\mathbf{1}$, are the same thing as objects in the coslice. This isomorphism commutes with the forgetful functors to **GMet**,[449] and we get that the monad $\widehat{\mathcal{T}}_{\Sigma,\hat{E}_{\mathbf{1}}}$ obtained via the existence of free algebras is isomorphic to the monad $- + \mathbf{1}$ which is obtained via the existence of free objects in $\mathbf{1}/\mathbf{GMet}$.[450]

[448] It exists because **GMet** has a terminal object (Proposition 2.33) and coproducts (Corollary 3.51).

[449] The functor $U : \mathbf{1}/\mathbf{GMet} \to \mathbf{GMet}$ sends the pair $(\mathbf{X}, f : \mathbf{1} \to \mathbf{X})$ to $\mathbf{X}$.

[450] You need to check that $\mathbf{X} + \mathbf{1}$ is indeed the free object on $\mathbf{X}$ in this coslice.

3.4 Lifting Presentations

Most examples of **GMet** presentations in the literature, e.g., [MPP16, MV20, MSV21, MSV22], including Examples 3.69, 3.71, and 3.72, are built on top of a **Set** presentation. In summary, there is a monad M on **Set** with a known algebraic presentation (Σ, E) (e.g. $\mathcal{P}_{\mathrm{ne}}$ and semilattices or $\mathcal{D}$ and convex algebras) and a lifting of every space (X,d) to a space $(MX, \widehat{d})$. Then, a quantitative algebraic theory $(\Sigma, \hat{E})$ over the same signature is generated by counterparts to the equations in E as well as new quantitative equations to model the liftings. Finally, it is shown how the theory axiomatizes the lifting, namely, the **GMet** monad induced by the theory is isomorphic to a monad whose action on objects is the assignment $(X,d) \mapsto (MX, \widehat{d})$.

In this section, we prove Theorem 3.84 which makes this process more automatic and gives necessary and sufficient conditions for when it can actually be done. Throughout, we fix a monad (M, η, μ) on **Set** and an algebraic theory (Σ, E) presenting M via an isomorphism $\rho : \mathcal{T}_{\Sigma,E} \cong M$. We first give multiple definitions to make precise what we mean by *lifting*.

Definition 3.73 (Liftings)**.** We have three different notions of lifting that we introduce from weakest to strongest.

- A **mere lifting** of M to **GMet** is an assignment $(X, d_{\mathbf{X}}) \mapsto (MX, \widehat{d_{\mathbf{X}}})$ defining a generalized metric on MX for every generalized metric on X.[451]

- A **functor lifting** of M to **GMet** is a functor $\widehat{M} : \mathbf{GMet} \to \mathbf{GMet}$ that makes the square below commute.

$$\begin{array}{ccc} \mathbf{GMet} & \xrightarrow{\widehat{M}} & \mathbf{GMet} \\ {\scriptstyle U}\downarrow & & \downarrow{\scriptstyle U} \\ \mathbf{Set} & \xrightarrow[M]{} & \mathbf{Set} \end{array} \tag{3.49}$$

  Note in particular that for every space $\mathbf{X}$, the carrier of $\widehat{M}\mathbf{X}$ is MX, so we obtain a mere lifting $\mathbf{X} \mapsto \widehat{M}\mathbf{X}$. Furthermore, given a nonexpansive map $f : \mathbf{X} \to \mathbf{Y}$, the underlying function of $\widehat{M}f$ is Mf, i.e. $Mf : \widehat{M}\mathbf{X} \to \widehat{M}\mathbf{Y}$ is nonexpansive.

  In fact, if we have a mere lifting $(X, d_{\mathbf{X}}) \mapsto (MX, \widehat{d_{\mathbf{X}}})$ such that for every nonexpansive map $f : \mathbf{X} \to \mathbf{Y}$, $Mf : (MX, \widehat{d_{\mathbf{X}}}) \to (MY, \widehat{d_{\mathbf{Y}}})$ is nonexpansive, we automatically get a functor lifting $\widehat{M}$ whose action on objects is given by the mere lifting.[452] We conclude that functor liftings are just mere liftings with that additional condition.

- A **monad lifting** of M to **GMet** is a monad $(\widehat{M}, \widehat{\eta}, \widehat{\mu})$ on **GMet** such that $\widehat{M}$ is a functor lifting of M and furthermore $U\widehat{\eta} = \eta U$ and $U\widehat{\mu} = \mu U$. These two equations mean that the underlying functions of the unit and multiplication $\widehat{\eta}_{\mathbf{X}}$ and $\widehat{\mu}_{\mathbf{X}}$ are η_X and μ_X for any space $\mathbf{X}$.[453] In particular, the maps

$$\eta_X : \mathbf{X} \to \widehat{M}\mathbf{X} \quad \text{and} \quad \mu_X : \widehat{M}\widehat{M}\mathbf{X} \to \widehat{M}\mathbf{X}$$

  are nonexpansive for every $\mathbf{X}$. In fact, since U is faithful, that completely determines $\widehat{\eta}_{\mathbf{X}}$ and $\widehat{\mu}_{\mathbf{X}}$, and we conclude as before that a monad lifting is just a mere lifting with three additional conditions:

  1. $Mf : (MX, \widehat{d_{\mathbf{X}}}) \to (MY, \widehat{d_{\mathbf{Y}}})$ is nonexpansive if $f : \mathbf{X} \to \mathbf{Y}$ is nonexpansive,
  2. $\eta_X : (X, d_{\mathbf{X}}) \to (MX, \widehat{d_{\mathbf{X}}})$ is nonexpansive for every $\mathbf{X}$, and
  3. $\mu_X : (MMX, \widehat{\widehat{d_{\mathbf{X}}}}) \to (MX, \widehat{d_{\mathbf{X}}})$ is nonexpansive for every $\mathbf{X}$.

In practice, when defining a monad lifting, we will define a mere lifting and check Items 1–3. Let us give an example.

[451] The name *lifting* more commonly refers to what we call functor lifting or monad lifting which require more conditions than a mere lifting, hence the name *mere lifting*.

[452] The action on morphisms is prescribed by (3.49), namely, the underlying function of $\widehat{M}f$ is Mf which is nonexpansive by hypothesis, and since U is faithful, that determines $\widehat{M}f$.

[453] In summary, the description of a monad M and its monad lifting $\widehat{M}$ are exactly the same after forgetting about distances. In particular, the action of $\widehat{M}$ on morphisms does not depend on the distances at the source or the target, and similarly, the unit and multiplication maps do not depend on the distance of the space.

Example 3.74. Given an L-space (X,d), we define an L-relation $\widehat{d}$ on $\mathcal{P}_{\mathrm{ne}}X$ as follows: for any non-empty finite $S,S' \subseteq X$,

$$\widehat{d}(S,S') = \begin{cases} \bot & S = S' \\ d(x,y) & S = \{x\} \text{ and } S' = \{y\} \\ \top & \text{otherwise} \end{cases}. \tag{3.50}$$

Instantiating **GMet** with the category of L-spaces that satisfy reflexivity ($x \vdash x =_\bot x$), (3.50) defines a mere lifting of $\mathcal{P}_{\mathrm{ne}}$ to **GMet** given by $(X,d) \mapsto (\mathcal{P}_{\mathrm{ne}}X, \widehat{d})$.[454] Viewing $\mathcal{P}_{\mathrm{ne}}$ as modelling nondeterminism, this lifting could model a system where nondeterministic processes cannot be meaningfully compared (they are put at maximum distance) unless the sets of possible outcomes are the same (distance is minimal) or both processes are deterministic (distance is inherited from the distance between the only possible outcomes).

[454] We need reflexivity to ensure the first and second cases do not clash. You can also check that whenever d is a metric space, $\widehat{d}$ is as well, so we get a mere lifting of $\mathcal{P}_{\mathrm{ne}}$ to **Met**.

We show this is a monad lifting of $(\mathcal{P}_{\mathrm{ne}}, \eta, \mu)$,[455] with Lemmas 3.75–3.77.

[455] The unit and multiplication of $\mathcal{P}_{\mathrm{ne}}$ were defined in Example 1.52.

Lemma 3.75. *If $f : (X,d) \to (Y,\Delta)$ is nonexpansive, then so is the direct image function $\mathcal{P}_{\mathrm{ne}}f : (\mathcal{P}_{\mathrm{ne}}X, \widehat{d}) \to (\mathcal{P}_{\mathrm{ne}}Y, \widehat{\Delta})$.*[456]

[456] We write $f(S)$ instead of $\mathcal{P}_{\mathrm{ne}}f(S)$ for better readability.

Proof. Let $S,S' \in \mathcal{P}_{\mathrm{ne}}X$. If $S = S'$, then $f(S) = f(S')$, so

$$\widehat{\Delta}(f(S), f(S')) = \bot \leq \bot = \widehat{d}(S,S').$$

If $S = \{x\}$ and $S' = \{y\}$, then $f(S) = \{f(x)\}$ and $f(S') = \{f(y)\}$, so[457]

[457] The inequality holds because f is nonexpansive.

$$\widehat{\Delta}(f(S), f(S')) = \Delta(f(x), f(y)) \leq d(x,y) = \widehat{d}(S,S').$$

Otherwise, $\widehat{d}(S,S') = \top$ and $\widehat{\Delta}(f(S), f(S'))$ is always less or equal to $\top$. $\square$

Lemma 3.76. *For any (X,d), the map $\eta_X : (X,d) \to (\mathcal{P}_{\mathrm{ne}}X, \widehat{d})$ is nonexpansive.*

Proof. Recall that $\eta_X(x) = \{x\}$. For any $x,y \in X$, $\widehat{d}(\{x\},\{y\}) = d(x,y)$, so η_X is even an isometry. $\square$

Lemma 3.77. *For any (X,d), the map $\mu_X : (\mathcal{P}_{\mathrm{ne}}\mathcal{P}_{\mathrm{ne}}X, \widehat{\widehat{d}}) \to (\mathcal{P}_{\mathrm{ne}}X, \widehat{d})$ is nonexpansive.*

Proof. Recall that $\mu_X(F) = \cup F$ and let $F,F' \in \mathcal{P}_{\mathrm{ne}}\mathcal{P}_{\mathrm{ne}}X$. The case $F = F'$ is dealt with like in Lemma 3.75, it implies $\cup F = \cup F'$, hence the distances on both sides are $\bot$. If $F = \{S\}$ and $F' = \{S'\}$, $\cup F = S$ and $\cup F' = S'$, then

$$\widehat{d}(\mu_X(F), \mu_X(F')) = \widehat{d}(S,S') = \widehat{\widehat{d}}(\{S\},\{S'\}).$$

Otherwise, $\widehat{\widehat{d}}(F,F') = \top$, so the inequality holds because $\widehat{d}(\mu_X(F), \mu_X(F'))$ is always less or equal to $\top$. $\square$

Here is an example of a functor lifting that is not a monad lifting.

Example 3.78. The **total variation** distance is a metric defined on probability distributions. For any X, we define $\mathrm{tv} : \mathcal{D}X \times \mathcal{D}X \to [0,1]$ by, for any $\varphi, \psi \in \mathcal{D}X$,[458]

[458] Since φ and ψ have finite support, we can restrict the quantification of the supremum to finite subsets of X, or even to subsets of the union of the supports of φ and ψ. Also, both $\varphi(S)$ and $\psi(S)$ are at most in $[0,1]$, so $\mathrm{tv}(\varphi,\psi)$ also takes values in $[0,1]$.

$$\mathsf{tv}(\varphi,\psi) = \sup_{S\subseteq X} |\varphi(S) - \psi(S)|.$$

Even though the assignment $(X,d) \mapsto (\mathcal{D}X, \mathsf{tv})$ is a mere lifting of the monad $\mathcal{D}$ to **Met**, namely, $(\mathcal{D}X, \mathsf{tv})$ is a metric whenever (X,d) is, it is not a monad lifting. One can show that Mf is nonexpansive whenever f is, so it is a functor lifting, and even that the multiplication is always nonexpansive, but the unit is not because if $x \neq y \in X$ are points at distance $d(x,y) < 1$, then $\mathsf{tv}(\delta_x, \delta_y) = 1 > d(x,y)$.

Many monads of interest on different **GMet** categories are monad liftings of **Set** monads which have an algebraic presentation. We already mentioned the Hausdorff and Kantorovich monad liftings in Examples 3.69 and 3.71, but there is also a combination of the two: the Hausdorff–Kantorovich monad lifting of the convex sets of distributions monad [MV20] to **Met**. In [MSV21], we further combined these with the maybe monad on **Met**. Another example is the formal ball monad on quasi-metric spaces [GL19] which is a monad lifting of a writer monad on **Set**. All of these happen to have a quantitative algebraic presentation,[459] and we will show that this is not a coincidence.

[459] Goubault-Larrecq does not talk about quantitative algebras in [GL19], but the quantitative writer monad of [BMPP21, §4.3.2] has a presentation which can easily be adapted to present the monad of [GL19].

Given a monad lifting $\widehat{M}$, we know that it acts on sets just like M does, and that can be described algebraically through the presentation $\rho : \mathcal{T}_{\Sigma,E} \cong M$. This can help to understand how $\widehat{M}$ acts on distances. For any space $\mathbf{X}$, we see the distance $\widehat{d_{\mathbf{X}}}$ on MX as a distance $\widehat{d}$ on terms modulo E via the bijection ρ_X:[460]

[460] Recall Proposition 2.47.

$$\widehat{d}([s]_E, [t]_E) = \widehat{d_{\mathbf{X}}}(\rho_X[s]_E, \rho_X[t]_E).$$

Can we find some quantitative equations $\hat{E}$ that axiomatize $\widehat{d}$, i.e. such that $d_{\hat{E}}$ and $\widehat{d}$ are isomorphic (uniformly for all $\mathbf{X}$)?

First of all, for the distances to be isomorphic, they need to be on the same set, namely, we need to have $\mathcal{T}_\Sigma X/{\equiv_E} \cong \mathcal{T}_\Sigma X/{\equiv_{\hat{E}}}$, or equivalently, $s \equiv_E t \iff s \equiv_{\hat{E}} t$. At once, this removes some options for which equations to add in $\hat{E}$. For instance, we cannot add $\mathbf{X} \vdash s = t$ if $X \vdash s = t$ does not already belong to $\mathfrak{Th}(E)$. Conversely, if $X \vdash s = t \in \mathfrak{Th}(E)$, we need to ensure $\mathbf{X} \vdash s = t$ belongs to $\mathfrak{QTh}(\hat{E})$. We can do this by adding $\mathbf{X}_\top \vdash s = t$ to $\hat{E}$ thanks to Example 3.57.

After that, we will have to add quantitative equations with quantities to axiomatize $\widehat{d}$, but we have to be careful not to break the equivalence we just obtained between $\equiv_E$ and $\equiv_{\hat{E}}$. For instance, if **GMet** = **Met**, $\mathsf{f} : 1 \in \Sigma$ and $E = \emptyset$, then we cannot have $x =_{\frac{1}{2}} y \vdash \mathsf{f}x =_0 \mathsf{f}y \in \hat{E}$, because using the equation $x =_0 y \vdash x = y$ that defines **Met**, we could conclude that $x =_{\frac{1}{2}} y \vdash \mathsf{f}x = \mathsf{f}y$ belongs to $\mathfrak{QTh}(\hat{E})$, which means $\mathsf{f}x \equiv_{\hat{E}} \mathsf{f}y$ whenever $d_{\mathbf{X}}(x,y) \leq \frac{1}{2}$ while $\mathsf{f}x \not\equiv_E \mathsf{f}y$.

The relation between $\hat{E}$ and E seems to mimic our intuition about mere liftings. We say that $\hat{E}$ extends E.

Definition 3.79 (Extension)**.** Given a class E of equations over Σ and a class $\hat{E}$ of quantitative equations over Σ, we say that $\hat{E}$ is an **extension** of E if for all $\mathbf{X} \in$ **GMet** and $s, t \in \mathcal{T}_\Sigma X$,

$$X \vdash s = t \in \mathfrak{Th}(E) \iff \mathbf{X} \vdash s = t \in \mathfrak{QTh}(\hat{E}). \tag{3.51}$$

Remark 3.80. Let us make two delicate points on the quantification of $\mathbf{X}$ in (3.51).

First, it happens *before* the equivalence. This means that equalities[461] that hold in $\mathcal{T}_{\Sigma,E}X$ coincide with the equalities that hold in $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$ for each $\mathbf{X}$ individually. In particular, if $\mathbf{X}$ and $\mathbf{X}'$ are spaces on the same set X, then the equalities that hold in $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$ and $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}'$ coincide. This intuitively corresponds to the fact that the action of $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}$ does not depend on distances.

[461] This is not a formal term: by *equalities that hold*, we mean which Σ-terms are in the same equivalence class.

If instead of (3.51) we had the following equivalence with the quantification inside,

$$X \vdash s = t \in \mathfrak{Th}(E) \iff \forall \mathbf{X} \in \mathbf{GMet}, \mathbf{X} \vdash s = t \in \mathfrak{QTh}(\hat{E}),$$

then the equalities in $\mathcal{T}_{\Sigma,E}X$ would be those that hold in all $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$ (for all spaces $\mathbf{X}$ with carrier X). In particular, $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$ and $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}'$ could have different equivalence classes. That is not desirable when defining a mere lifting.

Second, even though the context of a quantitative equation can be any L-space, $\mathbf{X}$ is only quantified over generalized metric spaces here. This implies that the equivalence classes of $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$ and $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}'$ may be different if $d_{\mathbf{X}}$ and $d'_{\mathbf{X}}$ are two different L-relations on X. This does not contradict our intuition about liftings because we only care about the action of $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}$ on L-spaces that belong to $\mathbf{GMet}$.

For instance, let $\Sigma = \{\mathsf{f}:1\}$, $E = \emptyset$, $\hat{E} = \emptyset$, and $\mathbf{GMet}$ be defined by the equation $x =_\perp y \vdash x = x$. If $X = \{x, y\}$ and $d_{\mathbf{X}}(x,y) = \perp$, then $\mathbf{X} \vdash \mathsf{f}x = \mathsf{f}y$ belongs to $\mathfrak{QTh}(\hat{E})$ while $\mathsf{f}x \not\equiv_E \mathsf{f}y$.[462] Still, it makes sense that $\hat{E}$ extend E since both have no equations.

[462] Here is the derivation (the application of $\mathbf{GMet}$ implicitly uses the fact that $x =_\perp y \vdash x = x$ is syntactic sugar for $\mathbf{X} \vdash x =_\perp y$):

$$\cfrac{\cfrac{}{\mathbf{X} \vdash x = y}\,\textsc{GMet}}{\mathbf{X} \vdash \mathsf{f}x = \mathsf{f}y}\,\textsc{Cong}$$

It turns out that extensions are stronger than mere liftings because we can show the monad we constructed via terms modulo $\hat{E}$ is a monad lifting of $\mathcal{T}_{\Sigma,E}$.

Proposition 3.81. *If $\hat{E}$ is an extension of E, then $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}$ is a monad lifting of $\mathcal{T}_{\Sigma,E}$.*

Proof. We need to check the following three equations where $U : \mathbf{GMet} \to \mathbf{Set}$ is the forgetful functor:

$$U\widehat{\mathcal{T}}_{\Sigma,\hat{E}} = \mathcal{T}_{\Sigma,E}U \qquad U\widehat{\eta}^{\Sigma,\hat{E}} = \eta^{\Sigma,E}U \qquad U\widehat{\mu}^{\Sigma,\hat{E}} = \mu^{\Sigma,E}U.$$

First, we have to show that for any space $\mathbf{X}$, $U\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} = \mathcal{T}_{\Sigma,E}U\mathbf{X}$. By definitions, the `L.H.S.` is $\mathcal{T}_\Sigma X/\equiv_{\hat{E}}$ and the `R.H.S.` is $\mathcal{T}_\Sigma X/\equiv_E$, so it boils down to showing that for all $s, t \in \mathcal{T}_\Sigma X$, $s \equiv_{\hat{E}} t \iff s \equiv_E t$. This readily follows from the definitions of $\equiv_{\hat{E}}$ and $\equiv_E$, and from (3.51):[463]

[463] Note again the importance of being able to do this for each $\mathbf{X}$ individually.

$$s \equiv_{\hat{E}} t \overset{(3.12)}{\iff} \mathbf{X} \vdash s = t \in \mathfrak{QTh}(\hat{E}) \overset{(3.51)}{\iff} X \vdash s = t \in \mathfrak{Th}(E) \overset{(1.21)}{\iff} s \equiv_E t.$$

Next, we have to show that $U\widehat{\mathcal{T}}_{\Sigma,\hat{E}}f = \mathcal{T}_{\Sigma,E}f$ for any $f : \mathbf{X} \to \mathbf{Y}$. This is done rather quickly by comparing their definitions, they make the same squares (1.23) and (3.18) commute now that we know $\equiv_{\hat{E}}$ and $\equiv_E$ coincide.

This takes care of the first equation, and the other two are done very similarly, we compare the definitions of $\widehat{\eta}^{\Sigma,\hat{E}}$ and $\eta^{\Sigma,E}$ (resp. $\widehat{\mu}^{\Sigma,\hat{E}}$ and $\mu^{\Sigma,E}$) and conclude they are the same when $\equiv_{\hat{E}}$ and $\equiv_E$ coincide.[464] □

[464] We defined $\widehat{\eta}^{\Sigma,\hat{E}}$ in (3.33), $\eta^{\Sigma,E}$ in Footnote 117, $\widehat{\mu}^{\Sigma,\hat{E}}$ in (3.21), and $\mu^{\Sigma,E}$ in (1.32).

So if we are able to construct an extension $\hat{E}$ of E, we can obtain a monad lifting of M by passing through the isomorphism $\rho : \mathcal{T}_{\Sigma,E} \cong M$.

Corollary 3.82. *If M is presented by (Σ, E), and $\hat{E}$ is an extension of E, then $\hat{E}$ presents a monad lifting of M.*

Proof. We first construct a monad lifting of (M, η, μ). For any space $\mathbf{X}$, we have an isomorphism $\rho_X^{-1} : MX \to \mathcal{T}_{\Sigma,E}X$, and a generalized metric $d_{\hat{E}}$ on $\mathcal{T}_{\Sigma,E}$ (since the underlying set of $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}$ is $\mathcal{T}_{\Sigma,E}$ by Proposition 3.81). We can define a generalized metric $\widehat{d_{\mathbf{X}}}$ on MX as we have done for Proposition 2.47 to guarantee that $\rho_X^{-1} : (MX, \widehat{d_{\mathbf{X}}}) \to \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X}$ is an isomorphism:[465]

$$\widehat{d_{\mathbf{X}}}(m, m') = d_{\hat{E}}(\rho_X^{-1}(m), \rho_X^{-1}(m')). \tag{3.52}$$

[465] In words, the distance between m and m' in MX is computed by viewing them as (equivalence classes of) terms in $\mathcal{T}_\Sigma X$, then using the distance between them given by $d_{\hat{E}}$.

This yields a mere lifting $(X, d_{\mathbf{X}}) \mapsto (MX, \widehat{d_{\mathbf{X}}})$.

In order to show this is a monad lifting, we use the following diagrams (quantified for all $\mathbf{X} \in \mathbf{GMet}$ and nonexpansive $f : \mathbf{X} \to \mathbf{Y}$) which commute because ρ is a monad isomorphism with inverse ρ^{-1}.[466]

$$\begin{tikzcd}
(MX, \widehat{d_{\mathbf{X}}}) \arrow[r, "\rho_X^{-1}"] \arrow[d, "Mf"'] & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[d, "\mathcal{T}_{\Sigma,E}f"] \\
(MY, \widehat{d_{\mathbf{Y}}}) & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{Y} \arrow[l, "\rho_Y"]
\end{tikzcd}
\qquad
\begin{tikzcd}
\mathbf{X} \arrow[r, "\eta_X^{\Sigma,E}"] \arrow[dr, "\eta_X"'] & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[d, "\rho_X"] \\
& (MX, \widehat{d_{\mathbf{X}}})
\end{tikzcd}$$

$$\begin{tikzcd}
(MMX, \widehat{\widehat{d_{\mathbf{X}}}}) \arrow[r, "\rho_{MX}^{-1}"] \arrow[d, "\mu_X"'] & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}(X, \widehat{d_{\mathbf{X}}}) \arrow[r, "\mathcal{T}_{\Sigma,E}\rho_X^{-1}"] & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[d, "\mu_X^{\Sigma,E}"] \\
(MX, \widehat{d_{\mathbf{X}}}) & & \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \arrow[ll, "\rho_X"]
\end{tikzcd}$$

[466] The first holds by naturality, the second by (1.51), and the third by (1.52). Moreover, all the functions in these diagrams are nonexpansive (with the sources and targets as drawn) by previous results:

- We just showed the components of ρ are isometries.
- We showed $\mathcal{T}_{\Sigma,E}f$ is the underlying function of $\widehat{\mathcal{T}}_{\Sigma,E}f$ because $\widehat{\mathcal{T}}_{\Sigma,E}$ is a monad lifting of $\mathcal{T}_{\Sigma,E}$ (Proposition 3.81), so $\mathcal{T}_\Sigma Ef$ is nonexpansive when f is nonexpansive.
- By the previous two points, $\mathcal{T}_{\Sigma,E}\rho_X^{-1}$ is nonexpansive.
- Again since $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}$ is a monad lifting of $\mathcal{T}_{\Sigma,E}$, $\eta_X^{\Sigma,E}$ and $\mu_X^{\Sigma,E}$ are nonexpansive.

These show (detailed in the footnote) that Mf, η_X and μ_X are compositions of nonexpansive maps, and hence are nonexpansive. We obtain a monad lifting $\widehat{M}$ of M to **GMet** which sends $(X, d_{\mathbf{X}})$ to $(MX, \widehat{d_{\mathbf{X}}})$.

It remains to show that $\widehat{M}$ is presented by $(\Sigma, \hat{E})$. By construction, we have the isomorphism $\widehat{\rho}_{\mathbf{X}} : \widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \to \widehat{M}\mathbf{X}$ whose underlying function is ρ_X for every $\mathbf{X}$. The fact that $\widehat{\rho}$ is a monad morphism follows from the facts that ρ is a monad morphism, and that $U : \mathbf{GMet} \to \mathbf{Set}$ is faithful so it reflects commutativity of diagrams.[467] □

[467] Let us detail the argument for naturality, the others would follow the same pattern. We need to show that $\widehat{\rho}_{\mathbf{Y}} \circ \widehat{M}f = \widehat{M}f \circ \widehat{\rho}_{\mathbf{X}}$. Applying U, we get $\rho_Y \circ Mf = Mf \circ \rho_X$ which is true because ρ is natural, hence $U(\widehat{\rho}_{\mathbf{Y}} \circ \widehat{M}f) = U(\widehat{M}f \circ \widehat{\rho}_{\mathbf{X}})$. Since U is faithful, and the desired equation holds.

Now, we would like to have a converse to Corollary 3.82. Namely, if $(X, d_{\mathbf{X}}) \mapsto (MX, \widehat{d_{\mathbf{X}}})$ is given by a monad lifting $\widehat{M}$ of M to **GMet**, our goal is to construct an extension $\hat{E}$ of E such that the monad lifting corresponding to $\hat{E}$ (given in Corollary 3.82) is $\widehat{M}$. There is no obvious reason this is even possible, maybe $\widehat{M}$ is a monad lifting that has no quantitative algebraic presentation.[468] Our next theorem shows that such an $\hat{E}$ always exists. In fact, it is constructed very naively.

[468] Or maybe $\widehat{M}$ has a presentation that is not an extension of E, but our informal discussion leading to the definition of extensions indicates that is less probable.

As discussed in Example 3.57, when $\hat{E}$ contains all the quantitative equations in

$$\hat{E}_1 = \{\mathbf{X}_\top \vdash s = t \mid X \vdash s = t \in E\}, \tag{3.53}$$

then we have at least one direction of (3.51), namely, that $X \vdash s = t \in \mathfrak{Th}(E)$ implies $\mathbf{X} \vdash s = t \in \mathfrak{QTh}(\hat{E})$ for all $\mathbf{X}$ and $s, t \in \mathcal{T}_\Sigma X$.[469] Next, we include in $\hat{E}$ all the possible

[469] We use Lemma 3.58.

equations $\mathbf{X} \vdash s =_\varepsilon t$ where ε is the distance between s and t when viewed inside $\widehat{M}\mathbf{X}$ (via ρ_X),[470] namely, $\hat{E}_2 \subseteq \hat{E}$ where

$$\hat{E}_2 = \left\{ \mathbf{X} \vdash s =_\varepsilon t \mid \mathbf{X} \in \mathbf{GMet}, s, t \in \mathcal{T}_\Sigma X, \varepsilon = \widehat{d_\mathbf{X}}(\rho_X[s]_E, \rho_X[t]_E) \right\}. \tag{3.54}$$

This is a very large bunch of equations (it is not even a set), but it leaves no stone unturned, meaning that the distance computed by $\hat{E}$ will always be smaller than the distance in $\widehat{M}\mathbf{X}$. Indeed, for any $m, m' \in MX$, letting $s, t \in \mathcal{T}_\Sigma X$ be such that $\rho_X[s]_E = m$ and $\rho_X[t]_E = m'$ (by surjectivity of ρ_X), we have[471]

$$\begin{aligned} \widehat{d_\mathbf{X}}(m, m') \leq \varepsilon &\implies \mathbf{X} \vdash s =_\varepsilon t \in \mathfrak{QTh}(\hat{E}) \\ &\iff d_{\hat{E}}([s]_E, [t]_E) \leq \varepsilon \\ &\iff d_{\hat{E}}(\rho_X^{-1}(m), \rho_X^{-1}(m')) \leq \varepsilon. \end{aligned}$$

In order to conclude that $\hat{E} = \hat{E}_1 \cup \hat{E}_2$ presents $\widehat{M}$, we need to show that $\hat{E}$ is an extension of E, i.e. the other direction of (3.51), and that the monad lifting defined in Corollary 3.82 coincides with $\widehat{M}$, i.e. the converse implication of the previous derivation holds. We will prove these by constructing a (family of) special algebras in $\mathbf{QAlg}(\Sigma, \hat{E})$.[472]

For any generalized metric space $\mathbf{A}$, we denote by $\mathbb{M}\mathbf{A}$ the quantitative Σ-algebra $(MA, [\![-]\!]_{\mu_A}, \widehat{d_\mathbf{A}})$, where

- $(MA, \widehat{d_\mathbf{A}})$ is the space obtained by applying $\widehat{M}$ to $\mathbf{A}$, and
- $(MA, [\![-]\!]_{\mu_A})$ is the Σ-algebra obtained by applying the isomorphism $\mathbf{Alg}(\Sigma, E) \cong \mathbf{EM}(M)$ (from the presentation) to the M-algebra (MA, μ_A) (from Example 1.57).

We can show that $\mathbb{M}\mathbf{A}$ belongs to $\mathbf{QAlg}(\Sigma, \hat{E}_1 \cup \hat{E}_2)$.

Lemma 3.83. *For all $\phi \in \hat{E}_1 \cup \hat{E}_2$, $\mathbb{M}\mathbf{A} \vDash \phi$.*

Proof. If $\phi = \mathbf{X}_\top \vdash s = t \in \hat{E}_1$, then by construction $(MA, [\![-]\!]_{\mu_A})$ satisfies $X \vdash s = t \in E$. So $\mathbb{M}\mathbf{A}$ satisfies ϕ by Lemma 3.32.

Suppose now that $\phi = \mathbf{X} \vdash s =_\varepsilon t \in \hat{E}_2$ with $\varepsilon = \widehat{d_\mathbf{X}}(\rho_X[s]_E, \rho_X[t]_E)$. A bit of unrolling[473] shows that for an assignment $\iota : X \to MA$, the interpretation $[\![-]\!]^\iota_{\mu_A}$ is the composite

$$\mathcal{T}_\Sigma X \xrightarrow{\mathcal{T}_\Sigma \iota} \mathcal{T}_\Sigma MA \xrightarrow{[-]_E} \mathcal{T}_{\Sigma,E} MA \xrightarrow{\rho_{MA}} MMA \xrightarrow{\mu_A} MA.$$

For later use, we apply the naturality of $[-]_E$ (1.23) and ρ to rewrite the composite as

$$[\![-]\!]^\iota_{\mu_A} = \mathcal{T}_\Sigma X \xrightarrow{[-]_E} \mathcal{T}_{\Sigma,E} X \xrightarrow{\rho_X} MX \xrightarrow{M\iota} MMA \xrightarrow{\mu_A} MA. \tag{3.55}$$

We conclude that $\mathbb{M}\mathbf{A} \vDash \phi$ with the following derivation which holds for all nonexpansive $\hat{\iota} : \mathbf{X} \to \widehat{M}\mathbf{A}$.[474]

$$\begin{aligned} \widehat{d_\mathbf{A}}([\![s]\!]^{\hat{\iota}}_{\mu_A}, [\![t]\!]^{\hat{\iota}}_{\mu_A}) &= \widehat{d_\mathbf{A}}\left(\mu_A(M\hat{\iota}(\rho_X[s]_E)), \mu_A(M\hat{\iota}(\rho_X[t]_E))\right) && \text{by (3.55)} \\ &\leq \widehat{\widehat{d_\mathbf{A}}}\left(M\hat{\iota}(\rho_X[s]_E), M\hat{\iota}(\rho_X[t]_E)\right) && \mu_A \text{ is nonexpansive} \\ &\leq \widehat{d_\mathbf{X}}\left(\rho_X[s]_E, \rho_X[t]_E\right) && M\hat{\iota} \text{ is nonexpansive} \\ &= \varepsilon && \square \end{aligned}$$

[470] We are essentially doing the opposite of (3.52).

[471] The implication follows because by definition, $\hat{E}$ will contain $\mathbf{X} \vdash s =_{d_\mathbf{X}(m,m')} t$, hence by the MAX rule, we will have $\mathbf{X} \vdash s =_\varepsilon t \in \mathfrak{QTh}(\hat{E})$. The first equivalence is (3.16), and the second holds because ρ_X^{-1} is the inverse of ρ_X.

[472] In turns out (after the rest of the proof) we are constructing the free algebra over $\mathbf{A}$, but we feel it is not necessary to make that explicit.

[473] Look at the definition of P^{-1} in Proposition 1.58, in particular what we proved in Footnote 168, and the definition of $-\rho$ in (1.56).

[474] Our hypothesis that $\widehat{M}$ is a monad lifting yields nonexpansiveness of μ_A and $M\hat{\iota}$.

Theorem 3.84. *Let $\widehat{M}$ be a monad lifting of M to* **GMet**, *and $\hat{E} = \hat{E}_1 \cup \hat{E}_2$. Then, $\hat{E}$ is an extension of E and it presents $\widehat{M}$.*

Proof. We already showed the forward implication of (3.51) when we defined $\hat{E}_1$ (3.53). For the converse, suppose that $\mathbf{X} \vdash s = t \in \mathfrak{QTh}(\hat{E})$, we saw in Lemma 3.83 that $\mathbb{M}\mathbf{X}$ satisfies $\mathbf{X} \vdash s = t$. Taking the assignment $\eta_X : \mathbf{X} \to \widehat{M}\mathbf{X}$ which is nonexpansive because $\widehat{M}$ is a monad lifting, we have $[\![s]\!]^{\eta_X}_{\mu_X} = [\![t]\!]^{\eta_X}_{\mu_X}$. Using (3.55) and the monad law $\mu_X \circ M\eta_X = \mathrm{id}_{MX}$ (left triangle in (1.40)), we find

$$\rho_X[s]_E = \mu_X(M\eta_X(\rho_X[s]_E)) = [\![s]\!]^{\eta_X}_{\mu_X} = [\![t]\!]^{\eta_X}_{\mu_X} = \mu_X(M\eta_X(\rho_X[t]_E)) = \rho_X[t]_E.$$

Finally, since ρ_X is a bijection, we have $[s]_E = [t]_E$, i.e. $X \vdash s = t \in \mathfrak{Th}(E)$.

We already showed that $\widehat{d_{\mathbf{X}}}(m, m') \geq d_{\hat{E}}(\rho_X^{-1}(m), \rho_X^{-1}(m'))$ when defining $\hat{E}_2$. For the converse, let $m = \rho_X[s]_E$ and $m' = \rho_X[t]_E$ for some $s, t \in \mathcal{T}_\Sigma X$ and suppose that $d_{\hat{E}}([s]_E, [t]_E) \leq \varepsilon$, or equivalently by (3.16), that $\mathbf{X} \vdash s =_\varepsilon t \in \mathfrak{QTh}(\hat{E})$. As above, Lemma 3.83 says that $\mathbb{M}\mathbf{X}$ satisfies that equation. Taking the assignment $\eta_X : \mathbf{X} \to \widehat{M}\mathbf{X}$ which is nonexpansive because $\widehat{M}$ is a monad lifting, we have[475]

$$\widehat{d_{\mathbf{X}}}(m, m') = \widehat{d_{\mathbf{X}}}(\rho_X[s]_E, \rho_X[t]_E) = \widehat{d_{\mathbf{X}}}\left([\![s]\!]^{\eta_X}_{\mu_X}, [\![t]\!]^{\eta_X}_{\mu_X}\right) \leq \varepsilon.$$

[475] The second inequality holds again by (3.55) and (1.40).

Comparing with (3.52), we conclude that $\widehat{M}$ is exactly the monad lifting from Corollary 3.82. In particular, $\hat{E}$ presents $\widehat{M}$ via $\widehat{\rho}$ whose component at $\mathbf{X}$ is ρ_X. □

Remark 3.85. A deeper result hides behind the last line. It follows from our constructions that if you start from an extension $\hat{E}$, build a monad lifting $\widehat{M}$ from $\hat{E}$ with Corollary 3.82, then build an extension $\hat{E}'$ from $\widehat{M}$ with Theorem 3.84, you obtain two *equivalent* classes of equations, i.e. $\mathfrak{QTh}(\hat{E}) = \mathfrak{QTh}(\hat{E}')$. Similarly, if you start with a monad lifting $\widehat{M}$, then build an extension $\hat{E}$, then build a monad lifting $\widehat{M}'$, then $\widehat{M} = \widehat{M}'$.[476]

This does not yield a bijection but almost. If you restrict extensions of E to those that are quantitative algebraic theories,[477] then you get a bijection with monad liftings of M.

I believe it is a simple exercise in categorical logic to transform this remark into an (dual) equivalence of categories. A slightly more challenging task would be to allow M and E to vary to get a (fibered) equivalence.[478]

[476] We have equality on the nose because monad liftings are defined with equality on the nose. One can probably relax the definition of monad lifting to be up to isomorphisms without breaking this correspondence.

[477] i.e. they are *saturated*, you cannot add more quantitative equations without changing the algebras

[478] A restricted version of this was done in the case of **GMet** = **Poset** in [ADV22, Theorem 49].

When constructing the extension $\hat{E} = \hat{E}_1 \cup \hat{E}_2$, $\hat{E}_1$ can be fairly small since it has the size of E, but $\hat{E}_2$ as defined is always huge (not even a set). In theory, some results in the literature could allow us to restrict the size of contexts to be of a moderate size only with mild size conditions on L and $\hat{E}_{\mathbf{GMet}}$.[479] In practice, we can sometimes find some simple set of quantitative equations which will be equivalent to $\hat{E}_2$ (when $\hat{E}_1$ is present), and we give a couple of examples below. They require some *clever* arguments that depend on the application, but there may be room for optimization in the definition of $\hat{E}_2$.

[479] I will not write the proofs because I am not confident enough with the literature on accessible and presentable categories, but I believe [FMS21, Propositions 3.8 and 3.9] make it possible to adapt the arguments of Remark 1.49 replacing $\aleph_0$ with a different cardinal (we implicitly used $\aleph_0$ because $\lambda < \aleph_0 \Leftrightarrow \lambda$ finite).

Example 3.86 (Trivial Lifting of $\mathcal{P}_{\mathrm{ne}}$)**.** Recall the monad lifting of $\mathcal{P}_{\mathrm{ne}}$ to **GMet** = **QAlg**$(\emptyset, \{x \vdash x =_\perp x\})$ from Example 3.74. Let us denote it by $\widehat{\mathcal{P}}$, and its action on objects by $(X,d) \mapsto (\mathcal{P}_{\text{ne}}X, \widehat{d_{\mathbf{X}}})$.[480] We also denote with ρ the monad isomorphism witnessing that $\mathcal{P}_{\text{ne}}$ is presented by the theory of semilattices $(\Sigma_{\mathbf{S}}, E_{\mathbf{S}})$ (recall Example 1.66). By Theorem 3.84, there is a quantitative algebraic presentation for $\widehat{\mathcal{P}}$ given by[481]

$$\hat{E}_1 = \{\mathbf{X}_\top \vdash s = t \mid X \vdash s = t \in E_{\mathbf{S}}\} \text{ and } \hat{E}_2 = \left\{\mathbf{X} \vdash s =_\varepsilon t \mid \varepsilon = \widehat{d_{\mathbf{X}}}\left(\rho_X[s]_{E_{\mathbf{S}}}, \rho_X[t]_{E_{\mathbf{S}}}\right)\right\}.$$

We claim that the equations in $\hat{E}_1$ are enough, namely, $\mathfrak{QTh}(\hat{E}_1 \cup \hat{E}_2) = \mathfrak{QTh}(\hat{E}_1)$.

First, since $\hat{E}_1 \subseteq \hat{E}_1 \cup \hat{E}_2$, we infer that $\mathfrak{QTh}(\hat{E}_1) \subseteq \mathfrak{QTh}(\hat{E}_1 \cup \hat{E}_2)$.[482]

Second, recall from Lemma 3.58 that with the equations in $\hat{E}_1$, we can already prove all the equations in the theory of semilattices. This means that for any $\mathbf{X} \vdash s =_\varepsilon t \in \hat{E}_2$ with $\varepsilon = \widehat{d_{\mathbf{X}}}\left(\rho_X[s]_{E_{\mathbf{S}}}, \rho_X[t]_{E_{\mathbf{S}}}\right)$, we have the three following cases.

- If $[s]_{E_{\mathbf{S}}} = [t]_{E_{\mathbf{S}}}$ and $\varepsilon = \bot$, i.e. s and t represent the same subset of X, then the equation $X \vdash s = t$ is in $\mathfrak{Th}(E_{\mathbf{S}})$ which implies $\mathbf{X} \vdash s = t$ is in $\mathfrak{QTh}(\hat{E}_1)$. It follows by the following derivation that $\mathbf{X} \vdash s =_0 t \in \mathfrak{QTh}(\hat{E}_1)$ as desired.[483]

$$\dfrac{\mathbf{X} \vdash s = t \qquad \dfrac{\sigma = x \mapsto s \qquad \dfrac{}{x \vdash x =_\bot x}\,\textsc{GMet} \qquad \dfrac{}{\mathbf{X} \vdash s =_\top s}\,\textsc{Top}}{\mathbf{X} \vdash s =_\bot s}\,\textsc{SubQ}}{\mathbf{X} \vdash s =_\bot t}\,\textsc{CompR}$$

- If $[s]_{E_{\mathbf{S}}} = [x]_{E_{\mathbf{S}}}$ and $[t]_{E_{\mathbf{S}}} = [y]_{E_{\mathbf{S}}}$ for some $x, y \in X$ and $\varepsilon = d_{\mathbf{X}}(x,y)$, then the equations $X \vdash s = x$ and $X \vdash y = t$ are in $\mathfrak{Th}(E_{\mathbf{S}})$ which implies $\mathbf{X} \vdash s = x$ and $\mathbf{X} \vdash y = t$ are in $\mathfrak{QTh}(\hat{E}_1)$. Furthermore, Lemma 3.27 implies $\mathbf{X} \vdash x =_\varepsilon y \in \mathfrak{QTh}(\hat{E}_1)$, and finally by Lemmas 3.24 and 3.25, $\mathbf{X} \vdash s =_\varepsilon t$ also belongs to $\mathfrak{QTh}(\hat{E}_1)$ as desired.

- Otherwise, $\varepsilon = \top$, so $\mathbf{X} \vdash s =_\varepsilon t$ belongs to $\mathfrak{QTh}(\hat{E}_1)$ by Lemma 3.26.

We have shown that $\hat{E}_2 \subseteq \mathfrak{QTh}(\hat{E}_1)$, and it follows that $\mathfrak{QTh}(\hat{E}_1 \cup \hat{E}_2) \subseteq \mathfrak{QTh}(\hat{E}_1)$.[484]

In conclusion, we found that $\widehat{\mathcal{P}}$ is presented by the equations in $\hat{E}_1$ which we rewrite below:

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

Remark 3.87. Compared to the presentation of $\mathcal{P}_{\text{ne}}^{\uparrow}$, we simply removed (3.6). These quantitative equations are included in the theory by default in the framework of [MPP16] because they only consider quantitative algebras with interpretations of operations that are nonexpansive with respect to the product metric (see Example 3.10). It is then natural to ask whether the monad lifting $\widehat{\mathcal{P}}$ we defined can be presented by a quantitative algebraic theory in the sense of [MPP16]. The answer is negative because of a property that all monads presented by such theories share: they are enriched over $(\mathbf{Met}, \otimes, \mathbf{1})$[485]

The monad $\widehat{\mathcal{P}}$ is not enriched because it does not satisfy (see [ADV23, Example 7.(1)])

$$\forall f, g : (X,d) \to (Y,\Delta),\ \sup_{x \in X} \Delta(f(x), g(x)) \geq \sup_{S \in \mathcal{P}X} \widehat{\Delta}(f(S), g(S)).$$

480 The distance $\widehat{d_{\mathbf{X}}}$ was defined in (3.50).

481 We are concise in the quantifications for $\hat{E}_2$.

482 There are two ways to understand this. Semantically, the equations that are satisfied by all algebras in $\mathbf{QAlg}(\Sigma, \hat{E}_1)$ are also satisfied by all algebras in $\mathbf{QAlg}(\Sigma, \hat{E}_1 \cup \hat{E}_2)$ because the second category is contained in the first. Syntactically, if you have more axioms, you can prove more things.

483 Recall that the context of $x \vdash x =_\bot x$, after unrolling the syntactic sugar, is the L-space with x at distance $\top$ from itself, so we only need to prove $\sigma(x)$ is also at distance $\top$ from itself (we do it with Top).

484 Again, there are two different ways to understand this. Semantically, if all algebras in $\mathbf{QAlg}(\Sigma, \hat{E}_1)$ satisfy $\hat{E}_2$, then $\mathbf{QAlg}(\Sigma, \hat{E}_1)$ and $\mathbf{QAlg}(\Sigma, \hat{E}_1 \cup \hat{E}_2)$ are the same categories. Syntactically, in any derivation with axioms $\hat{E}_1 \cup \hat{E}_2$, you can replace each axiom in $\hat{E}_2$ by a derivation using only axioms in $\hat{E}_1$.

485 See [ADV23, Full version, after Corollary 4.19].

Let f be the identity function on $[0, \frac{1}{2}]$ and g be the squaring function, then the left hand side is at most $\frac{1}{2}$ (Δ is bounded by $\frac{1}{2}$), and the right hand side is 1 as witnessed by $S = \{0, \frac{1}{2}\}$: $f(S) = S$ and $g(S) = \{0, \frac{1}{4}\}$, so $\widehat{\Delta}(f(S), g(S)) = 1$.

This enrichment property is also shared by the free algebra monads of [FMS21], as they prove in Corollary 4.14, so in this direction, our framework is more general than theirs.

In a sense, $\widehat{\mathcal{P}}$ can be seen as a *trivial* monad lifting of $\mathcal{P}_{\text{ne}}$ because we simply viewed the equations presenting $\mathcal{P}_{\text{ne}}$ as quantitative equations as we did in (3.43), and we added nothing else. After this example, you may want to conjecture that whenever $\hat{E}$ is constructed from E like that, then $\hat{E}$ presents a monad lifting of the $\mathcal{T}_{\Sigma,E}$, or equivalently thanks to Corollary 3.82 and Theorem 3.84, $\hat{E}$ is an extension of E. That is not true. We showed in [MSV21, Theorem 44] that $\hat{E}$ can sometimes prove more equations than E. This implies $U\widehat{\mathcal{T}}_{\Sigma,\hat{E}}\mathbf{X} \neq \mathcal{T}_{\Sigma,E}X$, so $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}$ is not a monad lifting of $\mathcal{T}_{\Sigma,E}$.

We end this chapter with a final example, the one that motivated a lot of ideas in this manuscript.

Example 3.88 (ŁK)**.** The ŁK distance on probability distributions defined in (3.3) defines a mere lifting $(X, d) \mapsto (\mathcal{D}X, d_{\text{ŁK}})$ of $\mathcal{D}$ to $\mathbf{GMet} = [0,1]\mathbf{Spa}$.[486] We show this is a monad lifting of $(\mathcal{D}, \eta, \mu)$ (as defined in Example 1.53) with Lemmas 3.89–3.91.

[486] Of course, you can take $[0,\infty]\mathbf{Spa}$ as well. You can also show that this mere lifting preserves the satisfaction of all the equations defining metric spaces except reflexivity ($x \vdash x =_0 x$). Indeed, we have $d_{\text{ŁK}}(\varphi, \varphi) = 0$ if and only if $d(x, y) = 0$ for all $x, y \in \text{supp}(\varphi)$ (if d is reflexive, this forces $\varphi = \delta_x$). For instance, you can take **GMet** to be the category of diffuse metric spaces as we did in [MSV22, §5.3].

Lemma 3.89. *If $f : (X, d) \to (Y, \Delta)$ is nonexpansive, then so is $\mathcal{D}f : (\mathcal{D}X, d_{\text{ŁK}}) \to (\mathcal{D}Y, \Delta_{\text{ŁK}})$.*

Proof. Let $\varphi, \psi \in \mathcal{D}X$, we have

$$\begin{aligned}
&d_{\text{ŁK}}(\mathcal{D}f(\varphi), \mathcal{D}f(\psi)) \\
&= \sum_{(y,y')} \mathcal{D}f(\varphi)(y)\mathcal{D}f(\psi)(y')\Delta(y, y') \\
&= \sum_{(y,y')} \left(\sum_{x \in f^{-1}(y)} \varphi(x) \right) \left(\sum_{x' \in f^{-1}(y')} \psi(x') \right) \Delta(y, y') && \text{definition of } \mathcal{D}f \text{ (1.42)} \\
&= \sum_{(y,y')} \sum_{x \in f^{-1}(y)} \sum_{x' \in f^{-1}(y')} \varphi(x)\psi(x')\Delta(y, y') \\
&= \sum_{(x,x')} \varphi(x)\psi(x')\Delta(f(x), f(x')) \\
&\leq \sum_{(x,x')} \varphi(x)\psi(x')d(f(x), f(x')) && f \text{ is nonexpansive} \\
&= d_{\text{ŁK}}(\varphi, \psi). && \text{definition of } d_{\text{ŁK}} \qquad \square
\end{aligned}$$

Lemma 3.90. *For any (X, d), the map $\eta_X : (X, d) \to (\mathcal{D}X, d_{\text{ŁK}})$ is nonexpansive.*

Proof. For any $a, a' \in X$, we have[487]

$$d_{\text{ŁK}}(\delta_a, \delta_{a'}) \stackrel{(3.3)}{=} \sum_{(x,x')} \delta_a(x)\delta_{a'}(x')d(x, x') = \delta_a(a)\delta_{a'}(a')d(a, a') = d(a, a'). \qquad \square$$

[487] Notice that η_X is even an isometric embedding.

Lemma 3.91. *For any (X, d), the map $\mu_X : (\mathcal{D}\mathcal{D}X, d_{\text{ŁKŁK}}) \to (\mathcal{D}X, d_{\text{ŁK}})$ is nonexpansive.*

Proof. For any $\Phi, \Psi \in \mathcal{D}\mathcal{D}X$, we have

$$\begin{aligned}
d_{ŁK}(\mu_X\Phi, \mu_X\Psi) &\overset{(3.3)}{=} \sum_{(x,x')} \mu_X\Phi(x)\mu_X\Psi(x')d(x,x') \\
&\overset{(1.43)}{=} \sum_{(x,x')} \left(\sum_{\varphi \in \mathrm{supp}(\Phi)} \Phi(\varphi)\varphi(x) \right) \left(\sum_{\psi \in \mathrm{supp}(\Psi)} \Psi(\psi)\psi(x') \right) d(x,x') \\
&= \sum_{(x,x')} \sum_{(\varphi,\psi)} \Phi(\varphi)\varphi(x)\Psi(\psi)\psi(x')d(x,x') \\
&= \sum_{(\varphi,\psi)} \Phi(\varphi)\Psi(\psi) \left(\sum_{(x,x')} \varphi(x)\psi(x')d(x,x') \right) \\
&\overset{(3.3)}{=} \sum_{\varphi,\psi} \Phi(\varphi)\Psi(\psi)d_{ŁK}(\varphi,\psi) \\
&\overset{(3.3)}{=} d_{ŁKŁK}(\Phi,\Psi)
\end{aligned}$$

□

Let us denote this monad lifting by $\mathcal{D}_{ŁK}$. In [MSV22, §5.3], we gave a relatively simple quantitative algebraic presentation for $\mathcal{D}_{ŁK}$, but Theorem 3.84 will help us find a simpler one. Since, by Example 1.67, the theory of convex algebras generated by $(\Sigma_{\mathbf{CA}}, E_{\mathbf{CA}})$ presents $\mathcal{D}$ (via a monad isomorphism that we write ρ), the theorem gives us a theory presenting $\mathcal{D}_{ŁK}$ generated by $\hat{E}_1 \cup \hat{E}_2$ where

$$\begin{aligned}
\hat{E}_1 &= \{\mathbf{X}_\top \vdash s = t \mid X \vdash s = t \in E_{\mathbf{CA}}\} \text{ and} \\
\hat{E}_2 &= \left\{(X,d) \vdash s =_\varepsilon t \mid \varepsilon = d_{ŁK}\left(\rho_X[s]_{E_{\mathbf{CA}}}, \rho_X[t]_{E_{\mathbf{CA}}}\right)\right\}.
\end{aligned}$$

In order to simplify $\hat{E}_2$, we rely on two property that $d_{ŁK}$ has (one symmetric to the other) : for any $\varphi, \varphi', \psi \in \mathcal{D}X$ and $p \in [0,1]$,

$$d_{ŁK}(p\varphi + \overline{p}\varphi', \psi) = pd_{ŁK}(\varphi,\psi) + \overline{p}d_{ŁK}(\varphi',\psi) \text{ and} \tag{3.56}$$

$$d_{ŁK}(\varphi, p\varphi + \overline{p}\varphi') = pd_{ŁK}(\psi,\varphi) + \overline{p}d_{ŁK}(\psi,\varphi'). \tag{3.57}$$

Intuitively, this means that we can compute the distance between s and t by decomposing the terms into their variables, computing simple distances, then combining them to get back to s and t.[488] Formally, we only need to keep the quantitative equations in $\hat{E}_2$ that belong to[489]

$$\begin{aligned}
\hat{E}_2' = &\{x =_{\varepsilon_1} y, x =_{\varepsilon_2} z \vdash x =_{p\varepsilon_1 + \overline{p}\varepsilon_2} y +_p z \mid \varepsilon_1, \varepsilon_2 \in [0,1], p \in (0,1)\} \\
&\cup \{y =_{\varepsilon_1} x, z =_{\varepsilon_2} x \vdash y +_p z =_{p\varepsilon_1 + \overline{p}\varepsilon_2} x \mid \varepsilon_1, \varepsilon_2 \in [0,1], p \in (0,1)\}.
\end{aligned}$$

We will prove that for any $\hat{\mathbb{A}} \in \mathbf{QAlg}(\Sigma_{\mathbf{CA}})$, $\hat{\mathbb{A}} \vDash \hat{E}_1 \cup \hat{E}_2'$ implies $\hat{\mathbb{A}} \vDash \hat{E}_1 \cup \hat{E}_2$.[490] Suppose $\hat{\mathbb{A}} \vDash \hat{E}_1 \cup \hat{E}_2'$, we proceed by induction on the structure of s and t to show that $\hat{\mathbb{A}}$ satisfies $(X,d) \vdash s =_\varepsilon t$, where $\varepsilon = d_{ŁK}\left(\rho_X[s]_{E_{\mathbf{CA}}}, \rho_X[t]_{E_{\mathbf{CA}}}\right)$.

If s and t are variables, then $\rho_X[s]_{E_{\mathbf{CA}}} = \delta_x$ and $\rho_X[t]_{E_{\mathbf{CA}}} = \delta_y$ for some $x, y \in X$, thus $\varepsilon = d(x,y)$ and $(X,d) \vdash x =_{d(x,y)} y$ is satisfied by $\hat{\mathbb{A}}$ (by 3.27).

[488] This is very similar to what happens for the Kantorovich distance and (3.10).

[489] If you have symmetry ($x =_\varepsilon y \vdash y =_\varepsilon x$) as an axiom in **GMet** already, you can keep only one of these sets.

[490] It follows that $\mathfrak{Q}\mathfrak{T}\mathfrak{h}(\hat{E}_1 \cup \hat{E}_2') = \mathfrak{Q}\mathfrak{T}\mathfrak{h}(\hat{E}_1 \cup \hat{E}_2)$ because we already have the $\supseteq$ inclusion as explained in Footnote 484.

Otherwise, without loss of generality,[491] we write $t = t_1 +_p t_2$, and let $\varepsilon_i = d_{ŁK}\left(\rho_X[s]_{E_{\mathbf{CA}}}, \rho_X[t_i]\right)$. By the induction hypothesis, $\hat{\mathbb{A}} \vDash (X,d) \vdash s =_{\varepsilon_i} t_i$ for $i = 1,2$. Then, we define a substitution map $\sigma : \{x,y,z\} \to \mathcal{T}_\Sigma X$ with $x \mapsto s$, $y \mapsto t_1$ and $z \mapsto t_2$, and since $\hat{\mathbb{A}}$ satisfies $x =_{\varepsilon_1} y, x =_{\varepsilon_2} z \vdash x =_{p\varepsilon_1 + \overline{p}\varepsilon_2} y +_p z \in \hat{E}_2'$, we can apply Lemma 3.34 to conclude $\hat{\mathbb{A}}$ satisfies $(X,d) \vdash s =_{\varepsilon'} t$ with

$$\begin{aligned}
\varepsilon' &= p d_{ŁK}\left(\rho_X[s]_{E_{\mathbf{CA}}}, \rho_X[t_1]\right) + \overline{p} d_{ŁK}\left(\rho_X[s]_{E_{\mathbf{CA}}}, \rho_X[t_2]\right) \\
&= d_{ŁK}\left(\rho_X[s]_{E_{\mathbf{CA}}}, p\rho_X[t_1] + \overline{p}\rho_X[t_2]\right) && \text{by (3.56)} \\
&= d_{ŁK}\left(\rho_X[s]_{E_{\mathbf{CA}}}, \rho_X[t_1 +_p t_2]\right) \\
&= d_{ŁK}\left(\rho_X[s]_{E_{\mathbf{CA}}}, \rho_X[t]_{E_{\mathbf{CA}}}\right) = \varepsilon.
\end{aligned}$$

We conclude that $\hat{E}_1 \cup \hat{E}_2'$ presents $\mathcal{D}_{ŁK}$.

[491] If s is a term of depth > 0 but t is a variable, you decompose s instead, and then you have to use a symmetric argument.

4 Conclusion

Edge of Existence

Yoko Shimomura

In [MPP16], the authors introduced a theoretical framework to reason algebraically about distances inside a metric space. We have made adjustments to their proposal with two main goals in mind:

1. replace metrics with a more general notion of distance, and

2. tighten the relationship with universal algebra.

The result is a theory of quantitative algebras which are algebras $(A, [\![-]\!]_A)$ paired with a distance function $d : A \times A \to \mathsf{L}$ valued in a complete lattice, and no hardcoded constraint on the interaction between $[\![-]\!]_A$ and d, in contrast with the nonexpansiveness requirement (0.1) of [MPP16].[492]

[492] It is still possible to enforce (0.1) and variants with supplementary axioms (see (3.8) and (3.9)).

We introduced a sound and complete deduction system (Figure 3.1) generalizing Birkhoff's equational logic. The judgments are quantitative equations, a closer analog to classical equations than the judgments of [MPP16].

We gave a construction for free quantitative $(\Sigma, \hat{E})$-algebras (Theorem 3.48) relative to any class $\hat{E}$ of quantitative equations, following that of free classical algebras (Proposition 1.40) almost to the T. This yielded a monad $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}$ on the category of generalized metric spaces **GMet**.

We showed that algebras for the monad $\widehat{\mathcal{T}}_{\Sigma,\hat{E}}$ coincide with the $(\Sigma, \hat{E})$-algebras (Theorem 3.66), justifying a search for quantitative algebraic presentations for monads on **GMet**, of which we gave several examples (Examples 3.69, 3.71, 3.72, 3.86, and 3.88).

Finally, we gave a sufficient condition for a distance on Σ-terms to be axiomatized with a quantitative algebraic theory (Theorem 3.84). More precisely, if M is a monad on **Set** with an algebraic presentation (Σ, E), and $\widehat{M}$ is a monad lifting of M to **GMet**, then we constructed a quantitative algebraic theory $\hat{E}$ that extends E and gives a presentation for $\widehat{M}$.

4.1 Future Work

We mention some lines of questioning that need further inverstigation.

Examples

In the original paper on quantitative algebras [MPP16], the authors gave theories axiomatizing the Hausdorff distance (Example 3.69) and the Kantorovich distance (Example 1.67). I think these are amazing examples to showcase the potential of quantitative algebraic reasoning, and I would like to find more. Several papers like [BMPP18, BMPP21, MSV21, MSV22, Ró24] contain additional examples, and most of them follow the leitmotif discussed in §3.4, namely, they are built on top of a classical algebraic theory. I believe that Theorem 3.84 will accelerate the process of developing similar examples, but some efforts are still needed.[493]

[493] I planned to include a chapter in this thesis with detailed examples and non-examples to help others in this search, but I ran out of time.

Quantitative Diagrammatic Reasoning

Diagrammatic reasoning is another generalization of algebraic reasoning that has been popular in recent years. Using string diagrams in particular, people have axiomatized languages for quantum processes [CK17], strochastic processes [Fri20], machine learning models [CGG⁺22], satisfaction of Boolean formulas [GPZ23], finite state automata [PZ23], and more. There is a gap in the literature on the combination of quantitative and diagrammatic reasoning. I am aware of only one paper [KTW17] going in this direction.

HSP Theorem

We mentioned in the introduction that Birkhoff's HSP theorem [Bir35] is a celebrated result in universal algebra. In [MPP17], the authors proved a variant of this theorem for the quantitative algebras in the original paper [MPP16]. The question of how to adapt their methods to our new framework is still open.[494] We can mention other variants of the HSP theorem in similar settings that are proven (with concrete methods) in [Wea95, BV05, Hin16, Hin17].

[494] After some unsuccessful attempts during my PhD.

In the process of abstracting universal algebra away from the category of sets, several abstract HSP theorems were proven (see, e.g., [BH76, Man76, Bar94, Bar02, ARV11, MU19]). In [MU19], Milius and Urbat prove one such result and apply it to the quantitative algebras of [MPP16]. They obtain a generalization of Mardare et al.'s result in [MPP17]. In [JMU24], the authors apply Milius and Urbat's result to a new class of algebras that are a mix between [FMS21]'s and [MSV22]'s, and it should apply to the quantitative algebras presented in this thesis,[495] but careful checks are needed.

[495] They consider arbitrary relational structures like in [FMS21], but the arities are restricted to be natural numbers only, so operations are not partial. They do not require operations to be nonexpansive in the sense of (0.1), but they achieve this with lifted signatures like in [MSV22].

There are other theoretical results that followed Mardare et al.'s introduction of quantitative algebras which could be generalized to the present work. I am most interested in their work on combining theories and monads [BMPP18, BMPP21], and in the characterization of monads which can be presented by a quantitative algebraic theory [AFMS21, FMS21, Adá22, ADV23].

Partial Operations

In classical universal algebra, a signature Σ is a set of operation symbols each equipped with an arity in $\mathbb{N}$. Then, the interpretation of an n-ary operation is a function $[\![\mathsf{op}]\!]_A : A^n \to A$, where A^n is the n-wise Cartesian product. Equivalently, we can see A^n as an exponential, namely, the set of functions from $\{1, \ldots, n\}$ to A.

In [FMS21], the arity of an operation is allowed to be an arbitrary generalized metric space on $[n] = \{1, \ldots, n\}$.[496] Then, the interpretation of a $([n], d)$-ary operation symbol is a nonexpansive map $[\![\mathsf{op}]\!]_A : \mathbf{A}^{([n],d)} \to \mathbf{A}$. The definition of $\mathbf{A}^{([n],d)}$ is out of scope (it is not an exponential in the sense of cartesian closed categories), but it is a generalized metric on the set of nonexpansive maps $([n], d) \to \mathbf{A}$ with two notable consequences.

[496] We are simplifying to keep things light and closer to our work. They actually allow infinite arities and arbitrary relational structures.

1. The carrier of $\mathbf{A}^{([n],d)}$ does not necessarily contain all the functions from $[n]$ to A, so $[\![\mathsf{op}]\!]_A$ may not be applicable to all n-tuples of elements in A.

2. When d is the discrete generalized metric on $[n]$ (recall Example 3.50), the carrier of $\mathbf{A}^{([n],d)}$ is all of A^n, and the nonexpansiveness of $[\![\mathsf{op}]\!]_A$ translates to the original requirement (0.1) of [MPP16].

It is not known how to keep the flexibility of Item 1 to deal with *partial* operations without the constraint of Item 2. Namely, $[\![\mathsf{op}]\!]_A$ should be a function from the carrier of $\mathbf{A}^{([n],d)}$ to the carrier of $\mathbf{A}$ that is not necessarily nonexpansive. This would combine the generality of both [FMS21]'s and our algebras.

Applications

Many ad-hoc methods for combining algebraic reasoning with various structures like metrics or orders to reason about program semantics already exist in the literature.[497] Our abstract framework could allow viewing several of these examples under the same lens, and facilitate the discovery of new similar methods.

[497] See, e.g., [CPV16, BBKK18, DLHLP22, Sch22a, Sch22b].

With applications in mind, we can mention term rewriting systems [BKdV03] which are a popular approach to compute *in an actual computer* with classical equations. Gavazzo and Di Florio gave a very elegant account of quantitative rewriting systems in [GD23]. It seems our approaches are complementary because they replaced $[0, \infty]$ with an arbitrary quantale (a kind of complete lattice), and they also rework the nonexpansiveness assumption (0.1) in [GD23, §6].

Bibliography

[ABH$^+$12] Jiří Adámek, Filippo Bonchi, Mathias Hülsbusch, Barbara König, Stefan Milius, and Alexandra Silva. A coalgebraic perspective on minimization and determinization. In Lars Birkedal, editor, *Foundations of Software Science and Computational Structures*, page 58–73, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[ACT10] Andrei Akhvlediani, Maria Manuel Clementino, and Walter Tholen. On the categorical meaning of hausdorff and gromov distances, i. *Topology and its Applications*, 157(8):1275–1295, 2010. Advances in Set-Theoretic Topology: Proceedings of the Conference in Honour of Professor Tsugunori Nogura on his Sixtieth Birthday (9–19 June 2008, Erice, Sicily, Italy).

[Adá22] Jiří Adámek. Varieties of quantitative algebras and their monads. In *Proceedings of the 37th Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS '22, New York, NY, USA, 2022. Association for Computing Machinery.

[ADV22] Jiří Adámek, Matej Dostál, and Jiří Velebil. A categorical view of varieties of ordered algebras. *Math. Struct. Comput. Sci.*, 32(4):349–373, 2022.

[ADV23] Jirí Adámek, Matej Dostál, and Jirí Velebil. Strongly finitary monads for varieties of quantitative algebras. In Paolo Baldan and Valeria de Paiva, editors, *10th Conference on Algebra and Coalgebra in Computer Science, CALCO 2023, June 19-21, 2023, Indiana University Bloomington, IN, USA*, volume 270 of *LIPIcs*, pages 10:1–10:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023.

[AFMS21] Jiří Adámek, Chase Ford, Stefan Milius, and Lutz Schröder. Finitary monads on the category of posets. *Math. Struct. Comput. Sci.*, 31(7):799–821, 2021.

[AHS06] Jiří Adámek, Horst Herrlich, and George E Strecker. Abstract and concrete categories; the joy of cats 2004. *Reprints in Theory and Applications of Categories*, 17:1–507, 2006. Originally published by: John Wiley and Sons, New York, 1990.

[AMMU15] Jiří Adámek, Stefan Milius, Lawrence S. Moss, and Henning Urbat. On finitary functors and their presentations. *J. Comput. Syst. Sci.*, 81(5):813–833, 2015.

[ANR85] Jiří Adámek, Evelyn Nelson, and Jan Reiterman. The birkhoff variety theorem for continuous algebras. *Algebra Universalis*, 20(3):328–350, 10 1985.

[ARV11] J. Adámek, J. Rosický, and E. M. Vitale. *Algebraic theories. A categorical introduction to general algebra. With a foreword by F. W. Lawvere*, volume 184 of *Camb. Tracts Math.* Cambridge: Cambridge University Press, 2011.

[AW20] Alessandro Aldini and Herbert Wiklicky, editors. *Proceedings 16th Workshop on Quantitative Aspects of Programming Languages and Systems, QAPL@ETAPS 2019, Prague, Czech Republic, 7th April 2019*, volume 312 of *EPTCS*, 2020.

[Awo10] Steve Awodey. *Category Theory*. Oxord University Press, 2010.

[Bar94] Michael Barr. Functorial semantics and HSP type theorems. *Algebra Univers.*, 31(2):223–251, 1994.

[Bar02] Michael Barr. HSP subcategories of Eilenberg-Moore algebras. *Theory Appl. Categ.*, 10:461–468, 2002.

[Bar06] Michael Barr. Relational algebras. In *Reports of the Midwest Category Seminar IV*, page 39–55. Springer, 2006.

[Bau19] Andrej Bauer. What is algebraic about algebraic effects and handlers?, 2019.

[BBKK18] Paolo Baldan, Filippo Bonchi, Henning Kerstan, and Barbara König. Coalgebraic behavioral metrics. *Log. Methods Comput. Sci.*, 14(3), 2018.

[BBLM18a] Giorgio Bacci, Giovanni Bacci, Kim G. Larsen, and Radu Mardare. A Complete Quantitative Deduction System for the Bisimilarity Distance on Markov Chains. *Logical Methods in Computer Science*, Volume 14, Issue 4, November 2018.

[BBLM18b] Giorgio Bacci, Giovanni Bacci, Kim G. Larsen, and Radu Mardare. Complete axiomatization for the total variation distance of markov chains. *Electronic Notes in Theoretical Computer Science*, 336:27–39, 2018. The Thirty-third Conference on the Mathematical Foundations of Programming Semantics (MFPS XXXIII).

[BD80] Francis Borceux and Brian Day. Universal algebra in a closed category. *Journal of Pure and Applied Algebra*, 16(2):133–147, 1980.

[BE99] Nuno Barreiro and Thomas Ehrhard. Quantitative semantics revisited. In Jean-Yves Girard, editor, *Typed Lambda Calculi and Applications*, pages 40–53, Berlin, Heidelberg, 1999. Springer Berlin Heidelberg.

[BG19] John Bourke and Richard Garner. Monads and theories. *Advances in Mathematics*, 351:1024–1071, 2019.

[BH76] B. Banaschewski and Horst Herrlich. Subcategories defined by implications. *Houston J. Math.*, 2:149–171, 1976.

[BHKR15] Marcello M. Bonsangue, Helle Hvid Hansen, Alexander Kurz, and Jurriaan Rot. Presenting Distributive Laws. *Logical Methods in Computer Science*, Volume 11, Issue 3, August 2015.

[Bir33] Garrett Birkhoff. On the combination of subalgebras. *Mathematical Proceedings of the Cambridge Philosophical Society*, 29(4):441–464, 1933.

[Bir35] Garrett Birkhoff. On the structure of abstract algebras. *Mathematical Proceedings of the Cambridge Philosophical Society*, 31(4):433–454, 1935.

[BKdV03] Marc Bezem, Jan Willem Klop, and Roel de Vrijer, editors. *Terese. Term rewriting systems*, volume 55 of *Camb. Tracts Theor. Comput. Sci.* Cambridge: Cambridge University Press, 2003.

[Blo76] Stephen L. Bloom. Varieties of ordered algebras. *Journal of Computer and System Sciences*, 13(2):200–212, 1976.

[BMPP18] Giorgio Bacci, Radu Mardare, Prakash Panangaden, and Gordon D. Plotkin. An algebraic theory of markov processes. In Anuj Dawar and Erich Grädel, editors, *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2018, Oxford, UK, July 09-12, 2018*, page 679–688. ACM, 2018.

[BMPP21] Giorgio Bacci, Radu Mardare, Prakash Panangaden, and Gordon D. Plotkin. Tensor of quantitative equational theories. In Fabio Gadducci and Alexandra Silva, editors, *9th Conference on Algebra and Coalgebra in Computer Science, CALCO 2021, August 31 to September 3, 2021, Salzburg, Austria*, volume 211 of *LIPIcs*, page 7:1–7:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021.

[Bor94] Francis Borceux. *Handbook of Categorical Algebra*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1994.

[Bou92] Dominique Bourn. Low dimensional geometry of the notion of choice. In *Category Theory 1991, CMS Conference Proceedings*, volume 13, page 55–73, 1992.

[BP98] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1):107–117, 1998. Proceedings of the Seventh International World Wide Web Conference.

[BP15] Filippo Bonchi and Damien Pous. Hacking nondeterminism with induction and coinduction. *Commun. ACM*, 58(2):87–95, jan 2015.

[Bra00] A. Branciari. A fixed point theorem of Banach-Caccioppoli type on a class of generalized metric spaces. *Publ. Math. Debr.*, 57(1-2):31–37, 2000.

[BSS21] Filippo Bonchi, Alexandra Silva, and Ana Sokolova. Distribution bisimilarity via the power of convex algebras. *Logical Methods in Computer Science*, Volume 17, Issue 3, July 2021.

[BSV19] Filippo Bonchi, Ana Sokolova, and Valeria Vignudelli. The theory of traces for systems with nondeterminism and probability. In *34th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2019, Vancouver, BC, Canada, June 24-27, 2019*, page 1–14. IEEE, 2019.

[BSV22] Filippo Bonchi, Ana Sokolova, and Valeria Vignudelli. The theory of traces for systems with nondeterminism, probability, and termination. *Logical Methods in Computer Science*, Volume 18, Issue 2, June 2022.

[BV05] Radim Bělohlávek and Vilém Vychodil. *Fuzzy equational logic*, volume 186 of *Stud. Fuzziness Soft Comput.* Berlin: Springer, 2005.

[BvBR98] Marcello M. Bonsangue, Franck van Breugel, and Jan J. M. M. Rutten. Generalized metric spaces: Completion, topology, and powerdomains via the yoneda embedding. *Theor. Comput. Sci.*, 193(1-2):1–51, 1998.

[BW05] Michael Barr and Charles Wells. Toposes, triples and theories. *Reprints in Theory and Applications of Categories*, 12:1–287, 2005. Originally published by: John Wiley and Sons, New York, 1990.

[CDL15] Raphaelle Crubille and Ugo Dal Lago. Metric reasoning about λ-terms: The affine case. In *Proceedings of the 2015 30th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, LICS '15, page 633–644, USA, 2015. IEEE Computer Society.

[CDL17] Raphaëlle Crubillé and Ugo Dal Lago. Metric reasoning about λ-terms: The general case. In Hongseok Yang, editor, *Programming Languages and Systems*, pages 341–367, Berlin, Heidelberg, 2017. Springer Berlin Heidelberg.

[CGG+22] Geoffrey S. H. Cruttwell, Bruno Gavranović, Neil Ghani, Paul Wilson, and Fabio Zanasi. Categorical foundations of gradient-based learning. In Ilya Sergey, editor, *Programming Languages and Systems*, pages 1–28, Cham, 2022. Springer International Publishing.

[CH06] Maria Manuel Clementino and Dirk Hofmann. Exponentiation in v-categories. *Topology and its Applications*, 153(16):3113–3128, 2006. Special Issue: Aspects of Contemporary Topology.

[Che16] Eugenia Cheng. *How to Bake PI*. Basic Books, 5 2016.

[CK17] Bob Coecke and Aleks Kissinger. *Picturing quantum processes. A first course in quantum theory and diagrammatic reasoning*. Cambridge: Cambridge University Press, 2017.

[CKPR21] Pablo Samuel Castro, Tyler Kastner, Prakash Panangaden, and Mark Rowland. Mico: Improved representations via sampling-based state similarity for markov decision processes. *Advances in Neural Information Processing Systems*, 34:30113–30126, 2021.

[CM22a] Davide Castelnovo and Marino Miculan. Fuzzy Algebraic Theories. In Florin Manea and Alex Simpson, editors, *30th EACSL Annual Conference on Computer Science Logic (CSL 2022)*, volume 216 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 13:1–13:17, Dagstuhl, Germany, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[CM22b] Alex Citkin and Alexei Muravitsky. *Consequence Relations: An Introduction to the Lindenbaum-Tarski Method*. Oxford University Press, 07 2022. Available at `https://arxiv.org/abs/2106.10966`.

[CMS23] Christopher Conlon, Julie Holland Mortimer, and Paul Sarkis. Estimating preferences and substitution patterns from second-choice data alone. Working Paper, Accessed on: 2023-12-08, 2023.

[Col24] Thomas Colcombet. knowledge. LaTeX package, 2024. `https://ctan.org/pkg/knowledge`.

[Con17] Gabriel Conant. Distance structures for generalized metric spaces. *Annals of Pure and Applied Logic*, 168(3):622–650, 2017.

[CPV16] Konstantinos Chatzikokolakis, Catuscia Palamidessi, and Valeria Vignudelli. Up-To Techniques for Generalized Bisimulation Metrics. In *27th International Conference on Concurrency Theory (CONCUR 2016)*, volume 59 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 35:1–35:14, Québec City, Canada, August 2016.

[Cur29] Haskell B. Curry. An analysis of logical substitution. *American Journal of Mathematics*, 51(3):363–384, 1929.

[DFM23] Francesco Dagnino, Amin Farjudian, and Eugenio Moggi. Robustness in metric spaces over continuous quantales and the hausdorff-smyth monad. In Erika Ábrahám, Clemens Dubslaff, and Silvia Lizeth Tapia Tarifa, editors, *Theoretical Aspects of Computing – ICTAC 2023*, page 313–331, Cham, 2023. Springer Nature Switzerland.

[DGY19] Ugo Dal Lago, Francesco Gavazzo, and Akira Yoshimizu. Differential logical relations, part i: The simply-typed case. In Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi, editors, *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*, volume 132 of *Leibniz International Proceedings in Informatics (LIPIcs)*, page 111:1–111:14, Dagstuhl, Germany, 2019. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[DLHLP22] Ugo Dal Lago, Furio Honsell, Marina Lenisa, and Paolo Pistone. On Quantitative Algebraic Higher-Order Theories. In Amy P. Felty, editor, *7th International Conference on Formal Structures for Computation and Deduction (FSCD 2022)*, volume 228 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 4:1–4:18, Dagstuhl, Germany, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[DLPS07] Christian Delhommé, Claude Laflamme, Maurice Pouzet, and Norbert Sauer. Divisibility of countable metric spaces. *European Journal of Combinatorics*, 28(6):1746–1769, 2007.

[DP02] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, 2 edition, 2002.

[DPS18] Fredrik Dahlqvist, Louis Parlant, and Alexandra Silva. Layer by layer – combining monads. In Bernd Fischer and Tarmo Uustalu, editors, *Theoretical Aspects of Computing – ICTAC 2018*, page 153–172, Cham, 2018. Springer International Publishing.

[Dub70] E. J. Dubuc. Enriched semantics-structure (meta) adjointness. *Revista de la Unión Matemática Argentina*, 25:5–26, 1970.

[Dwo06] Cynthia Dwork. Differential privacy. In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, editors, *Automata, Languages and Programming*, pages 1–12, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[EE86] Paul Erdös and Marcel Erné. Clique numbers of graphs. *Discrete Mathematics*, 59(3):235–242, 1986.

[EGP07] Marcel Erné, Mai Gehrke, and Aleš Pultr. Complete congruences on topologies and down-set lattices. *Applied Categorical Structures*, 15(1):163–184, Apr 2007.

[EM45] Samuel Eilenberg and Saunders MacLane. General theory of natural equivalences. *Transactions of the American Mathematical Society*, 58(2):231–294, 1945.

[EM65] Samuel Eilenberg and John C Moore. Adjoint functors and triples. *Illinois Journal of Mathematics*, 9(3):381–398, 1965.

[Esc99] Martin Escardo. A metric model of pcf. In *Workshop on Realizability Semantics and Applications*, volume 417, page 418. Citeseer, 1999.

[FGSW21] Uli Fahrenberg, Mai Gehrke, Luigi Santocanale, and Michael Winter, editors. *Relational and Algebraic Methods in Computer Science*, Lecture Notes in Computer Science. Springer Cham, 2021.

[FH11] Marcelo Fiore and Chung-Kil Hur. On the mathematical synthesis of equational logics. *Logical Methods in Computer Science*, Volume 7, Issue 3, September 2011.

[Fla97] R. C. Flagg. Quantales and continuity spaces. *Algebra Universalis*, 37(3):257–276, Jun 1997.

[FMS21] Chase Ford, Stefan Milius, and Lutz Schröder. Behavioural preorders via graded monads. In *Proceedings of the 36th Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS '21, New York, NY, USA, 2021. Association for Computing Machinery.

[FMS21] Chase Ford, Stefan Milius, and Lutz Schröder. Monads on Categories of Relational Structures. In Fabio Gadducci and Alexandra Silva, editors, *9th Conference on Algebra and Coalgebra in Computer Science (CALCO 2021)*, volume 211 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 14:1–14:17, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[FP19] Tobias Fritz and Paolo Perrone. A probability monad as the colimit of spaces of finite samples. *Theory and Applications of Categories*, 34(7):170–220, 2019.

[Fré06] Maurice Fréchet. Sur quelques points du calcul fonctionnel. *Rend. Circ. Mat. Palermo*, 22:1–74, 1906.

[Fri20] Tobias Fritz. A synthetic approach to markov kernels, conditional independence and theorems on sufficient statistics. *Advances in Mathematics*, 370:107239, 2020.

[FSW+23] Jonas Forster, Lutz Schröder, Paul Wild, Harsh Beohar, Sebastian Gurke, Barbara König, and Karla Messing. Graded semantics and graded logics for eilenberg-moore coalgebras. *arXiv preprint arXiv:2307.14826*, 2023.

[GD23] Francesco Gavazzo and Cecilia Di Florio. Elements of quantitative rewriting. *Proc. ACM Program. Lang.*, 7(POPL), jan 2023.

[Gir82] Michèle Giry. A categorical approach to probability theory. In B. Banaschewski, editor, *Categorical Aspects of Topology and Analysis*, page 68–85, Berlin, Heidelberg, 1982. Springer Berlin Heidelberg.

[Gir88] Jean-Yves Girard. Normal functors, power series and λ-calculus. *Annals of Pure and Applied Logic*, 37(2):129–177, 1988.

[GL19] Jean Goubault-Larrecq. Formal ball monads. *Topology and its Applications*, 263:372–391, 2019.

[GMS+23] Sergey Goncharov, Stefan Milius, Lutz Schröder, Stelios Tsampas, and Henning Urbat. Towards a higher-order mathematical operational semantics. *Proc. ACM Program. Lang.*, 7(POPL), jan 2023.

[GP98] R. Gordon and A.J. Power. Algebraic structure for bicategory enriched categories. *Journal of Pure and Applied Algebra*, 130(2):119–132, 1998.

[GP21] Guillaume Geoffroy and Paolo Pistone. A partial metric semantics of higher-order types and approximate program transformations. In *CSL 2021 - Computer Science Logic*, Lubjana, Slovenia, January 2021.

[GPA21] Alexandre Goy, Daniela Petrişan, and Marc Aiguier. Powerset-like monads weakly distribute over themselves in toposes and compact hausdorff spaces. In Nikhil Bansal, Emanuela Merelli, and James Worrell, editors, *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*, volume 198 of *Leibniz International Proceedings in Informatics (LIPIcs)*, page 132:1–132:14, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[GPR16] Mai Gehrke, Daniela Petrisan, and Luca Reggio. The schützenberger product for syntactic spaces. In Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani, and Davide Sangiorgi, editors, *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*, volume 55 of *Leibniz International Proceedings in Informatics (LIPIcs)*, page 112:1–112:14, Dagstuhl, Germany, 2016. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[GPZ23] Tao Gu, Robin Piedeleu, and Fabio Zanasi. A Complete Diagrammatic Calculus for Boolean Satisfiability. *Electronic Notes in Theoretical Informatics and Computer Science*, Volume 1 - Proceedings of MFPS XXXVIII, February 2023.

[GS21] Fabio Gadducci and Alexandra Silva, editors. *LIPIcs, Volume 211, CALCO 2021, Complete Volume*, volume 211 of *Leibniz International Proceedings in Informatics (LIPIcs)*, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[GTWW77] J. A. Goguen, J. W. Thatcher, E. G. Wagner, and J. B. Wright. Initial algebra semantics and continuous algebras. *J. ACM*, 24(1):68–95, jan 1977.

[Hau14] Felix Hausdorff. *Grundzüge der Mengenlehre.* 1914.

[HHL22] André Hirschowitz, Tom Hirschowitz, and Ambroise Lafont. Modules over monads and operational semantics (expanded version). *Logical Methods in Computer Science*, Volume 18, Issue 3, August 2022.

[Hin16] Wataru Hino. Varieties of metric and quantitative algebras, 2016.

[Hin17] Wataru Hino. Continuous varieties of metric and quantitative algebras, 2017.

[HP07] Martin Hyland and John Power. The category theoretic understanding of universal algebra: Lawvere theories and monads. *Electronic Notes in Theoretical Computer Science*, 172:437–458, 2007. Computation, Meaning, and Logic: Articles dedicated to Gordon Plotkin.

[HPP06] Martin Hyland, Gordon Plotkin, and John Power. Combining effects: Sum and tensor. *Theoretical Computer Science*, 357(1):70–99, 2006. Clifford Lectures and the Mathematical Foundations of Programming Semantics.

[HR13] Dirk Hofmann and C.D. Reis. Probabilistic metric spaces as enriched categories. *Fuzzy Sets and Systems*, 210:1–21, 2013. Theme : Topology and Algebra.

[HS00] Pascal Hitzler and Anthony Karel Seda. Dislocated topologies. *J. Electr. Eng*, 51(12):3–7, 2000.

[HST14] Dirk Hofmann, Gavin J Seal, and Walter Tholen. *Monoidal Topology: A Categorical Approach to Order, Metric, and Topology*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2014.

[Jac16] Bart Jacobs. *Introduction to Coalgebra: Towards Mathematics of States and Observation*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 2016.

[JMU24] Jan Jurka, Stefan Milius, and Henning Urbat. Algebraic reasoning over relational structures. *CoRR*, abs/2401.08445, 2024.

[KP93] G.M. Kelly and A.J. Power. Adjunctions whose counits are coequalizers, and presentations of finitary enriched monads. *Journal of Pure and Applied Algebra*, 89(1):163–179, 1993.

[KP22] H. Kunzi and H. Pajoohesh. b-metrics. *Topology and its Applications*, 309:107905, 2022. Remembering Hans-Peter Kunzi (1955-2020).

[KS18] Bartek Klin and Julian Salamanca. Iterated covariant powerset is not a monad. *Electronic Notes in Theoretical Computer Science*, 341:261–276, 2018. Proceedings of the Thirty-Fourth Conference on the Mathematical Foundations of Programming Semantics (MFPS XXXIV).

[KTW17] Aleks Kissinger, Sean Tull, and Bas Westerbaan. Picture-perfect quantum key distribution, 2017.

[KV14] Delia Kesner and Daniel Ventura. Quantitative types for the linear substitution calculus. In Josep Diaz, Ivan Lanese, and Davide Sangiorgi, editors, *Theoretical Computer Science*, pages 296–310, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

[KV17] Alexander Kurz and Jiří Velebil. Quasivarieties and varieties of ordered algebras: regularity and exactness. *Mathematical Structures in Computer Science*, 27(7):1153–1194, 2017.

[KWG15] Bil Kleb, Bill Wood, and Kevin Godby. tufte-latex. LaTeX package, 2015.

[Kwi07] Marta Kwiatkowska. Quantitative verification: models techniques and tools. In *Proceedings of the the 6th Joint Meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on The Foundations of Software Engineering*, ESEC-FSE '07, page 449–458, New York, NY, USA, 2007. Association for Computing Machinery.

[KyKK+21] Yuichi Komorida, Shin ya Katsumata, Clemens Kupke, Jurriaan Rot, and Ichiro Hasuo. Expressivity of quantitative modal logics : Categorical foundations via codensity and approximation. In *2021 36th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, page 1–14, 2021.

[Lac07] S. Lack. Homotopy-theoretic aspects of 2-monads. *Journal of Homotopy and Related Structures*, 2(2):229–260, 2007.

[Law62] F. W. Lawvere. The category of probabilistic mappings. *preprint*, 1962.

[Law63] F. W. Lawvere. Functorial semantics of algebraic theories. *Proc. Natl. Acad. Sci. USA*, 50:869–872, 1963.

[Law02] F William Lawvere. Metric spaces, generalized logic, and closed categories. *Reprints in Theory and Applications of Categories*, 1:1–37, 2002. Originally published in Rendiconti del seminario matématico e fisico di Milano, XLIII (1973).

[Lin66] F. E. J. Linton. Some aspects of equational categories. In S. Eilenberg, D. K. Harrison, S. MacLane, and H. Röhrl, editors, *Proceedings of the Conference on Categorical Algebra*, page 84–94, Berlin, Heidelberg, 1966. Springer Berlin Heidelberg.

[Lin69] F. E. J. Linton. An outline of functorial semantics. In B. Eckmann, editor, *Seminar on Triples and Categorical Homology Theory*, page 7–52, Berlin, Heidelberg, 1969. Springer Berlin Heidelberg.

[LJ11] Weisi Lin and C.-C. Jay Kuo. Perceptual visual quality metrics: A survey. *Journal of Visual Communication and Image Representation*, 22(4):297–312, 2011.

[LMMP13] Jim Laird, Giulio Manzonetto, Guy McCusker, and Michele Pagani. Weighted relational models of typed lambda-calculi. In *Proceedings of the 2013 28th Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS '13, page 301–310, USA, 2013. IEEE Computer Society.

[LP08] Stephen Lack and Simona Paoli. 2-nerves for bicategories. *K-Theory: interdisciplinary journal for the development, application and influence of K-theory in the mathematical sciences*, 38(2):153–175, January 2008.

[LP23] Rory B. B. Lucyshyn-Wright and Jason Parker. Diagrammatic presentations of enriched monads and varieties for a subcategory of arities. *Appl. Categorical Struct.*, 31(5):40, 2023.

[Łuk04] Szymon Łukaszyk. A new concept of probability metric and its applications in approxiomation of scattered data sets. *Computational Mechanics*, 33:299–304, 2004.

[LW16] Rory B. B. Lucyshyn-Wright. Enriched algebraic theories and monads for a system of arities. *Theory and Applications of Categories*, 31(5):101–137, 2016. Published 2016-01-31.

[LY16] Shou Lin and Ziqiu Yun. *Generalized metric spaces and mappings*, volume 6 of *Atlantis Studies in Mathematics*. Atlantis Press, [Paris], chinese edition, 2016. With a foreword by Alexander V. Arhangel'skiĭ, With a preface by Guoshi Gao.

[Mac71] Saunders Mac Lane. *Categories for the Working Mathematician*. Springer-Verlag, 2nd edition, 1971.

[Man76] Ernest G. Manes. *Algebraic theories*, volume 26 of *Grad. Texts Math.* Springer, Cham, 1976.

[Mat94] S. G. Matthews. Partial metric topology. *Annals of the New York Academy of Sciences*, 728(1):183–197, 1994.

[MB99] Saunders Mac Lane and Garrett Birkhoff. *Algebra*, volume 330. American Mathematical Soc., 1999.

[Mim20] Samuel Mimram. *PROGRAM = PROOF*. 2020. Available at `https://www.lix.polytechnique.fr/Labo/Samuel.Mimram/teaching/INF551/course.pdf`.

[MMM12] Annabelle McIver, Larissa Meinicke, and Carroll Morgan. A kantorovich-monadic powerdomain for information hiding, with probability and nondeterminism. In *2012 27th Annual IEEE Symposium on Logic in Computer Science*, pages 461–470, 2012.

[Mog89] E. Moggi. Computational lambda-calculus and monads. In *Proceedings. Fourth Annual Symposium on Logic in Computer Science*, pages 14,15,16,17,18,19,20,21,22,23, Los Alamitos, CA, USA, jun 1989. IEEE Computer Society.

[Mog91] Eugenio Moggi. Notions of computation and monads. *Information and Computation*, 93(1):55–92, 1991. Selections from 1989 IEEE Symposium on Logic in Computer Science.

[MPP17] Radu Mardare, Prakash Panangaden, and Gordon D. Plotkin. On the axiomatizability of quantitative algebras. In *32nd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2017, Reykjavik, Iceland, June 20-23, 2017*, page 1–12. IEEE Computer Society, 2017.

[MPP18] Radu Mardare, Prakash Panangaden, and Gordon D. Plotkin. Free complete wasserstein algebras. *Log. Methods Comput. Sci.*, 14(3), 2018.

[MPP21] Radu Mardare, Prakash Panangaden, and Gordon D. Plotkin. Fixed-points for quantitative equational logics. In *36th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2021, Rome, Italy, June 29 - July 2, 2021*, page 1–13. IEEE, 2021.

[MPP16] Radu Mardare, Prakash Panangaden, and Gordon D. Plotkin. Quantitative algebraic reasoning. In Martin Grohe, Eric Koskinen, and Natarajan Shankar, editors, *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science, LICS '16, New York, NY, USA, July 5-8, 2016*, page 700–709. ACM, 2016.

[MSV21] Matteo Mio, Ralph Sarkis, and Valeria Vignudelli. Combining nondeterminism, probability, and termination: Equational and metric reasoning. In *2021 36th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, page 1–14, 2021.

[MSV22] Matteo Mio, Ralph Sarkis, and Valeria Vignudelli. Beyond nonexpansive operations in quantitative algebraic reasoning. In *Proceedings of the 37th Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS '22, New York, NY, USA, 2022. Association for Computing Machinery.

[MSV23] Matteo Mio, Ralph Sarkis, and Valeria Vignudelli. Universal quantitative algebra for fuzzy relations and generalised metric spaces, July 2023.

[MU19] Stefan Milius and Henning Urbat. Equational axiomatization of algebras with structure. In Mikolaj Bojanczyk and Alex Simpson, editors, *Foundations of Software Science and Computation Structures - 22nd International Conference, FOSSACS 2019, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2019, Prague, Czech Republic, April 6-11, 2019, Proceedings*, volume 11425 of *Lecture Notes in Computer Science*, pages 400–417. Springer, 2019.

[MV20] Matteo Mio and Valeria Vignudelli. Monads and quantitative equational theories for nondeterminism and probability. In Igor Konnov and Laura Kovács, editors, *31st International Conference on Concurrency Theory, CONCUR 2020, September 1-4, 2020, Vienna, Austria (Virtual Conference)*, volume 171 of *LIPIcs*, page 28:1–28:18. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.

[Mé11] Facundo Mémoli. Gromov–wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics*, 11(4):417–487, Aug 2011.

[Ngu10] L. Nguyen Van Thé. *Structural Ramsey theory of metric spaces and topological dynamics of isometry groups*. American Mathematical Soc., 2010.

[Par22] Jason Parker. Extensivity of categories of relational structures. *Theory and Applications of Categories*, 38(23):898–912, 2022.

[Par23] Jason Parker. Exponentiability in categories of relational structures. *Theory and Applications of Categories*, 39(16):493–518, 2023.

[Pis21] Paolo Pistone. On generalized metric spaces for the simply typed lambda-calculus. In *Proceedings of the 36th Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS '21, New York, NY, USA, 2021. Association for Computing Machinery.

[Pow99] John Power. Enriched lawvere theories. *Theory and Applications of Categories*, 6(7):83–93, 1999.

[Pow05] John Power. Discrete lawvere theories. In *Proceedings of the First International Conference on Algebra and Coalgebra in Computer Science*, CALCO'05, page 348–363, Berlin, Heidelberg, 2005. Springer-Verlag.

[PP01a] Gordon Plotkin and John Power. Adequacy for algebraic effects. In Furio Honsell and Marino Miculan, editors, *Foundations of Software Science and Computation Structures*, page 1–24, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg.

[PP01b] Gordon Plotkin and John Power. Semantics for algebraic operations. In *MFPS 2001. Papers from the 17th conference on the mathematical foundations of programming semantics, Aarhus University, Aarhus, Denmark, May 23–26, 2001.*, pages 332–345. Amsterdam: Elsevier, 2001.

[PP02] Gordon Plotkin and John Power. Notions of computation determine monads. In Mogens Nielsen and Uffe Engberg, editors, *Foundations of Software Science and Computation Structures*, pages 342–356, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg.

[PRSW20] Louis Parlant, Jurriaan Rot, Alexandra Silva, and Bas Westerbaan. Preservation of Equations by Monoidal Monads. In Javier Esparza and Daniel Král', editors, *45th International Symposium on Mathematical Foundations of Computer Science (MFCS 2020)*, volume 170 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 77:1–77:14, Dagstuhl, Germany, 2020. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[PS21] Daniela Petrisan and Ralph Sarkis. Semialgebras and weak distributive laws. In Ana Sokolova, editor, *Proceedings 37th Conference on Mathematical Foundations of Programming Semantics, MFPS 2021, Hybrid: Salzburg, Austria and Online, 30th August - 2nd September, 2021*, volume 351 of *EPTCS*, pages 218–241, 2021.

[PZ23] Robin Piedeleu and Fabio Zanasi. A Finite Axiomatisation of Finite-State Automata Using String Diagrams. *Logical Methods in Computer Science*, Volume 19, Issue 1, February 2023.

[QIF20] *The Science of Quantitative Information Flow*. Information Security and Cryptography. Springer Cham, 1 edition, 2020.

[Rie17] Emily Riehl. *Category Theory in Context*. Dover Publications, 2017.

[Rob02] Edmund Robinson. Variations on algebra: Monadicity and generalisations of equational theories. *Formal Aspects of Computing*, 13(3):308–326, 07 2002.

[Ros21] Jiří Rosický. Metric monads. *Mathematical Structures in Computer Science*, 31(5):535–552, 2021.

[Ros24] Jiří Rosický. Discrete equational theories. *Mathematical Structures in Computer Science*, 34(2):147–160, 2024.

[RT23] Jiří Rosický and Giacomo Tendas. Enriched universal algebra, 2023.

[Rut96] J.J.M.M. Rutten. Elements of generalized ultrametric domain theory. *Theoretical Computer Science*, 170(1):349–381, 1996.

[RZHE24] Aloïs Rosset, Maaike Zwart, Helle Hvid Hansen, and Jörg Endrullis. Correspondence between composite theories and distributive laws, 2024.

[Ró24] Wojciech Różowski. A complete quantitative axiomatisation of behavioural distance of regular expressions, 2024.

[SB06] H.R. Sheikh and A.C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, 2006.

[Sch22a] Todd Schmid. A (co)algebraic framework for ordered processes, 2022.

[Sch22b] Todd Schmid. Presenting with quantitative inequational theories, 2022.

[Sco76] Dana Scott. Data types as lattices. *SIAM Journal on Computing*, 5(3):522–587, 1976.

[SS71] Dana Scott and Christopher Strachey. Toward a mathematical semantics for computer languages. In *Proceedings of the Symposium on Computers and Automata*, volume 21, 1971.

[SW18] Ana Sokolova and Harald Woracek. Termination in convex sets of distributions. *Logical Methods in Computer Science*, Volume 14, Issue 4, November 2018.

[Tho12] Walter Tholen. Kleisli enriched. *Journal of Pure and Applied Algebra*, 216(8):1920–1931, 2012. Special Issue devoted to the International Conference in Category Theory 'CT2010'.

[TP97] Daniele Turi and Gordon D. Plotkin. Towards a mathematical operational semantics. In *Proceedings of the Twelfth Annual IEEE Symposium on Logic in Computer Science (LICS 1997)*, pages 280–291. IEEE Computer Society Press, June 1997.

[vB01] Franck van Breugel. An introduction to metric semantics: operational and denotational models for programming and specification languages. *Theoretical Computer Science*, 258(1):1–98, 2001.

[vB05] Franck van Breugel. The metric monad for probabilistic nondeterminism. Available at `http://www.cse.yorku.ca/~franck/research/drafts/monad.pdf`, 2005.

[vBW01] Franck van Breugel and James Worrell. Towards quantitative verification of probabilistic transition systems. In Fernando Orejas, Paul G. Spirakis, and Jan van Leeuwen, editors, *Automata, Languages and Programming*, page 421–432, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg.

[Vig23] Ignacio Viglizzo. Basic constructions in the categories of sets, sets with a binary relation on them, preorders, and posets, June 2023.

[Vil09] Cédric Villani. *Optimal transport: old and new*, volume 338 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

[VK11] Jiří Velebil and Alexander Kurz. Equational presentations of functors and monads. *Mathematical Structures in Computer Science*, 21(2):363–381, 2011.

[VW06] Daniele Varacca and Glynn Winskel. Distributing probability over non-determinism. *Mathematical Structures in Computer Science*, 16(1):87–113, 2006.

[Wea93] Nikolai Weaver. Generalized varieties. *Algebra Universalis*, 30(1):27–52, 03 1993.

[Wea95] Nikolai Weaver. Quasi-varieties of metric algebras. *Algebra Universalis*, 33(1):1–9, 03 1995.

[Wec12] Wolfgang Wechler. *Universal algebra for computer scientists*, volume 25. Springer Science & Business Media, 2012.

[Wil31a] Wallace Alvin Wilson. On quasi-metric spaces. *American Journal of Mathematics*, 53(3):675–684, 1931.

[Wil31b] Wallace Alvin Wilson. On semi-metric spaces. *American Journal of Mathematics*, 53(2):361–373, 1931.

[WS20] Paul Wild and Lutz Schröder. Characteristic logics for behavioural metrics via fuzzy lax extensions. In Igor Konnov and Laura Kovács, editors, *31st International Conference on Concurrency Theory (CONCUR 2020)*, volume 171 of *Leibniz International Proceedings in Informatics (LIPIcs)*, page 27:1–27:23, Dagstuhl, Germany, 2020. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[ZK22] Linpeng Zhang and Benjamin Lucien Kaminski. Quantitative strongest post: a calculus for reasoning about the flow of quantitative information. *Proc. ACM Program. Lang.*, 6(OOPSLA1), apr 2022.

[ZM22] Maaike Zwart and Dan Marsden. No-Go Theorems for Distributive Laws. *Logical Methods in Computer Science*, Volume 18, Issue 1, January 2022.

The following is an abridged and unfinished copy of a textbook I am writing on category theory. It is included in the manuscript only as a target for the many `knowldege` links throughout the thesis. The standalone and up to date file can be found here.

My First Category Theory Textbook

An abridged and unfinished copy.

Ralph Sarkis

A Preliminaries

Our main goal here is to introduce enough notation and terminology so that this book is self-contained.[0]

We assume you are familiar and comfortable with basic concepts about sets (e.g.: subsets, union, Cartesian product, cardinality, equivalence classes, quotients, etc.), functions (e.g.: injectivity, surjectivity, inverses, (pre)image, etc.), logic (e.g.: quantifiers, implication) and proofs (e.g.: you can write, read and understand proofs),[1] and we will not recall anything here. However, we need to have a little talk about foundations.

Several times in our coverage of category theory, we will use the term **collection** in order to avoid set-theoretical paradoxes. Collections are supposed to behave just like sets except that we will never consider collections containing other collections. We do not make it more formal because there are many ways to do it (dealing with

[0] Especially with the heavy use of the `knowledge` package, I felt it was necessary to cover enough background material in order to have the least amount of external links in the book.

[1] The very first things usually taught in early undergraduate mathematics courses.

so-called **size issues**),[2] and none of them are relevant to this course.

Still, you need to know why we cannot use sets as is usual in all other courses. In short, there exist collections of objects that cannot be sets.[3] In our case, we will need to talk about the collection of all sets and the collection of all groups (among others) and they cannot form sets. For the former, it is easy to see because if S is the set of all sets, then it contains all its subsets and hence $\mathcal{P}(S) \subseteq S$, this leads to the contradiction $|\mathcal{P}(S)| \leq |S| < |\mathcal{P}(S)|$.[4]

In the rest of this chapter, we cover the necessary background that we will use in the rest of the book. It is supposed to be a quick and (unfortunately) dry overview of stuff you may or may not have seen, so we will not dwell on explanations, intuitions and motivations.[5] You can safely skip these sections and come back whenever you click on a word or symbol that is defined here. We hope that this will save you from several trips to Wikipedia.

[2] Most commonly, people use classes or Grothendieck universes. If this sticky point worries you, I suggest you keep it in the back of your mind and go read `https://arxiv.org/pdf/0810.1279.pdf` when you are a bit more comfortable with category theory.

[3] Famous examples include the collection of ordinal numbers which, by the Burali–Forti paradox, cannot be a set and the collection of all sets that do not contain themselves which, by the Russel paradox, cannot be a set.

[4] For a set X, $|X|$ denotes the **cardinal** of X and $\mathcal{P}(X)$ denotes the **powerset** of X, i.e. the set of all subsets of X. The strict inequality $|S| < |\mathcal{P}(S)|$ is due to Georg Cantor's famous diagonalization argument.

[5] Contrarily to the other chapters of this book.

A.1 Abstract Algebra

Here we recall definitions, examples and results you may have seen in classes on abstract algebra or linear algebra.[6]

[6] Monoids are not commonly covered, but they are simpler than groups and we need them at one point so we present them here.

Monoids

Definition A.1 (Monoid)**.** A **monoid** is a set M equipped with a binary operation $\cdot : M \times M \to M$ (written infix) called **multiplication** and an **identity** element[7] 1_M satisfying for all $x, y, z \in M$

$$(x \cdot y) \cdot z = x \cdot (y \cdot z) \quad \text{and} \quad 1_M \cdot x = x = x \cdot 1_M.$$

If it satisfies $\forall x, y \in M, x \cdot y = y \cdot x$, M is a **commutative monoid**.

[7] Some authors call 1_M the **unit** or the **neutral** element.

Depending on the context, we will refer to a monoid either as M or $(M, \cdot)$ or $(M, \cdot, 1_M)$.

Remark A.2. We will quickly drop the $\cdot$ symbol and denote multiplication with plain juxtaposition (i.e. $xy := x \cdot y$) for monoids and other algebraic structures with a multiplication.

Examples A.3. 1. For any set S, the set of function from S to itself forms a monoid with the multiplication being composition of functions and the identity being the identity function $s \mapsto s$. We denote this monoid by S^S.

2. The sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$[8] equipped with the operation of addition are all commutative monoids.

3. For any set S, the powerset $\mathcal{P}(S)$ has two simple monoid structures: one where the multiplication is $\cup$ and the identity is $\emptyset \subseteq S$, and the other where multiplication is $\cap$ and the identity is $S \subseteq S$.

[8] The symbols $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ denote respectively the sets of natural numbers, integers, rationals and real numbers.

Definition A.4 (Submonoid)**.** Given a monoid M, a **submonoid** of M is a subset $N \subseteq M$ containing 1_M that is closed under multiplication (i.e. $\forall x, y \in N, x \cdot y \in N$).[9]

[9] This implies N is also a monoid with the multiplication and identity inherited from M.

Example A.5. For any set S, the set of bijections from S to itself, denoted by Σ_S, is a submonoid of S^S because the composition of two bijections is bijective.

Definition A.6 (Homomorphism)**.** Let M and N be two monoids, a **monoid homomorphism** from M to N is a function $f : M \to N$ satisfying the following property:

$$f(1_M) = 1_N \quad \text{and} \quad \forall x, y \in M, f(xy) = f(x)f(y).$$

When f is a bijection, we call it a **monoid isomorphism**, say that M and N are **isomorphic**, and write $M \cong N$.

Definition A.7 (Kernel)**.** The **kernel** of a homomorphism $f : M \to N$ is the preimage of 1_N: $\ker(f) := f^{-1}(1_N)$. For any homomorphism f, $\ker(f)$ is a submonoid of M.[10]

[10] Similarly, the image of a homomorphism is also a submonoid.

Example A.8. The inclusions $(\mathbb{N}, +) \to (\mathbb{Z}, +) \to (\mathbb{Q}, +) \to (\mathbb{R}, +)$ are all monoid homomorphisms with trivial kernel.[11] This implies this is also a chain of inclusions as submonoids.

[11] i.e. the kernel only contains the identity.

Definition A.9 (Monoid action)**.** Let M be a monoid and S a set, an (left) **action** of M on S is an operation $\star : M \times S \to S$ satisfying for all $x, y \in M$ and $s \in S$

$$(x \cdot y) \star s = x \star (y \star s) \quad \text{and} \quad 1_M \star s = s.$$

The data $(M, S, \star)$ will also be called an M–**set** and we may refer to it abusively with S.

Any monoid action has a **permutation representation** defined to be the map

$$\sigma_\star : M \to S^S = x \mapsto (s \mapsto x \star s).$$

The properties of the action imply $\sigma_\star$ is a homomorphism. Conversely, given a homomorphism $\sigma : M \to S^S$ (i.e. $\sigma(1_M)$ is the identity function and $\sigma(xy) = \sigma(x) \circ \sigma(y)$ for any $x, y \in M$), there is a monoid action $\star_\sigma$ defined by $x \star_\sigma s = \sigma(x)(s)$.[12]

[12] These are inverse operations, i.e.

$$\sigma_{\star_\sigma} = \sigma \quad \text{and} \quad \star_{\sigma_\star} = \star.$$

Example A.10. Any monoid M has a canonical left action on itself defined by $x \star m = xm$ for all $x, m \in M$.

Groups

Definition A.11 (Group)**.** A **group** is set G equipped with a binary operation $\cdot : G \times G \to G$ called **multiplication**, an **inverse** operation $(-)^{-1} : G \to G$ and an **identity** element 1_G such that $(G, \cdot, 1_G)$ is a monoid and for all $x \in G$

$$x \cdot x^{-1} = 1_G = x^{-1} \cdot x.$$

If $(G, \cdot, 1_G)$ is a commutative monoid, we say that G is an **abelian group**.

Examples A.12. 1. For any set S, we saw Σ_S was a submonoid of S^S, and it is in fact a group where the inverse of a function f is f^{-1} (it exists because f is bijective). We denote this group $\boldsymbol{\Sigma}_S$ and call it the group of **permutations** of S.[13]

[13] For $n \in \mathbb{N}$, Σ_n denotes the group of permutations of $\{1, \ldots, n\}$.

2. The monoids on $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$ and $(\mathbb{R}, +)$ are also abelian groups with the inverse of x being $-x$.

3.

Definition A.13 (Subgroup)**.** Given a group G, a **subgroup** of G is a submonoid H of G closed under taking inverses (i.e. $\forall x \in H, x^{-1} \in H$).[14]

[14] This implies H is also a group with the multiplication, inverse and identity inherited from G.

Example A.14. For any group G and subset $S \subseteq G$, the subgroup **generated** by S inside G, denoted by $\langle S \rangle$ is the smallest subgroup containing S.[15]

[15] An explicit construction is

$$\langle S \rangle = \{x_1 \cdots x_n \mid n \in \mathbb{N}, x_1, \ldots, x_n \in S \cup \{1_G\}\}.$$

Definition A.15 (Homomorphism)**.** Let G and H be two groups, a **group homomorphism** from G to H is a monoid homomorphism $f : G \to H$. It follows that[16]

[16] For this, you need to show that inverses are unique.

$$\forall x \in G, f(x^{-1}) = f(x)^{-1}.$$

When f is a bijection, we call it a **group isomorphism**, say that G and H are **isomorphic**, and write $G \cong H$.

Example A.16. For any group G and element $g \in G$, we call **conjugation** by g the homomorphism $\mathsf{c}_g : G \to G$ defined by $\mathsf{c}_g(x) = gxg^{-1}$.[17]

[17] It is a homomorphism as $g1_Gg^{-1} = gg^{-1} = 1_G$ and

$$gxyg^{-1} = gx1_Gyg^{-1} = gxg^{-1}gyg^{-1}.$$

Definition A.17 (Kernel)**.** The **kernel** of a homomorphism $f : G \to H$ is the preimage of 1_H: $\ker(f) := f^{-1}(1_H)$. For any homomorphism f, $\ker(f)$ is a subgroup of G.[18]

[18] Similarly, the image of a homomorphism is also a subgroup.

Example A.18. For any group G and element $g \in G$, $\ker(\mathsf{c}_g) = \{1_G\}$. Indeed, if $gxg^{-1} = 1_G$, conjugating by g^{-1} on both sides yields $x = 1_G$.

Definition A.19 (Normal subgroup)**.** A subgroup N of G is called **normal** if for any $g \in G$ and $n \in N$, $gng^{-1} \in N$. In words, N is closed under conjugation by G. We write $N \lhd G$ when N is a normal subgroup of G.[19]

[19] The kernel of any homomorphism f is a normal subgroup as for any $h \in \ker f$ and any $g \in G$, we have

$$f(ghg^{-1}) = f(g)f(h)f(g)^{-1} = f(g)1f(g^{-1}) = 1.$$

Proposition A.20. *For any subgroup H of G, the relation $\sim_H$ defined by*

$$g \sim_H g' \Leftrightarrow \exists h \in H, gh = g'$$

is an equivalence relation.

Proof. Any subgroup contains 1_G, so $g \sim_H g$ is witnessed by $g1_G = g$, hence $\sim_H$ is reflexive. If $gh = g'$, then $g = ghh^{-1} = g'h^{-1}$, thus $\sim_H$ is symmetric. If $gh = g'$ and $g'h' = g''$, then $ghh' = g''$ and since H is a subgroup $hh' \in H$, we conclude $\sim_H$ is transitive. □

Definition A.21 (Quotient)**.** Let G be a group and N a normal subgroup of G, the multiplication of G is well-defined on equivalence classes of $\sim_N$, namely, if $g \sim_N g'$ and $h \sim_N h'$, then $gh \sim_N g'h'$.[20] The **quotient** G/N is the group whose elements are equivalence classes of $\sim_N$ with the multiplication $[g] \cdot [h] := [g \cdot h]$ and identity $1_{G/N} = [1_G]$ (where $[g]$ denotes the equivalence class of $\sim_N$ containing g).

[20] Suppose $gn = g'$ and $hn' = h'$ for $n, n' \in N$, then using the fact that $h^{-1}nh \in N$, we let $n'' := h^{-1}nhn' \in N$ and we find

$$g'h' = gnhn' = ghh^{-1}nhn' = ghn'',$$

thus $gh \sim_N g'h'$.

Definition A.22 (Group action)**.** Let G be a group and S a set, an (left) **action** of G on S is a (left) monoid action of G on S. A set S equipped with action of G is called a **G–set**. It follows from the properties of an action that the function $s \mapsto g \star s$ is a bijection, hence the permutation representation $\sigma_\star$ is a homomorphism $G \to \Sigma_S$.

Example A.23. Any group G has a canonical left action on itself defined by $x \star m = xm$ for all $x, m \in G$.

Definition A.24 (Orbit)**.** Let S be a G–set, an **orbit** of S is a maximal subset of S closed under the action of G. Namely, it is a subset $A \subset S$ such that $g \star a \in A$ for any $g \in G$ and $a \in A$, and no subset strictly including A and strictly included in S ($A \subset A' \subset S$) has this property.

Rings

Definition A.25 (Ring)**.** A **ring** is a set R equipped with a monoid structure $(R, \cdot, 1_R)$ and an abelian group structure $(R, +, 0_R)$[21] such that for all $x, y, z \in R$

$$x \cdot (y + z) = (x \cdot y) + (x \cdot z).$$

If $(R, \cdot, 1_R)$ is a commutative monoid, we say that R is commutative.

[21] We call $\cdot$ the **multiplication** and $+$ the **addition** of the ring.

Examples A.26. 1. The abelian groups $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$ and $(\mathbb{R}, +)$ are also commutative rings with multiplication being the standard multiplication of numbers.

2. For any ring R and any $n \in \mathbb{N}$, the set of matrices $R^{n\times n}$ is a ring where addition is done pointwise, multiplication is the standard multiplication of matrices, $1_{R^{n\times n}}$ is the matrix with 1_R in each diagonal entry and 0_R everywhere else, and $0_{R^{n\times n}}$ is the matrix with 0_R everywhere.

Proposition A.27. *Let R be a ring, for any $r \in R$, $0_R \cdot r = 0_R = r \cdot 0_R$.*

Proof. Here is the derivation for one equality (the other is symmetric):

$$0_R \cdot r = (1_R - 1_R) \cdot r = 1_R \cdot r - 1_R \cdot r = r - r = 0_R.$$

□

Definition A.28 (Subring)**.** Given a ring R, a **subring** of R is a subset $S \subseteq R$ that is both a submonoid for $\cdot$ and a subgroup for $+$.[22]

[22] This implies S is also a ring with the multiplication and addition inherited from R.

Definition A.29 (Homomorphism)**.** Let R and S be two rings, a **ring homomorphism** from R to S is a function $f : R \to S$ that is both a monoid homomorphism for the operation $\cdot$ and a group homomorphism for the operation $+$. Namely, it satisfies

$$\begin{aligned} &\forall x, y \in R, f(x \cdot y) = f(x) \cdot f(y) && f(1_R) = 1_S \\ &\forall x, y \in R, f(x + y) = f(x) + f(y) && f(0_R) = 0_S. \end{aligned}$$

When f is a bijection, we call it a **ring isomorphism**, say that R and S are **isomorphic**, and write $R \cong S$.

Definition A.30 (Kernel)**.** The **kernel** of a homomorphism $f : R \to S$ is the preimage of 0_S: $\ker f := f^{-1}(0_S)$. For any homomorphism, $\ker f$ is a subring of S.

As for monoids and groups, the image of a homomorphism is a subring, and as for groups the kernel satisfies an additional property: it is an ideal.

Definition A.31 (Ideal)**.** Given a ring R, an **ideal** of R is a subring I such that for any $i \in I$ and $r, s \in R$, $ris \in I$.[23]

[23] An ideal is not only closed under multiplication but it is also preserved by multiplication by elements outside of the ideal.

Proposition A.32. *For any subring S of R, the relation $\sim_S$ defined by*

$$r \sim_S r' \Leftrightarrow \exists s \in S, r + s = r'$$

is an equivalence relation.[24]

[24] Apply Proposition A.20 to the group $(R, +)$ and its subgroup $(S, +)$.

Definition A.33 (Quotient)**.** Let R be a ring and I be an ideal of R, the addition and multiplication of R are well-defined on equivalence classes of $\sim_I$, namely, if $r \sim_I r'$ and $s \sim_I s'$, then $r + s \sim_I r' + s'$ and $rs \sim_I r's'$.[25] The quotient R/I is the ring whos elements are equivalence classes of $\sim_I$ with the addition $[r] + [s] := [r + s]$, the multiplication $[r] \cdot [s] := [r \cdot s]$, $0_{R/I} := [0_R]$, and $1_{R/I} := [1_R]$.

[25] For addition, we can use the same proof as for quotient groups because I is a normal subgroup of $(R, +)$ (any subgroup of an abelian group is normal). For multiplication, suppose $r + i = r'$ and $s + j = s'$ for $i, j \in I$, then

$$r's' = (r + i)(s + j) = rs + rj + is + ij,$$

and since I is an ideal, $rj + is + ij \in I$. We conclude $rs \sim_I r's'$.

Definition A.34 (Units)**.** An element of a ring is called a **unit** if it has a multiplicative inverse. Namely, $x \in R$ is a unit if there exists x^{-1} such that $xx^{-1} = 1_R = x^{-1}x$. We denote by $R^\times$ the set of units of R, it is a group with the multiplication inherited from R.

Example A.35. The group of unit of $R^{n \times n}$ is called the **general linear group** over R and denoted by $\mathrm{GL}_n(R)$. It contains all the invertible[26] $n \times n$ matrices with entries in R.

[26] Sometimes called non-singular.

Proposition A.36. *Any ring homomorphism $f : R \to S$ sends units of R to units of S.*[27]

[27] By restricting f to $R^\times$, we obtain a group homomorphism
$$f^\times : R^\times \to S^\times.$$

Proof. If $x \in R$ has a multiplicative inverse x^{-1}, then the homomorphism properties imply

$$f(x)f(x^{-1}) = f(xx^{-1}) = f(1_R) = 1_S = f(1_R) = f(x^{-1}x) = f(x^{-1})f(x),$$

thus $f(x^{-1})$ is the multiplicative inverse of $f(x)$. □

Fields

Definition A.37 (Field)**.** A **field** is a commutative ring where every non-zero element is a unit.

Example A.38. The rings $\mathbb{Q}$ and $\mathbb{R}$ are fields, but $\mathbb{Z}$ is not since the $\mathbb{Z}^\times = \{-1, 1\}$.

Definition A.39 (Characteristic)**.** The *characteristic* of a field k is the minimum $n \in \mathbb{N}$ such that $1_k + \overset{n}{\cdots} + 1_k = 0_K$. If no such n exists, the characteristic of k is infinite.[28]

[28] One can show the characteristic of a field is never a composite number, it is either prime or infinite.

Examples A.40. Fix a prime number p. The set $p\mathbb{Z}$ of multiples of p is an ideal of the ring $\mathbb{Z}$ and $\mathbb{Z}/p\mathbb{Z}$ is a field of characteristic p. The field $\mathbb{Q}$ has infinite characteristic.

Vector Spaces

Fix a field k.

Definition A.41 (Vector space)**.** A **vector space** over k is a set an abelian group $(V, +, 0)$ along with an operation $\cdot : k \times V \to V$ called **scalar multiplication** such that the following holds for any $x, y \in k$ and $u, v \in V$:[29]

$$(xy) \cdot v = x \cdot (y \cdot v) \qquad 1 \cdot v = v$$
$$(x + y) \cdot v = x \cdot v + y \cdot v \qquad x \cdot (u + v) = x \cdot u + x \cdot v.$$

It follows that $0 \cdot v = 0$. We call elements of V **vectors**.

[29] We will not distinguish between the additions and zeros in k and V.

Example A.42. For any $n \in \mathbb{N}$, the set k^n has a vector space structure, where addition and scalar multiplication are done pointwise, i.e.:

$$(u_1, \ldots, u_n) + (v_1, \ldots, v_n) = (u_1 + v_1, \ldots, u_n + v_n) \quad x \cdot (v_1, \ldots, v_n) = (xv_1, \ldots, xv_n).$$

Definition A.43 (Subspace)**.** Given a vector space V, a **subspace** of V is a subset $W \subseteq V$ such that $0 \in W$, and for any $x \in k$ and $u, w \in W$, $x \cdot w \in W$ and $u + w \in W$.

Definition A.44 (Linear map)**.** Let V and W be two vector spaces over k, a **linear map** from V to W is a function $T : V \to W$ satisfying

$$\forall x \in k, \forall u, v \in V, \quad T(x \cdot v) = x \cdot T(v) \qquad T(u + v) = T(u) + T(v).$$

When T is a bijection, we call it a **linear isomorphism**, say that R and S are **isomorphic**, and write $V \cong W$.

Definition A.45 (Linear combination)**.** Let V be a vector space and $v_1, \ldots, v_n \in V$, a **linear combination** of these vectors is a sum

$$\sum_{i=1}^{n} a_i v_i = a_1 \cdot v_1 + \cdots + a_n v_n,$$

where $a_1, \ldots, a_n \in k$ are called the **coefficients**.

Definition A.46 (Basis)**.** Let V be a vector space and $S \subseteq V$. We say that S is **linearly independent** if a linear combination of vectors in S is the zero vector if and only if all coefficients are zero. We say that S is **generating** if any $v \in V$ is a linear combination of vectors in S. We say that S is a **basis** of V if it is linearly independent and generating. The cardinality of a basis S of V is called the **dimension** of V.[30]

[30] Using the axiom of choice, one can show a basis always exists and all bases must have the same cardinality, hence the dimension of a vector space is well-defined.

Proposition A.47. *A linear map $T : V \to W$ is completely determined by where it sends a basis of V.*

Proposition A.48. *If a vector space V over k has dimension $n \in \mathbb{N}$, then $V \cong k^n$.*

Definition A.49 (Dual)**.**

A.2 Order Theory

In this section, we briefly cover some early definitions and results from order theory. Since this subject is not usually taught in undergraduate courses, we spend a bit more time. In fact, we even introduce stuff we will not use later to make sure readers can get more familiar with the most important objects: posets and monotone functions.

Definition A.50 (Poset)**.** A **poset** (short for partially ordered set) is a pair $(A, \leq)$ comprising a set A and a binary relation $\leq \subseteq A \times A$ that is

1. **reflexive** ($\forall x \in A, x \leq x$),
2. **transitive** ($\forall x, y, z \in A$ if $x \leq y$ and $y \leq z$ then $x \leq z$), and
3. **antisymmetric** ($\forall x, y \in A$ if $x \leq y$ and $y \leq x$ the $x = y$).

The relation is also called a **partial order**.[31]

[31] If antisymmetry is not satisfied, $\leq$ is called a **preorder**.

Examples A.51. 1. The usual non-strict orders ($\leq$ and $\geq$) on $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ are all partial orders. The strict orders do not satisfy reflexivity.

2. The divisibility relation $\mid$ on $\mathbb{N}$ ($n \mid m$ if and only if n divides m) is a partial order.
3. For any set S, the powerset of S equipped with the subset relation ($\subseteq$) is a poset.
4. Any subset of a poset inherits a poset structure by restricting the partial order.

For any monoid M, there are three preorders defined by the so-called Green's relations:

$$\forall x, y \in M, x \leq_L y \Leftrightarrow \exists m \in M, x = my$$
$$\forall x, y \in M, x \leq_R y \Leftrightarrow \exists m \in M, x = ym$$
$$\forall x, y \in M, x \leq_J y \Leftrightarrow \exists m, m' \in M, x = mym'$$

Definition A.52 (Monotone)**.** A function $f : (A, \leq_A) \to (B, \leq_B)$ between posets is **monotone** (or **order-preserving**) if for any $a, a' \in A$, $a \leq_A a' \implies f(a) \leq_B f(a')$.

Example A.53. You probably already know lots of monotone functions, but let us give two less intuitive examples. Let $f : S \to T$ be a function, the **image map** of f[32] is the function $\mathcal{P}(S) \to \mathcal{P}(T)$ defined by $S \supseteq X \mapsto f(X) := \{f(x) \mid x \in X\}$. When both powersets are equipped with the inclusion partial order, the image map is monotone because $X \subseteq X' \subseteq S$ implies $f(X) \subseteq f(X')$.

[32] Which we abusively denote by f.

The **preimage map** is

$$f^{-1} : \mathcal{P}(T) \to \mathcal{P}(S) = T \supseteq Y \mapsto f^{-1}(Y) := \{y \in S \mid f(y) \in Y\}.$$

It is also order-preserving because $Y \subseteq Y' \subseteq T$ implies $f^{-1}(Y) \subseteq f^{-1}(Y')$.

Proposition A.54. *The composition of monotone functions between posets is monotone.*

Definition A.55 (Dual)**.** The **dual order**[33] of a poset $(A, \leq)$, denoted by $(A, \leq)^{\mathrm{op}}$, is the same set equipped with the converse relation $\geq$ defined by

$$\forall x, y \in A, x \geq y \Leftrightarrow y \leq x.$$

[33] This definition lets us avoid many symmetric arguments.

Definition A.56 (Bounds)**.** Let $(A, \leq)$ be a poset and $S \subseteq A$, then $a \in A$ is an **upper bound** of S if $\forall s \in S, s \leq a$. Moreover, $a \in A$ is a **supremum** of S, if it is a least upper bound, that is, a is an upper bound of S and for any upper bound a' of S, $a \leq a'$. A supremum of S is denoted by $\vee S$, but when S contains only two elements, we use the infix notation $s_1 \vee s_2$ and call this a **join**.

A **lower bound** (resp. **infimum/meet**) of S is an upper bound (resp. supremum/join) of S in the dual order $(A, \leq)^{\mathrm{op}}$.[34] An infimum of S is denoted by $\wedge S$ or $s_1 \wedge s_2$ in the binary case.

[34] Explicitly, $a \in A$ is a lower bound of S if $\forall s \in S, a \leq s$. It is an infimum of S if, in addition to being a lower bound of S, any lower bound a' of S satisfies $a' \leq a$.

Proposition A.57. *Infimums and supremums are unique when they exist.*[35]

[35] This holds by antisymmetry.

Definition A.58 (Complete lattice)**.** A **complete lattice** is a poset $(L, \leq)$ where every subset has a supremum and an infimum.[36] In particular, L has a smallest element $\vee\emptyset$ and a largest element $\wedge\emptyset$ (they are usually called **top** and **bottom** respectively).

[36] Notice that, we can see $\vee$ and $\wedge$ as monotone maps from $(\mathcal{P}(L), \subseteq)$ to $(L, \leq)$.

Examples A.59. 1. For any set S, $(\mathcal{P}(S), \subseteq)$ is a complete lattice. the supremum of a family of subsets is their union and the infimum is their intersection.

2. Defining supremums and infimums on the poset $(\mathbb{N}, |)$ is subtle. When $S \subseteq \mathbb{N}$ is non-empty, $\wedge S$ is the greatest common divisor of all elements in S and $\wedge\emptyset$ is 0 because any integer divides 0. For a finite and non-empty $S \subseteq \mathbb{N}$, $\vee S$ is the least common multiple of all elements in S. If S is infinite, then $\vee S$ is 0 and the supremum of the empty set is 1 because 1 divides any integer.

You might be wondering about possible posets where all infimums exist but not necessarily all supremums or vice-versa, it turns out that this is not possible as shown below.

Proposition A.60. *Let $(L, \leq)$ be a poset, then the following are equivalent:*

(i) *$(L, \leq)$ is a complete lattice.*

(ii) *Any $S \subseteq L$ has a supremum.*

(iii) *Any $S \subseteq L$ has an infimum.*

Proof. (i) $\implies$ (ii), (i) $\implies$ (iii) and (ii) + (iii) $\implies$ (i) are all trivial. Also, by using duality, we only need to prove (ii) $\implies$ (iii).[37] For that, it suffices to note that, for any $S \subseteq L$, we can define $\wedge S$ to be the least upper bound for lower bounds of S. Formally,

$$\wedge S = \bigvee \{a \in L \mid \forall s \in S, a \leq s\}.$$

Defined that way, $\wedge S$ is a lower bound of S because if $s \in S$, then $s \geq a$ for every lower bound a of S, thus $\wedge S$, being the least upper bound of the lower bounds, is smaller than s. By definition, $\wedge S$ is greater than any other lower bound of S, hence it is indeed the infimum of S. $\square$

[37] If this implication is true for any $(L, \leq)$, then it is true, in particular, for $(L, \geq)$. This implication for $(L, \geq)$ is equivalent to the converse implication for $(L, \leq)$.

Definition A.61 (Fixpoints)**.** Let $f : (L, \leq) \to (L, \leq)$, a **pre-fixpoint** of L is an element $x \in L$ such that $f(x) \leq x$. A **post-fixpoint** is an element $x \in L$ such that $x \leq f(x)$. A **fixpoint** (or **fixed point**) of f is a pre- and post-fixpoint.

Theorem A.62 (Knaester–Tarski[38])**.** *Let $(L, \leq)$ be a complete lattice and $f : L \to L$ be monotone.*

[38] This is actually a weaker version of the Knaester-Tarski theorem. The latter states that the fixpoints of a monotone function f form a complete lattice.

1. *The least fixpoint of f is the least pre-fixpoint $\mu f := \wedge\{a \in L \mid f(a) \leq a\}$.*

2. *The greatest fixpoint of $f$is the greatest post-fixpoint $\nu f := \vee\{a \in L \mid a \leq f(a)\}$.*

Proof. 1. Any fixpoint of f is in particular a pre-fixpoint, thus μf, being a lower bound of all pre-fixpoints, is smaller than all fixpoints. Moreover, because for any pre-fixpoint $a \in L$, $f(\mu f) \leq f(a) \leq a$, $f(\mu f)$ is also a lower bound of the pre-fixpoints, so $f(\mu f) \leq \mu f$. We infer that $f(f(\mu f)) \leq f(\mu f)$, so $f(\mu f)$ is a pre-fixpoint and $\mu f \leq f(\mu f)$. We conclude that μf is a fixpoint by antisymmetry.

The proof of the second item is the proof of the first item done in the dual order.

2. Any fixpoint of f is in particular a post-fixpoint, thus νf, being an upper bound of post-fixpoints, is bigger than all fixpoints. Moreover, because for any post-fixpoint $a \in L$, $a \leq f(a) \leq f(\nu f)$, $f(\nu f)$ is an upper bound of the post-fixpoints, so $\nu f \leq f(\nu f)$. We infer that $f(\nu f) \leq f(f(\nu f))$, so $f(\nu f)$ is a post-fixpoint and $f(\nu f) \leq \nu f$. We conclude that νf is a fixpoint by antisymmetry.

$\square$

Definition A.63 (Closure operator)**.** Let $(A, \leq)$ be a poset, a **closure operator** on A is a map $c : A \to A$ that is

1. monotone,
2. extensive ($\forall a \in A, a \leq c(a)$), and
3. idempotent ($\forall a \in A, c(a) = c(c(a))$).

Example A.64. The floor ($\lfloor - \rfloor$) and ceiling ($\lceil - \rceil$) operations are closure operators on $(\mathbb{R}, \geq)$ and $(\mathbb{R}, \leq)$ respectively.

Definition A.65 (Galois connection)**.** Given two posets $(A, \leq)$ and $(B, \sqsubseteq)$, a **Galois connection** is a pair of monotone functions $l : A \to B$ and $r : B \to A$ such that for any $a \in A$ and $b \in B$,

$$l(a) \sqsubseteq b \Leftrightarrow a \leq r(b).$$

For such a pair, we write $l \dashv r : A \to B$.

Proposition A.66. *Let $l \dashv r : A \to B$ be a Galois connection, then l and r are monotone.*

Proof. Suppose $a \leq a'$, we will show $l(a) \sqsubseteq l(a')$. Since $l(a') \sqsubseteq l(a')$, using $\Rightarrow$ of the Galois connection yields $a' \leq r(l(a'))$, and, by transitivity, we have $a \leq r(l(a'))$. Then, using $\Leftarrow$ of the Galois connection, we find $l(a) \sqsubseteq l(a')$. We conclude that l is monotone.

A symmetric argument works to show r is monotone. $\square$

Example A.67.

Proposition A.68. *Let $l \dashv r : A \to B$ be a Galois connection, then $r \circ l : A \to A$ is a closure operator.*

Proof. Since r and l are monotone, $r \circ l$ is monotone. Also, for any $a \in A$, $l(a) \sqsubseteq l(a)$ implies $a \leq r(l(a))$, so $r \circ l$ is extensive.

Now, in order to prove $r \circ l$ is idempotent, it is enough to show that[39]

$$r(l(a)) \geq r(l(r(l(a)))).$$

Observe that since $r(b) \leq r(b)$ for any $b \in B$, we have $l(r(b)) \leq b$, thus in particular, with $b = l(a)$, we have $l(r(l(a))) \leq l(a)$. Applying r which is monotone yields the desired inequality. $\square$

[39] The $\leq$ inequality follows by extensiveness.

Proposition A.69. *Let $l \dashv r : A \to B$ and $l' \dashv r : A \to B$ be Galois connections, then $l = l'$.*

Proposition A.70. *Let $l \dashv r : A \to B$ and $l \dashv r' : A \to B$ be Galois connections, then $r = r'$.*

A.3 Topology

In this section, we introduce the basic terminology of topological spaces. Again we go a bit further than needed to help readers that first learn about topology here. We end this section by recalling some definitions about metric spaces.

Definition A.71. A **topological space** is a pair (X, τ), where X is a set and $\tau \subseteq \mathcal{P}(X)$ is a family of subsets of X closed under arbitrary unions and finite intersections[40] whose elements are called **open sets** of X. We call τ a **topology** on X.

The **complement** of an open set U, denoted by U^c, is said to be **closed**.[41]

Example A.72. On any set X, there are two trivial and extreme topologies.[42] The **discrete topology** $\tau_\top := \mathcal{P}X$ contains all the subsets of X. We can view $(X, \tau_\top)$ as a space where all points of X are separated from each other. The **codiscrete topology** $\tau_\bot := \{\varnothing, X\}$ contains only the subsets that must be open by definition of a topology. We can view $(X, \tau_\bot)$ as a space where all points of X are glued together with no space in-between.

In the sequel, fix a topological space (X, τ).

Proposition A.73. *Let $(C_i)_{i\in I}$ be a family of closed sets of X, then $\cap_{i\in I} C_i$ is closed and if I is finite, $\cup_{i\in I} C_i$ is also closed.*[43]

Proof. Both statements readily follow from DeMorgan's laws and the fact that the complement of a closed set is open and vice-versa. For the first one, DeMorgan's laws yield

$$\bigcap_{i\in I} C_i = \left(\bigcup_{i\in I} C_i^c\right)^c,$$

and the LHS is the complement of a union of opens, so it is closed. For the second one, DeMorgan's laws yield

$$\bigcup_{i\in I} C_i = \left(\bigcap_{i\in I} C_i^c\right)^c,$$

and the LHS is the complement of a finite intersection of opens, so it is closed. □

Proposition A.74. *A subset $A \subseteq X$ is open if and only if for any $x \in A$, there exists an open $U \subseteq A$ such that $x \in U$.*

Proof. ($\Rightarrow$) For any $x \in A$, set $U = A$.

($\Leftarrow$) For each $x \in X$, pick an open $U_x \subseteq A$ such that $x \in A$, then we claim $A = \cup_{x\in A} U_x$ which is open[44]. The $\subseteq$ inclusion follows because each $x \in A$ has a set U_x in the union that contains x. The $\supseteq$ inclusion follows because each term of the union is a subset of A by assumption. □

Proposition A.75. *A subset $A \subseteq X$ is closed if and only if for any $x \notin A$, there exists an open U such that, $x \in U$ and $U \cap A = \varnothing$.*[45]

[40] For any family of open sets $\{U_i\}_{i\in I} \subseteq \tau$,
$$\bigcup_{i\in I} U_i \in \tau,$$
and if I is finite,
$$\bigcap_{i\in I} U_i \in \tau.$$

[41] Observe that both the empty set and the whole space are open and closed (sometimes referred to as **clopen**) because
$$\varnothing = \bigcup_{U\in\varnothing} U \text{ and } X = \bigcap_{U\in\varnothing} U \text{ and } \varnothing = X^c.$$

[42] Trivial because

[43] This lemma gives an alternative to the axioms of Definition A.71. Indeed, it is sometimes more convenient to define a topological space by giving its closed sets, and you can show the axioms about open sets still hold.

[44] Arbitrary unions of opens are open.

[45] This result is simply a restatement of the last one by setting $A = A^c$.

Definition A.76. Given $A \subseteq X$, the **closure** of A, denoted by $\overline{A}$ is the intersection of all closed sets containing A. One can show that $\overline{A}$ is the smallest closed set containing A.[46] Then, it follows that A is closed if and only if $\overline{A} = A$.

[46] $\overline{A}$ is closed because it is an intersection of closed sets and any closed sets containing A also contains $\overline{A}$ by definition.

Here are more easy results on the closure of a subset.

Proposition A.77. *Given $A, B \subseteq X$ then the following statements hold:*

1. $A \subseteq B \implies \overline{A} \subseteq \overline{B}$
2. $A \subseteq \overline{A}$
3. $\overline{\overline{A}} = \overline{A}$
4. $\overline{\emptyset} = \emptyset$
5. $\overline{(A \cup B)} = \overline{A} \cup \overline{B}$

Proof of Lemma A.77. 1. By definition, $\overline{B}$ contains B, thus A, but $\overline{B}$ is closed, so it must contain $\overline{A}$.

2. By definition.
3. $\overline{A}$ is closed, so its closure is itself.
4. 3 applied to $\emptyset$.
5. $\subseteq$ follows because the LHS is the smallest closed set containing $A \cup B$ and the RHS is closed and contains $A \cup B$.
   $\supseteq$: Since the RHS is closed, we have $\overline{(\overline{A} \cup \overline{B})} = \overline{A} \cup \overline{B}$ implying that the RHS is the smallest closed set containing $\overline{A} \cup \overline{B}$. Then, since the LHS is a closed set containing A and B, it contains $\overline{A}$ and $\overline{B}$ and hence must contain the RHS.

□

Remark A.78. If we view $\mathcal{P}(X)$ as partial order equipped with the inclusion relation, the previous lemma is about good properties of the function $\overline{(-)} : \mathcal{P}(X) \to \mathcal{P}(X)$. Namely, we showed in the first three points that it is a monotone, extensive and idempotent, and therefore it is a closure operator.[47]

[47] In fact, this is where the terminology comes from.

Definition A.79 (Dense). A subset $A \subseteq X$ is said to be **dense** (in X) if any non-empty open set intersects A non-trivially, that is, $\forall \emptyset \neq U \in \tau, A \cap U \neq \emptyset$.

Proposition A.80 (Decomposition). *Let $A \subseteq X$, then $A = \overline{A} \cap (A \cup (\overline{A})^c)$, where $\overline{A}$ is closed and $A \cup (\overline{A})^c$ is dense. This results says that any subset of X can be decomposed into a closed and a dense set.*

Proof. The equality is clear[48] and $\overline{A}$ is closed by definition. It is left to show that $A \cup (\overline{A})^c$ is dense. Let $U \neq \emptyset$ be an open set. If U intersects A, we are done. Otherwise, we have the following equivalences:

$$U \cap A = \emptyset \Leftrightarrow A \subseteq U^c \Leftrightarrow \overline{A} \subseteq U^c \Leftrightarrow U \subseteq (\overline{A})^c,$$

where the second $\Rightarrow$ holds because U^c is closed. We conclude $U \cap (\overline{A})^c \neq \emptyset$. □

[48] We use (in this order) distributivity of $\cap$ over $\cup$, the fact that a set and its complement intersect trivially and the inclusion $A \subseteq \overline{A}$:

$$\begin{aligned} \overline{A} \cap (A \cup (\overline{A})^c) &= (\overline{A} \cap A) \cup (\overline{A} \cap (\overline{A})^c) \\ &= A \cup \emptyset \\ &= A \end{aligned}$$

Proposition A.81. *A subset $A \subseteq X$ is dense if and only if $\overline{A} = X$.*

Proof. ($\Rightarrow$) Since $(\overline{A})^c$ is open but it intersects trivially the dense set A, it must be empty, thus $\overline{A}$ is the whole space.

($\Leftarrow$) Let U be an open set such that $U \cap A = \emptyset$, then A is contained in the closed set U^c, but this implies $\overline{A} \subseteq U^c$,[49] thus U is empty. □

[49] Recall that the closure of A is the smallest closed set containing A.

Definition A.82 (Interior). Let $A \subseteq X$, the **interior** of A, denoted by A° is the union of all open sets contained in A. Similarly to the closure, we can check that that A° is the largest open subset of A and thus that A is open if and only if $A = A^\circ$.[50]

[50] It also follows that $A \subseteq B \implies A^\circ \subseteq B^\circ$ and that $A^{\circ\circ} = A^\circ$.

We end this section by presenting a largely preferred way of defining a topology that avoid describing all open sets.

Definition A.83 (Base)**.** Let X be a set, a **base** B is a set $B \subseteq \mathcal{P}(X)$ such that $X = \cup_{U \in B} U$ and any finite intersection of sets in B can be written as a union of sets in B.

Proposition A.84. *Let X and $B \subseteq \mathcal{P}(X)$. If τ is the set of all unions of sets in B, then it is a topology on X. We say that τ is the topology **generated** by B.*

Proof. By assumption, we know that unions of opens are open and finite intersections of sets in B are open. It remains to show that finite intersections of unions of sets in B are also open. Let $U = \cup_{i \in I} U_i$ and $V = \cup_{j \in J} V_j$ with $U_i \in B$ and $V_j \in B$, then by distributivity, we obtain

$$U \cap V = \cup_{i \in I} U_i \bigcap \cup_{j \in J} V_j = \bigcup_{i \in I, j \in J} U_i \cap V_j,$$

so $U \cap V$ is open.[51] The lemma then follows by induction. □

[51] It is a union of opens.

In practice, instead of generating a topology from a base B, we start with any family $B_0 \subseteq \mathcal{P}(X)$ and let B be its closure under finite intersections, which satisfies the axioms of a base. Such a B_0 is often called a **subbase** for the topology generated by B.

Another very useful way to define topological spaces is to consider the topology induced by a metric.

Definition A.85 (Metrics space)**.** A **metric space** (X, d) is a set X together with a function $d : X \times X \to \mathbb{R}$ called a **metric** with the following properties for $x, y, z \in X$:

1. $d(x, y) \geq 0$
2. $d(x, y) = 0 \Leftrightarrow x = y$
3. $d(x, y) = d(y, x)$
4. $d(x, y) \leq d(x, z) + d(z, y)$

Definition A.86 (Non-expansive)**.** A function between metric spaces $f : (X, d_X) \to (Y, d_Y)$ is said to be **non-expansive**[52] if for all $x, x' \in X$,

$$d_Y(f(x), f(x')) \leq d_X(x, x').$$

[52] Also called 1–**Lipschitz** or **short**.

Proposition A.87. *The composition of any two non-expansive maps is non-expansive.*

Definition A.88 (Open ball)**.** Let (X, d) be a metric space. Given a point $x \in X$ and a non-negative radius $r \in [0, \infty)$, the **open ball** of radius r centered at x is

$$B_r(x) := \{y \in X \mid d(x, y) < r.$$

Definition A.89 (Induced topology)**.** Any metric space (X, d) has an *induced topology* generated by the set of all open balls of X.[53]

In this topology, a set $S \subseteq X$ is open if and only if every point $x \in S$ is contained in an open ball which is contained in S.[54]

[53] This topology is sometimes called the open ball topology.

[54] Equivalently, $\forall x \in S, \exists r > 0, B_r(x) \subseteq S$.

Definition A.90 (Convergence)**.** Let (X, d) be a metric space, a sequence $\{p_n\}_{n\in\mathbb{N}} \subseteq X$ **converges** to $p \in X$ if

$$\forall \varepsilon > 0, \exists N \in \mathbb{N}, \forall n \geq N, d(p_n, p) < \varepsilon.$$

Definition A.91 (Cauchy sequence)**.** Let (X, d) be a metric space, a sequence $\{p_n\}_{n\in\mathbb{N}} \subseteq X$ is called **Cauchy** if

$$\forall \varepsilon > 0, \exists N \in \mathbb{N}, \forall m, n \geq N \implies d(p_n, p_m) < \varepsilon.$$

Definition A.92 (Completeness)**.** A metric space in which every Cauchy sequence converges is called **complete**.

B Categories and Functors

As you will soon realize, many common mathematical objects can be viewed as categories or parts of a category, and often in several ways. Hence, there can be many starting points to motivate category theory even after restricting ourselves to the background of an undergraduate student in mathematics (see Chapter A). I do not want to spend much time in the realm of informal explanations, so we will start from the notion of directed graphs, quickly get to the definition of a category and begin an enumeration of examples which will carry on (implicitly) for the rest of the book. We will also define functors which are to categories what homomorphisms are to groups (or rings, etc.), and list a bunch of examples.

B.1 Categories

Definition B.1 (Directed graph)**.** A **directed graph** G consists of a collection of **nodes** or **objects** denoted G_0 and a collection of **arrows** or **morphisms** denoted G_1 along with two maps $s, t : G_1 \to G_0$, so that each arrow $f \in G_1$ has a **source** $s(f)$ and a **target** $t(f)$.

We draw a morphism as an arrow, the source being its tail and target being its head:

$$s(f) \xrightarrow{f} t(f)$$

Definition B.2 (Paths)**.** A **path** in a directed graph G is a sequence of arrows $(f_1, \dots, f_k)$ that are **composable** in the sense that $t(f_i) = s(f_{i-1})$ for $i = 2, \dots, k$ as drawn below in (0). The collection of paths of length k in G will be denoted G_k.[55]

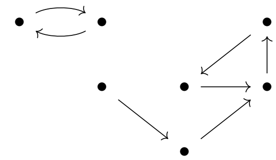
$$\bullet \xrightarrow{f_k} \bullet \xrightarrow{f_{k-1}} \bullet \cdots \bullet \xrightarrow{f_2} \bullet \xrightarrow{f_1} \bullet \tag{0}$$

[55] The **length** of a path is the number of arrows in it. It is fitting that G_1 denotes the arrows of G and the paths of length 1 in G as they are the same thing.

Observe that when referring to a path as $(f_1, \dots, f_k)$ or drawing it as in (0), there is a mismatch in the ordering of the arrows. The order as drawn — also called the diagrammatic order — agrees with the usual notation in graph theory (the branch of mathematics concerned with studying graphs), and it is arguably a more intuitive representation of the word "path". The other order will be motivated when we will define the composition of arrows in a category. The main idea is that, conceptually, arrows coincide more closely with functions between mathematical objects, and if we see the arrows in (0) as functions, their composition is most of the time denoted by $f_1 \circ \cdots \circ f_k$.

Examples B.3. It is very simple to give an example of a directed graph by drawing a bunch of nodes and arrows between them as in (1), G_0 is the collection of nodes, G_1

is the collection of arrows and s and t can be inferred from looking at the head and tail of each arrow. Let us give more examples to motivate the next definition.



(1)

1. For any set X, there is a trivial directed graph with X as its collection of nodes and no arrows. The source and target maps are the unique functions $\emptyset \to X$. You can represent it by drawing a node for each element of X.[56]

   There is a slightly more complex directed graph whose nodes are the elements of X. For each pair $(x, x') \in X \times X$, we can add an arrow with source x and target x'. Drawing it is still fairly simple[57]: you draw a node for each element of X and an arrow from x to x' for each pair (x, x').[58]

[56] This is a very uninteresting directed graph.

[57] Provided the set X is finite

[58] Note that there are so-called **loops** which are arrows from a node to itself because (x, x) is in $X \times X$.

2. Starting from a set X, we can define another directed graph by letting X be its only node and the collection of arrows be the set of functions from X to itself. The source and target maps are uniquely determined again, this time by their codomain that contains only the node X. This graph is already more interesting since the collection of arrows has a monoid structure. Indeed, the operation of composition of functions is associative, and the identity function is the identity for this operation.

3. Taking inspiration from the previous examples, we define a directed graph **Set**. It contains one node for every set, i.e., $\mathbf{Set}_0$ is the collection of all sets, [59] and one arrow with source X and target Y for every function $f : X \to Y$.

   Similarly to the last example, we recognize that the collection of arrows has a novel kind of structure induced by composition of functions and identity functions. It is not a monoid because you can only compose functions when one's source is the target of the other. In other words, composition of functions is not a binary operation $\circ : \mathbf{Set}_1 \times \mathbf{Set}_1 \to \mathbf{Set}_1$, it is of type $\mathbf{Set}_2 \to \mathbf{Set}_1$. Nonetheless, we still have associativity and identities which are at the core of the definition of a monoid. Since the theory of monoids is extremely rich and ubiquitous in mathematics, it is daring to study this seemingly more complex variant. We first need to make this structure abstract in the definition of a category.

[59] Notice how we could not have defined this graph if we required G_0 to be a set.

Definition B.4 (Category)**.** A directed graph $\mathbf{C}$ along with a **composition** map $\circ : \mathbf{C}_2 \to \mathbf{C}_1$ is a **category** if it satisfies the following properties:

1. For any $(f, g) \in \mathbf{C}_2$, $s(f \circ g) = s(g)$ and $t(f \circ g) = t(f)$. This is more naturally understood visually in (2).

2. For any $(f, g, h) \in \mathbf{C}_3$, $f \circ (g \circ h) = (f \circ g) \circ h$, namely, composition is associative. Again, the graphic representation in (3) may be more revealing.

3. For any object $A \in \mathbf{C}_0$, there exists an **identity** morphism $u_{\mathbf{C}}(A) \in \mathbf{C}_1$ with A as its source and target that satisfies $u_{\mathbf{C}}(A) \circ f = f$ and $g \circ u_{\mathbf{C}}(A) = g$, for any $f, g \in \mathbf{C}_1$ where $t(f) = A$ and $s(g) = A$.

$$\begin{tikzcd} \bullet \arrow[r, "g"'] \arrow[rr, bend left, "f \circ g"] & \bullet \arrow[r, "f"'] & \bullet \end{tikzcd} \tag{2}$$

$$\begin{tikzcd} \bullet \arrow[r, "h"] \arrow[rrr, bend left, "(f \circ g) \circ h"] \arrow[rr, bend right, "g \circ h"'] \arrow[rrr, bend right=50, "f \circ (g \circ h)"'] & \bullet \arrow[r, "g"] \arrow[rr, bend left, "f \circ g"] & \bullet \arrow[r, "f"] & \bullet \end{tikzcd} \tag{3}$$

If the third property of Definition B.4 is not satisfied, $\mathbf{C}$ is referred to as a **semicategory**. In rare occasions, authors choose to explicit when a category *does* satisfy this property, qualifying it as unital.

Remark B.5 (Notation). In general, we will refer to categories with bold uppercase letters typeset with `\mathbf` ($\mathbf{C}$, $\mathbf{D}$, $\mathbf{E}$, etc.), their objects with uppercase letters (A, B, X, Y, Z, etc.) and their morphisms with lowercase letters (f, g, h, etc.). When the category is clear from the context, we denote the identity morphisms id_A instead of $u_{\mathbf{C}}(A)$. We say that two morphisms are **parallel** if they have the same source and target. Given morphisms f and g in a category $\mathbf{C}$, we say that f **factors through** g if there exists $h \in \mathbf{C}_1$ such that $f = g \circ h$ or $f = h \circ g$.

Observe that since $\circ$ is associative, it induces a unique composition map on paths of any finite lengths, which we abusively denote $\circ : \mathbf{C}_k \to \mathbf{C}_1$.[60] This lets us write $f_1 \circ f_2 \circ \cdots \circ f_k$ with no parentheses. Occasionally, we will refer to the image of the path under this map as the **composition of the path** or the **morphism that a path composes to**.

[60] Another abuse we make is to define $\circ : \mathbf{C}_0 \to \mathbf{C}_1$ by $X \mapsto \mathrm{id}_X$. That is, we identify objects of $\mathbf{C}$ with empty paths (of length 0) starting and ending at that object, and we consider the composition of an empty path to be the identity.

Examples B.6 (Boring examples)**.** It can be really easy to construct a category by drawing its underlying directed graph and inferring the definition of the composition from it. Starting from the very simple graph depicted in (4), we can infer the definition of a category with a single object and its identity morphism. This category is denoted $\mathbf{1}$, the composition is trivial since $\mathrm{id}_\bullet \circ \mathrm{id}_\bullet = \mathrm{id}_\bullet$.

$$\begin{tikzcd} \bullet \arrow[loop above] \end{tikzcd} \tag{4}$$

Similarly, we construct from the graph in (5) a category with two objects, their identity morphisms and nothing else. The composition is again trivial. This category will be denoted $\mathbf{1} + \mathbf{1}$.[61] More generally, for any collection $\mathbf{C}_0$, there is a category $\mathbf{C}$ whose collection of objects is $\mathbf{C}_0$ and whose collection of morphisms is $\mathbf{C}_1 := \{\mathrm{id}_X \mid X \in \mathbf{C}_0\}$. The composition map is completely determined by the third property in Definition B.4.[62] A category without non-identity morphisms is called a **discrete category**.

$$\begin{tikzcd} \bullet_1 \arrow[loop above] & \bullet_2 \arrow[loop above] \end{tikzcd} \tag{5}$$

[61] This notation is cleared up in Definition D.20.

[62] i.e. the only elements of $\mathbf{C}_2$ are pairs $(\mathrm{id}_X, \mathrm{id}_X)$ for some $X \in \mathbf{C}_0$ and $\mathrm{id}_X \circ \mathrm{id}_X$ must be equal to id_X.

The graph in (6) corresponds to the category with objects $\{A, B\}$ and morphisms $\{\mathrm{id}_A, \mathrm{id}_B, f\}$.

$$\begin{tikzcd} A \arrow[loop left, "\mathrm{id}_A"] \arrow[r, "f"] & B \arrow[loop right, "\mathrm{id}_B"] \end{tikzcd} \tag{6}$$

The composition map is then completely determined by the properties of identity morphisms.[63] This category is called the interval category or the **walking arrow**, and it is denoted $\mathbf{2}$. Note however that $\mathbf{1} + \mathbf{1} \neq \mathbf{2}$.

[63] i.e., $f \circ \mathrm{id}_A = f$, $\mathrm{id}_B \circ f = f$, $\mathrm{id}_A \circ \mathrm{id}_A = \mathrm{id}_A$ and $\mathrm{id}_B \circ \mathrm{id}_B = \mathrm{id}_B$

Starting now, we will omit the identity morphisms from the diagrams (as is usual in the literature) for clarity reasons: they would hinder readability without adding information.

It is not always as straightforward to construct a category from a directed graph. For instance, if two distinct arrows have the same source and target, they must be explicitly drawn and the ambiguity in the composition must be dealt with. The graph in (7) is problematic in this way: it has two distinct paths of length two starting at the top-left corner and ending at the bottom-right corner. Since the composition of these paths can be equal to any of the two distinct morphisms between these corners, there is no category obviously corresponding to this graph.

$$\begin{tikzcd} \bullet \arrow[r] \arrow[d] \arrow[dr, shift left] \arrow[dr, shift right] & \bullet \arrow[d] \\ \bullet \arrow[r] & \bullet \end{tikzcd} \tag{7}$$

Since categories can be quite huge, it is rare that we draw all of a category at once. We will often draw diagrams with (labelled) nodes and arrows to represent the objects and morphisms within a category that we are focusing on. We also omit from our diagrams morphisms that can be inferred from the categorical structure. For instance, if we draw two composable morphisms as in (8), we do not draw the identity morphisms nor the composition $g \circ f$.

$$A \xrightarrow{f} B \xrightarrow{g} C \tag{8}$$

In many cases, not drawing all morphisms can lead to ambiguities like for (7). We have to be careful to avoid these, but sometimes we can resolve the problem by stating that the diagram is **commutative**.

Definition B.7 (Commutativity)**.** Given a diagram representing objects and morphisms in a category, we say that it is **commutative** if the composition of any path of length greater than one is equal to the composition of any other path with the same source and target. The morphism resulting from the composition may or may not be depicted.

Examples B.8. Arguably the most frequently used commutative diagram is the commutative square drawn in (9).

$$\begin{tikzcd} \bullet \arrow[r] \arrow[d] & \bullet \arrow[d] \\ \bullet \arrow[r] & \bullet \end{tikzcd} \tag{9}$$

We say the square commutes when the bottom and top paths compose to the same (omitted in the diagram) morphism. The commutative square can also be seen as a category by inferring the missing morphism and the composition from commutativity. We can denote it $\mathbf{2} \times \mathbf{2}$.[64]

[64] This notation is explained in Definition B.40.

Assuming that (10) commutes, we can infer that $f' \circ h = h' \circ f$, $g' \circ h' = h'' \circ g$, and $g' \circ f' \circ h = h'' \circ g \circ f$. Observe that the last equation can be derived from the first two which are equivalent to the commutativity of the two squares in (10). More generally, combining commutative diagrams in this way yields commutative diagrams, and this is the core of a powerful proof method called diagram paving that we introduce at the end of this chapter.

$$\begin{tikzcd} \bullet \arrow[r, "f"] \arrow[d, "h"'] & \bullet \arrow[r, "g"] \arrow[d, "h'"'] & \bullet \arrow[d, "h''"] \\ \bullet \arrow[r, "f'"'] & \bullet \arrow[r, "g'"'] & \bullet \end{tikzcd} \tag{10}$$

$$\begin{tikzcd} A \arrow[r, shift left, "f"] & B \arrow[l, shift left, "g"] \end{tikzcd} \tag{11}$$

Stating that (11) commutes is equivalent to stating that $f \circ g \circ f = f$ and $g \circ f \circ g = g$. We can also infer that $f \circ g \circ f \circ g = f \circ g$ and $g \circ f \circ g \circ f = g \circ f$, but this follows from the first two equality.

It would be odd to require that (7) commutes. It would imply that the two parallel morphisms are equal because they are both equal to the composition of the bottom and top paths. We will never draw parallel morphisms when they are supposed to be equal.

To assert that two morphisms $f, g : A \to B$ are equal using a diagram, we can say that either of the following is commutative, with a preference for the third one.[65]

[65] The equal sign in the third one can be read as id_A going in either direction.

$$\begin{tikzcd} A \arrow[d, "\mathrm{id}_A"'] \arrow[dr, "f"] & \\ A \arrow[r, "g"'] & B \end{tikzcd} \qquad \begin{tikzcd} A \arrow[dr, "f"] & \\ A \arrow[u, "\mathrm{id}_A"] \arrow[r, "g"'] & B \end{tikzcd} \qquad \begin{tikzcd} A \arrow[d, equal] \arrow[dr, "f"] & \\ A \arrow[r, "g"'] & B \end{tikzcd} \tag{12}$$

Remark B.9 (Convention). Reasoning with commutative diagrams is an acquired skill we will practice quite a lot in the following chapters. Yet there is no standard definition that everyone systematically uses.[66] For this reason, I decided to pick my favorite definition of commutativity which is uncommon.[67] In most cases, a diagram is called commutative when any two paths compose to the same morphism, but in practice, there are two exceptions handled by Definition B.7:

1. Two parallel morphisms are not always equal in a commutative diagram. In fact, when parallel morphisms are drawn, it is usually to emphasize that they are distinct.

2. Unless otherwise stated, an endomorphism[68] drawn in a commutative diagram is not equal to the identity morphism (the composition of the empty path).

[66] This does not really lead to many misunderstandings anyway because what is meant by a diagram is usually made clear by the context.

[67] I have not seen the constraint on the length anywhere else.

[68] An **endomorphism** is a morphism whose source and target coincide.

Warning B.10. Diagrams are not commutative by default. We will always specify when a diagram commutes. As our usage of commutative diagrams ramps up in the following chapters, you have to try to remember that.

Before moving on to more interesting categories, we introduce the Hom notation.

Definition B.11 (Hom)**.** Let $\mathbf{C}$ be a category and $A, B \in \mathbf{C}_0$ be objects, the collection of all morphisms going from A to B is

$$\mathrm{Hom}_{\mathbf{C}}(A,B) := \{f \in \mathbf{C}_1 \mid s(f) = A \text{ and } t(f) = B\}.$$

This leads to an alternative way of defining the morphisms of $\mathbf{C}$, namely, one can describe $\mathrm{Hom}_{\mathbf{C}}(A,B)$ for all $A, B \in \mathbf{C}_0$ instead of describing $\mathbf{C}_1$ all at once. Defining the morphisms this way also takes care of the source and target functions implicitly.

Remark B.12 (Notation). Some authors choose to denote the collection of morphisms between A and B with $\mathbf{C}(A,B)$. I prefer to use the latter notation when working with 2–categories[69] to highlight the fact that $\mathbf{C}(A,B)$ has more structure. Other authors use hom with a lowercase "h", my choice here is arbitrary.

[69] see Definition F.33.

Definition B.13 (Smallness)**.** A category $\mathbf{C}$ is called **small** if the collections of objects and morphisms are sets. If for all objects $A, B \in \mathbf{C}_0$, $\mathrm{Hom}_{\mathbf{C}}(A,B)$ is a set, $\mathbf{C}$ is said to be **locally small** and $\mathrm{Hom}_{\mathbf{C}}(A,B)$ is called a **hom-set**. A category that is not small can be referred to as **large**.

The following three examples will follow us throughout the book.

Example B.14 (**Set**)**.** The category **Set** has the collection of sets as its objects and for any sets X and Y, $\mathrm{Hom}_{\mathbf{Set}}(X,Y)$ is the set of all the functions from X to Y.[70] The composition map is given by composition of functions (which is associative) and the identity maps serve as the identity morphisms. This category is locally small but not small.[71]

We will carry out many examples using **Set** because it is elaborate enough to be interesting, yet it is easy to understand because we are (assumed to be) very familiar with sets and functions.

[70] We already saw this directed graph in Example B.3.3.

[71] By our argument at the start of Chapter A: the collection of all sets cannot be a set.

Example B.15. Let $(X, \leq)$ be a partially ordered set, it can be viewed as a category with elements of X as its objects. For any $x, y \in X$, the hom-set $\mathrm{Hom}_X(x, y)$ contains a single morphism if $x \leq y$ and is empty otherwise. The identity morphisms arise from the reflexivity of $\leq$. Since every hom-set contains at most one element and $\leq$ is transitive, the composition map is completely determined. Detailing this out, if $f : x \to y$ and $g : y \to z$ are morphisms, then we know that $x \leq y$ and $y \leq z$. Thus, transitivity implies that $x \leq z$ and there is a unique morphism $x \to z$, so it must be $g \circ f$.[72]

[72] Note that antisymmetry was not used in this argument, so one can more generally construct a category starting from a preorder. Such categories are called **thin** because each hom-set contains at most one morphism. It is straightforward to show the identities and composition ensure that any thin category $\mathbf{C}$ is constructed from the preorder $(\mathbf{C}_0, \leq)$ with

$$X \leq Y \Leftrightarrow \mathrm{Hom}_{\mathbf{C}}(X, Y) \neq \emptyset.$$

If a category corresponds to this construction for some poset, it is called **posetal**. In (13), we depict the posetal category associated to $(\mathbb{N}, \leq)$. The arrows between numbers n and $n + k$ are omitted for $k > 1$ as they can be inferred by the composition $n \leq n + 1 \leq n + 2 \leq \cdots \leq n + k$.

$$\overset{0}{\bullet} \longrightarrow \overset{1}{\bullet} \longrightarrow \overset{2}{\bullet} \longrightarrow \cdots \tag{13}$$

As a particular case of posetal categories, let (X, τ) be a topological space and note that the inclusion relation on open sets is a partial order on τ. Thus, X has a corresponding posetal category. More explicitly, the objects are open sets and for any $U, V \in \tau$, the hom-set $\mathrm{Hom}_X(U, V)$ contains the inclusion map i_{UV} if $U \subseteq V$ and is empty otherwise. This category will be denoted $\mathcal{O}(X, \tau)$ or $\mathcal{O}(X)$.

We will carry out many examples using posetal categories because it avoids difficulties arising from having different parallel morphisms.[73] In particular, every diagram drawn with objects and morphisms from a posetal category is commutative because the composition of any path is equal to the unique morphism between the source and target of that path. This also means some important aspects of a concept can be trivial when instantiating it for a posetal category.

[73] For the same reason, thin categories are also simple cases to carry out examples with.

Example B.16 (Single object categories)**.** If a category $\mathbf{C}$ has a single object $*$, then all morphisms go from $*$ to $*$. In particular, $\mathbf{C}_1 = \mathrm{Hom}_{\mathbf{C}}(*, *)$ and $\mathbf{C}_2 = \mathbf{C}_1 \times \mathbf{C}_1$. Then, the associativity of $\circ$ and existence of id_* make $(\mathbf{C}_1, \circ)$ into a monoid.

Conversely, a monoid $(M, \cdot)$ can be represented by a single object category M, where $\mathrm{Hom}_M(*, *) = M$ and the composition map is the monoid operation.

Since many algebraic structures have an associative operation with an identity element, this yields a fairly general construction. The single object category associated to a monoid or group G will be denoted by $\mathbf{B}G$ and referred to as the **delooping** of G.

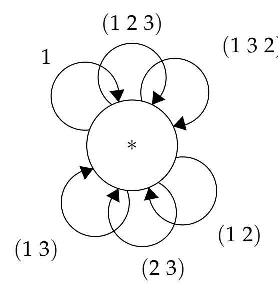


Figure B.1: The delooping of the symmetric group S_3, a.k.a. $\mathbf{B}S_3$.

The natural numbers can also be endowed with the monoid structure of addition, hence a particular instance of a single object category is the delooping of $(\mathbb{N}, +)$. Notice that this category is very different from the posetal category $(\mathbb{N}, \leq)$. In the former, $\mathbb{N}$ is in correspondence with the morphisms while in the latter, it is in correspondence with the objects.

We will carry out many examples using deloopings of monoids or groups because it avoids difficulties arising from having two different objects.

Several simple examples of large categories arise as subcategories of **Set**.

Definition B.17 (Subcategory)**.** Let $\mathbf{C}$ be a category, a category $\mathbf{C}'$ is a **subcategory** of $\mathbf{C}$ if, the following properties are satisfied.

1. The objects and morphisms of $\mathbf{C}'$ are objects and morphisms of $\mathbf{C}$ (i.e., $\mathbf{C}'_0 \subseteq \mathbf{C}_0$ and $\mathbf{C}'_1 \subseteq \mathbf{C}_1$).
2. The source and target maps of $\mathbf{C}'$ are the restrictions of the source and target maps of $\mathbf{C}$ on $\mathbf{C}'_1$ and for every morphism $f \in \mathbf{C}'_1$, $s(f), t(f) \in \mathbf{C}'_0$.
3. The composition map of $\mathbf{C}'$ is the restriction of the composition map of $\mathbf{C}$ on $\mathbf{C}'_2$ and for any $(f,g) \in \mathbf{C}'_2$, $f \circ_{\mathbf{C}'} g = f \circ_{\mathbf{C}} g \in \mathbf{C}'_1$.
4. The identity morphisms of objects in $\mathbf{C}'_0$ are the identity morphisms of objects in $\mathbf{C}_0$, i.e., $u_{\mathbf{C}}(A) = u_{\mathbf{C}'}(A)$ when $A \in \mathbf{C}'_0$.

Intuitively, one can see $\mathbf{C}'$ as being obtained from $\mathbf{C}$ by removing some objects and morphisms, but making sure that no morphism is left with no source or no target and that no path is left without its composition.

OL **Exercise B.18** (NOW!)**.** Find an example of a category $\mathbf{C}$ and a category $\mathbf{C}'$ that satisfy the first three conditions but not the fourth.

Definition B.19 (Full and wide)**.** A subcategory $\mathbf{C}'$ of $\mathbf{C}$ is called **full** if for any objects $A, B \in \mathbf{C}'_0$, $\mathrm{Hom}_{\mathbf{C}'}(A,B) = \mathrm{Hom}_{\mathbf{C}}(A,B)$. It is called **wide** if $\mathbf{C}'_0 = \mathbf{C}_0$.[74]

[74] In words, a subcategory is full if the morphisms that were removed had their source or target removed as well, and it is wide if no objects were removed.

Examples B.20 (Subcategories of **Set**)**.** We can selectively remove some objects and morphisms in **Set** to obtain the following categories.

1. Since the composition of injective functions is again injective, the restriction of morphisms in **Set** to injective functions yields a wide subcategory of **Set**, denoted by **SetInj**. Unsurprisingly, **SetSurj** can be constructed similarly.
2. Removing all infinite sets from **Set** yields the full subcategory of finite sets denoted **FinSet**.[75]
3. Further removing sets from **FinSet** and keeping only $\emptyset$, $\{1\}$, $\{1,2\}$, $\{1,2,3\}$, etc., we obtain the category **FinOrd** which is a small full subcategory of **Set**.[76]
4. Since the composition of monotone maps is monotone and the identity function is monotone, we can view each set $\{1,\ldots,n\}$ as ordered with $\leq$ and remove all morphisms that are not monotone from **FinOrd**. The resulting category is called the **simplex category** and denoted by $\mathbf{\Delta}$.

[75] This category is not small because there is no set of all finite sets.

[76] The name **FinOrd** is an abbreviation of finite ordinals, because we can also define **FinOrd** as the category of finite ordinals and functions between them.

Examples B.21 (Concrete categories)**.** This second list of examples contains so-called concrete categories. Informally, they are categories of sets with extra structure, where morphisms are functions that preserve that extra structure.[77]

[77] Formally, see Definition B.35.

1. The category $\mathbf{Set}_*$ is the category of **pointed** sets. Its objects are sets with a distinguished element, and its morphisms are functions that map distinguished elements to distinguished elements. The distinguished element of a pointed set is

the extra structure on top of the set, and morphisms between pointed set must preserve that structure. In more details, $(\mathbf{Set}_*)_0$ is the collection of pairs (X, x) where X is a set and $x \in X$, and for any two pointed sets (X, x) and (Y, y),

$$\mathrm{Hom}_{\mathbf{Set}_*}((X, x), (Y, y)) = \{f : X \to Y \mid f(x) = y\}.$$

The identity morphisms and composition are defined as in **Set**, so the axioms of a category clearly hold after checking that if $f : (X, x) \to (Y, y)$ satisfies $f(x) = y$ and $g : (Y, y) \to (Z, z)$ satisfies $g(y) = z$, then $(g \circ f)(x) = z$.

2. The category **Mon** is the category of monoids and their homomorphisms, let us be more explicit.[78] The objects are monoids, so $\mathbf{Mon}_0$ is the collection of all monoids, and the morphisms are monoid homomorphisms, so for any $M, N \in \mathbf{Mon}_0$, $\mathrm{Hom}_{\mathbf{Mon}}(M, N)$ is the set of homomorphisms from M to N. The composition in **Mon** is given by the composition of homomorphisms, we know it is well-defined because the composition of two homomorphisms is a homomorphism. Also, the composition is associative and the identity functions are homomorphisms, so we can define $u_{\mathbf{Mon}}(M) = \mathrm{id}_M$.

[78] These technicalities are essentially the same for the categories in the remainder of Example B.21.

3. Similarly, the category of groups (resp. rings or fields) where the morphisms are group (resp. ring or field) homomorphisms is **Grp** (resp. **Ring** or **Field**). The category of abelian groups (resp. commutative monoids or rings) is a full subcategory of **Grp** (resp. **Mon** or **Ring**) denoted by **Ab** (resp. **CMon** or **CRing**).[79]

[79] Defining a category by saying it is a full subcategory of another one is a compact way of saying that we remove all the objects we do not want (e.g., the non-abelian groups) and nothing else.

4. Let k be a fixed field, the category of vector spaces over k where the morphisms are linear maps is $\mathbf{Vect}_k$. The full subcategory of $\mathbf{Vect}_k$ consisting only of finite dimensional vector spaces is $\mathbf{FDVect}_k$.

5. The category of partially ordered sets where morphisms are order-preserving functions is denoted by **Poset**. It is a full subcategory of **Pre**, the category of preorders.

   A poset is a set A equipped with a binary relation $\leq \subseteq A \times A$ (the extra structure) that satisfies some axioms (reflexivity, transitivity and antisymmetry). In some sense, we can see the axioms as structure on top of the extra structure that is $\leq$. For example, we can consider the category **2Rel** of sets equipped with a binary relation (we do not require the axioms of posets to hold). An object of 2**Rel** is a pair (A, R) where A is a set and $R \subseteq A \times A$ is a binary relation on A.[80] A morphism $(A, R_A) \to (B, R_B)$ is defined like order-preserving functions: it is a function $f : A \to B$ satisfying $\forall x, y \in A, (x, y) \in R_A \implies (f(x), f(y)) \in R_B$.

   The categories **Poset** and **Pre** are both full subcategories of 2**Rel** where we only keep the relations satisfying the appropriate axioms.

[80] We use a nondescript letter for the relation instead of a symbol like $\leq$ to avoid being misled by the intuitions we have for partial orders.

6. The category of topological spaces where morphisms are continuous functions is denoted by **Top**.

7. The category of metric spaces where morphisms are nonexpansive functions is denoted by **Met**.

In these last two examples, the choice of morphisms to take between spaces is not as clear cut as for the previous examples. For instance, one could ask the morphism between metric spaces to be continuous also, or for morphisms between topological spaces to map open sets to open sets (those are called open maps). In the end, the choice made depends on the context where the category is used.

OL **Exercise B.22.** An n–ary relation on a set A is a subset of A^n. Define the category n**Rel**.

Our next example is a large category that is neither a subcategory of **Set** nor a concrete category.

Example B.23 (Rel). The category of sets and relations, denoted by **Rel**,[81] has as objects the collection of all sets, and for any sets X and Y, $\mathrm{Hom}_{\mathbf{Rel}}(X, Y)$ is the set of relations between X and Y, that is, the powerset of $X \times Y$. The composition of two relations $R \subseteq X \times Y$ and $S \subseteq Y \times Z$ is defined by

$$S \circ R = R\,;S := \{(x,z) \in X \times Z \mid \exists y \in Y, (x,y) \in R, (y,z) \in S\} \subseteq X \times Z.$$

One can check that this composition is associative and that, for any set X, the **diagonal relation** $\Delta_X = \{(x,x) : x \in X\} \subseteq X \times X$ is the identity with respect to this composition.

[81] The notations for **Rel** and n**Rel** look close, but these categories see relations from very different points of view.

If you are not familiar with composition of relations, try to understand it visually. Draw the sets X, Y and Z as regions with dots inside, the relation R as wires connecting some dots in X and Y, and the relation S as wires connecting some dots in Y and Z. The relation $R\,;S$ relates a dot $x \in X$ to a dot $z \in Z$ if you can follow a wire in R and a wire in S to go from x to z.

Examples can also be helpful. Let $X = Y = Z$ be the set of all humans, R be the "cousin" relation (i.e., $(x,y) \in R$ whenever x and y are cousins) and S be the "sibling" relation. You can verify that $R;S = R$, $S;S = S$, but $R;R \neq R$.

Remark B.24. You can view **Set** as a wide subcategory of **Rel** where you only take the relations $R \subseteq X \times Y$ satisfying for any $x \in X$,

$$|\{y \in Y \mid (x,y) \in R\}| = 1.$$

B.2 Functors

The list above is far from exhaustive; there are many more mathematical objects that can fit in a category and this is a main reason for studying this subject. Indeed, categories encapsulate a natural structure that accurately represents the heart of several mathematical theories from a global and abstract perspective.

If we were to develop category theory by mirroring the curriculum of most textbooks introducing abstract algebra, the rest of this chapter would be dedicated to exploring the insides of a category. We could talk about monomorphisms, epimorphisms, initial and terminal objects, subobjects, and even (co)limits inside a category. All these words will be defined in due time,[82] but not before explaining a guiding principle in category theory and setting an example by following it.

[82] Without relying on the rest of this chapter.

If we spend some more time studying Definition B.4, we realize that the objects of a category carry little to no structure, and they are way less important than the morphisms. For example, the categories **Set**, **SetInj**, **SetSurj**, and **Rel** all have the same collection of objects, but they are very dissimilar.[83] As a matter of fact, there are alternative (albeit more messy) definitions of categories that do not refer to objects.

[83] We do not have enough tools yet to formally point out their differences.

Furthermore, a category only has superficial information about what its objects and morphisms are. For example, the category **Grp** is only a bunch of nodes and arrows, identities and a composition map. We cannot recover the definition of a group or a group homomorphism from that information. At first, this might seem detrimental: how can we prove things about groups if we do not know what they are? A good chunk of category theorists' mindset is contained in this snarky response.

We do not need to know what they are, only how they interact with each other.

As we advance through this book, we will get more sense of how true and powerful this idea can be.[84] We quickly start this journey by defining functors which are how categories interact with each other.

Informally, a functor is a morphism of categories. Thus, to motivate the definition, we can look at other morphisms we have encountered. A clear similarity between categories like **Mon**, **Grp**, **Ring** or **Poset** is that all the objects are sets with some sort of structure that the morphisms preserve. In the first three categories, the structure on an object is the operations and identity elements that are preserved under homomorphisms, and in the last one, the structure on a poset is a relation that is preserved by order-preserving maps.[85] Hence, we go back to Definition B.4, and we see that the structure of a category consists of the source and target maps, the composition map and the identities.

[84] One could argue the culminating point of this book (and any introduction to category theory) is the Yoneda lemma (see Chapter G) which beautifully formalizes this idea.

[85] Not all morphisms are functions that preserve structure, see e.g. morphisms in posetal categories.

Definition B.25 (Functor)**.** Let $\mathbf{C}$ and $\mathbf{D}$ be categories, a **functor** $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ is a pair of maps $F_0 : \mathbf{C}_0 \to \mathbf{D}_0$ and $F_1 : \mathbf{C}_1 \to \mathbf{D}_1$ such that diagrams (14), (15) and (16) commute where F_2 is induced by the definition of F_1 with $F_2 = (f, g) \mapsto (F_1(f), F_1(g))$.[86]

$$\begin{tikzcd} \mathbf{C}_0 \arrow[d, "F_0"'] & \mathbf{C}_1 \arrow[l, "s"'] \arrow[r, "t"] \arrow[d, "F_1"] & \mathbf{C}_0 \arrow[d, "F_0"] \\ \mathbf{D}_0 & \mathbf{D}_1 \arrow[l, "s"] \arrow[r, "t"'] & \mathbf{D}_0 \end{tikzcd} \tag{14}$$

$$\begin{tikzcd} \mathbf{C}_2 \arrow[r, "F_2"] \arrow[d, "\circ_{\mathbf{C}}"'] & \mathbf{D}_2 \arrow[d, "\circ_{\mathbf{D}}"] \\ \mathbf{C}_1 \arrow[r, "F_1"'] & \mathbf{D}_1 \end{tikzcd} \tag{15}$$

$$\begin{tikzcd} \mathbf{C}_0 \arrow[r, "F_0"] \arrow[d, "u_{\mathbf{C}}"'] & \mathbf{D}_0 \arrow[d, "u_{\mathbf{D}}"] \\ \mathbf{C}_1 \arrow[r, "F_1"'] & \mathbf{D}_1 \end{tikzcd} \tag{16}$$

[86] It is the first time we use commutative diagrams and we are already cheating a bit. Indeed, these diagrams do not represent objects and morphisms of a category we know. They could live in the category **Set** if $\mathbf{C}$ and $\mathbf{D}$ were small, but in the general case, we would need a category of collections and functions. It does not exist because there is no collection of all collections. Fortunately, this does not impact how we read these commutative diagrams.

Remark B.26 (Digesting diagrams). Once again, we emphasize that commutative diagrams will be heavily employed to make clearer and more compact arguments,[87] and that it will take time to get used to them. For now, let us unpack the definition above to ease its comprehension.

Commutativity of these diagrams is equivalent to having the following equalities:

$$s \circ F_1 = F_0 \circ s \qquad t \circ F_1 = F_0 \circ t \qquad F_1 \circ \circ_{\mathbf{C}} = \circ_{\mathbf{D}} \circ F_2 \qquad F_1 \circ u_{\mathbf{C}} = u_{\mathbf{D}} \circ F_0$$

Unrolling further, a functor $F : \mathbf{C} \rightsquigarrow \mathbf{D}$[88] must satisfy the following properties.

i. For any $A, B \in \mathbf{C}_0$ and $f \in \mathrm{Hom}_{\mathbf{C}}(A, B)$, $F(f) \in \mathrm{Hom}_{\mathbf{D}}(F(A), F(B))$. This is equivalent to the commutativity of (14) which says $F_0(s(f)) = s(F_1(f))$ and $F_0(t(f)) = t(F_1(f))$.

ii. If $f, g \in \mathbf{C}_1$ are composable, then $F(f)$ and $F(g)$ are composable by i. and $F(f \circ_{\mathbf{C}} g) = F(f) \circ_{\mathbf{D}} F(g)$ by commutativity of (15).

iii. If $A \in \mathbf{C}_0$, then $u_{\mathbf{D}}(F(A)) = F(u_{\mathbf{C}}(A))$ by commutativity of (16).[89]

The subscript on F is often omitted, as is common in the literature, when it is clear whether F is applied to an object or a morphism. We will also denote application of F with juxtaposition instead of parentheses, i.e., we can write FA and Ff instead of $F(A)$ and $F(f)$.

[87] This is especially true when using a blackboard or pen and paper because it makes it easier to point at things. Sadly, I cannot point at things on this PDF you are reading.

[88] The $\rightsquigarrow$ (`\rightsquigarrow`) notation for functors is not that common, they are usually denoted with plain arrows because they are morphisms. Nonetheless, I feel it is useful to have a special treatment for functors until you get accustomed to them. The squiggly arrow notation is sometimes used for Kleisli morphisms which we cover in Chapter **??**.

[89] Alternatively, $\mathrm{id}_{F(A)} = F(\mathrm{id}_A)$.

Examples B.27 (Boring examples)**.** As usual, a few trivial constructions arise.

1. For any category $\mathbf{C}$, the **identity functor** $\mathrm{id}_{\mathbf{C}} : \mathbf{C} \rightsquigarrow \mathbf{C}$ is defined by letting $(\mathrm{id}_{\mathbf{C}})_0$ and $(\mathrm{id}_{\mathbf{C}})_1$ be identity maps on $\mathbf{C}_0$ and $\mathbf{C}_1$ respectively.

When the source and target of a functor coincide, we may refer to it as an **endofunctor**.

2. Let $\mathbf{C}$ be a category and $\mathbf{C}'$ a subcategory of $\mathbf{C}$, the **inclusion functor** $\mathcal{I} : \mathbf{C}' \rightsquigarrow \mathbf{C}$ is defined by letting $\mathcal{I}_0$ be the inclusion map $\mathbf{C}'_0 \hookrightarrow \mathbf{C}_0$ and $\mathcal{I}_1$ be the inclusion map $\mathbf{C}'_1 \hookrightarrow \mathbf{C}_1$.
3. Let $\mathbf{C}$ and $\mathbf{D}$ be categories and X be an object in $\mathbf{D}$, the **constant functor** $\Delta(X) : \mathbf{C} \rightsquigarrow \mathbf{D}$ sends every object to X and every morphism to id_X, i.e., $\Delta(X)_0(A) = X$ for any $A \in \mathbf{C}_0$ and $\Delta(X)_1(f) = \mathrm{id}_X$ for any $f \in \mathbf{C}_1$.

Examples B.28 (Less boring)**.** Functors with the source being one of $\mathbf{1}$, $\mathbf{2}$ or $\mathbf{2} \times \mathbf{2}$[90] are a bit less boring. Let the target be a category $\mathbf{C}$ and let us analyze these functors.

[90] $\mathbf{2} \times \mathbf{2}$ is the commutative square in (9)

- Let $F : \mathbf{1} \rightsquigarrow \mathbf{C}$, F_0 assigns to the single object $\bullet \in \mathbf{1}_0$ an object $F(\bullet) \in \mathbf{C}_0$. Then, by commutativity of (16), F_1 is completely determined by $\mathrm{id}_\bullet \mapsto \mathrm{id}_{F(\bullet)}$. We conclude that functors of this type are in correspondence with objects of $\mathbf{C}$.
- Let $F : \mathbf{2} \rightsquigarrow \mathbf{C}$, F_0 assigns to A and B, two objects $FA, FB \in \mathbf{C}_0$ and F_1's action on identities is fixed. Still, there is one choice to make for $F_1(f)$ which must be a morphism in $\mathrm{Hom}_{\mathbf{C}}(FA, FB)$. Therefore, F sums up to a choice of two objects in $\mathbf{C}$ and a morphism between them. In other words, functors of this type are in correspondence with morphisms in $\mathbf{C}$.[91]
- Similarly (we leave the details as an exercise), functors of type $F : \mathbf{2} \times \mathbf{2} \rightsquigarrow \mathbf{C}$ are in correspondence with commutative squares inside the category $\mathbf{C}$.[92]

[91] After picking a morphism, the source and target are determined.

[92] i.e., pairs of pairs of composable morphisms $((f,g),(f',g')) \in \mathbf{C}_2 \times \mathbf{C}_2$ satisfying $f \circ g = f' \circ g'$.

Remark B.29 (Functoriality)*.* We will use the term **functorial** as an adjective to qualify transformations that behave like functors and **functoriality** to refer to the property of behaving like a functor.

Throughout the rest of this book, the goal will essentially be to grow our list of categories and functors with more and more examples and perhaps exploit their properties wisely. Before pursuing this objective, we give important definitions analogous to injectivity and surjectivity of functions.

Definition B.30 (Full and faithful)**.** Let $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ be a functor. For $A, B \in \mathbf{C}_0$, denote the restriction of F_1 to $\mathrm{Hom}_{\mathbf{C}}(A,B)$ with

$$F_{A,B} : \mathrm{Hom}_{\mathbf{C}}(A,B) \to \mathrm{Hom}_{\mathbf{D}}(F(A), F(B)).$$

- If $F_{A,B}$ is injective for any $A, B \in \mathbf{C}_0$, then F is **faithful**.
- If $F_{A,B}$ is surjective for any $A, B \in \mathbf{C}_0$, then F is **full**.
- If $F_{A,B}$ is bijective for any $A, B \in \mathbf{C}_0$, then F is **fully faithful**.

OL **Exercise B.31** (NOW!)**.** Show that the inclusion functor $\mathcal{I} : \mathbf{C}' \rightsquigarrow \mathbf{C}$ is faithful. Show it is full if and only if $\mathbf{C}'$ is a full subcategory.

OL **Exercise B.32.** Let $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ and $G : \mathbf{D} \rightsquigarrow \mathbf{E}$. Show that

- if $G \circ F$ is faithful, then F is faithful, and
- if $G \circ F$ is full, then G is full.

As a generalization of the previous exercise, we note that a functor is full if and only if its image is a full subcategory of the target category.[93]

[93] The **image** of a functor $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ is the subcategory of $\mathbf{D}$ containing all objects and morphisms in the image of F_0 and F_1.

Remark B.33. While bijectivity is very strong to compare sets — it morally says that the elements of one set can be identified with the elements of another set — fully faithful functors are not as powerful. For instance, all functors between thin categories are fully faithful (because all the hom-sets are singletons). It should not be surprising that some fully faithful functors can be between two wildly unrelated categories because this property does not restrict the action on objects. We will see later what properties ensure that a functor strongly links the source and target category.

Examples B.34. For all but the first example, we leave you to prove functoriality.[94] In the literature, a lot of functors are given only with their action on objects and the reader is supposed to figure out the action on morphisms. Not everyone has the same innate ability to do this, but I hope this book can give you enough experience to overcome this difficulty.

[94] It is an elementary task that is mostly relevant to the field of mathematics the functor comes from.

1. The **powerset functor** $\mathcal{P} : \mathbf{Set} \rightsquigarrow \mathbf{Set}$ sends a set X to its powerset $\mathcal{P}(X)$[95] and a function $f : X \to Y$ to the image map $\mathcal{P}(f) : \mathcal{P}(X) \to \mathcal{P}(Y)$. The latter sends a subset $S \subseteq X$ to
$$\mathcal{P}(f)(S) = f(S) := \{f(s) \mid s \in S\} \subseteq Y.$$
In order to prove that $\mathcal{P}$ is a functor, we need to show it makes diagrams (14), (15), and (16) commute. Equivalently, we can show it satisfies the three conditions in Remark B.26.

[95] The powerset of X is the set of all subsets of X.

   i. For any function $f : X \to Y$, the source and target of the image map $\mathcal{P}f$ are $\mathcal{P}X$ and $\mathcal{P}Y$ respectively as required.
   ii. Given two functions $f : X \to Y$ and $g : Y \to Z$, we can verify that $\mathcal{P}g \circ \mathcal{P}f = \mathcal{P}(g \circ f)$ by looking at the action of both sides on a subset $S \subseteq X$.
$$\begin{aligned}\mathcal{P}g(\mathcal{P}f(S)) &= \{g(y) \mid y \in \mathcal{P}f(S)\}\\ &= \{g(y) \mid y \in \{f(x) \mid x \in S\}\}\\ &= \{g(f(x)) \mid x \in S\}\end{aligned} \qquad \begin{aligned}\mathcal{P}(g \circ f)(S) &= \{(g \circ f)(x) \mid x \in S\}\\ &= \{g(f(x)) \mid x \in S\}\end{aligned}$$
   iii. Finally, the image map of id_X is the identity on $\mathcal{P}X$ because
$$\mathcal{P}\mathrm{id}_X(S) = \{\mathrm{id}_X(x) \mid x \in S\} = \{x \mid x \in S\} = S.$$

The powerset functor is faithful because the same image map cannot arise from two different functions[96], it is not full because lots of functions $\mathcal{P}(X) \to \mathcal{P}(Y)$ are not image maps. A cardinality argument suffices: when $|X|, |Y| \geq 2$,
$$|\mathrm{Hom}_{\mathbf{Set}}(X, Y)| = |Y|^{|X|} < |\mathcal{P}(Y)|^{|\mathcal{P}(X)|} = |\mathrm{Hom}_{\mathbf{Set}}(\mathcal{P}(X), \mathcal{P}(Y))|.$$

[96] Indeed, if $f(x) \neq g(x)$, then $f(\{x\}) \neq g(\{x\})$.

2. The concrete categories of Examples B.21 are defined using a functor.

   Definition B.35 (Concrete category)**.** We call a category $\mathbf{C}$ **concrete** if it is paired (generally implicitly) with a faithful functor $U : \mathbf{C} \rightsquigarrow \mathbf{Set}$. In most cases, U is called the **forgetful functor** because it sends objects and morphisms of $\mathbf{C}$ to sets and functions by *forgetting* additional structure.

   The forgetful functor $U : \mathbf{Grp} \rightsquigarrow \mathbf{Set}$ sends a group $(G, \cdot, 1_G)$ to its underlying set G, *forgetting about the operation and identity*. It sends a group homomorphism $f : G \to H$ to the underlying function, *forgetting about the homomorphism properties*. It is faithful since if two homomorphisms have the same underlying function, then they are equal.[97]

   Briefly, functoriality of U follows from the facts that the underlying function of a homomorphism $f : G \to H$ goes between the underlying sets of G and H, the underlying function of a composition of homomorphisms is the composition of the underlying functions, and the underlying function of the identity homomorphism is the identity map.

3. It is also sometimes useful to consider *intermediate* forgetful functors. For example, $U : \mathbf{Ring} \rightsquigarrow \mathbf{Ab}$ sends a ring $(R, +, \cdot, 1_R, 0_R)$ to the abelian group $(R, +, 0_R)$, *forgetting about multiplication and* 1_R. It sends a ring homomorphism $f : R \to S$ to the same underlying function seen as a group homomorphism.[98] Not any old functor $\mathbf{Ring} \rightsquigarrow \mathbf{Ab}$ can be considered an intermediate forgetful functor. The key property is that forgetting about multiplication and 1_R ($\mathbf{Ring} \rightsquigarrow \mathbf{Ab}$) and then forgetting about the addition and 0_R ($\mathbf{Ab} \rightsquigarrow \mathbf{Set}$) is the same thing as forgetting all the ring structure at once ($\mathbf{Ring} \rightsquigarrow \mathbf{Set}$).

   The inclusion functor of **Poset** into $2\mathbf{Rel}$ is also an intermediate forgetful functor. It forgets about all the properties of the partial order, but it does not forget about the binary relation.

4. In some cases, there is a canonical way to go in the opposite direction to the forgetful functor, it is called the free functor. For **Mon**, the free functor $F : \mathbf{Set} \rightsquigarrow \mathbf{Mon}$ sends a set X to the free monoid generated by X and a function $f : X \to Y$ to the unique group homomorphism $F(X) \to F(Y)$ that restricts to f on the set of generators.[99]

   In Chapter H, when covering adjunctions, we will study a strong relation between the forgetful functor U and the free functor F that will generalize to other mathematical structures.

5. Let $(X, \leq)$ and $(Y, \sqsubseteq)$ be posets, and $F : X \rightsquigarrow Y$ be a functor between their posetal categories. For any $a, b \in X$, if $a \leq b$, then $\mathrm{Hom}_X(a, b)$ contains a single element, thus $\mathrm{Hom}_Y(F(a), F(b))$ must contain a morphism as well,[100] or equivalently $F(a) \sqsubseteq F(b)$. This shows that F_0 is an order-preserving function on the posets.

   Conversely, any order-preserving function between X and Y will correspond to a unique functor as there is only one morphism in all the hom-sets.[101]

[97] We leave you the repetitive task to describe the forgetful functor for every concrete category in Examples B.21.

[98] It can do that because part of the requirements for ring homomorphisms is to preserve the underlying additive group structure.

[99] More details about free monoids are in Chapter E.

[100] The image of the element in $\mathrm{Hom}_X(a, b)$ under F.

[101] Given $f : (X, \leq) \to (Y, \sqsubseteq)$ order-preserving, the corresponding functor between the posetal categories of X and Y acts like f of the objects and sends a morphism $a \to b$ to the unique morphism $f(a) \to f(b)$ which exists because $a \leq b \implies f(a) \sqsubseteq f(b)$.

SOL Exercise B.36. Let A and B be two sets, their powersets can be seen as posets with the order $\subseteq$. Thus, we can view $\mathcal{P}(A)$ and $\mathcal{P}(B)$ as posetal categories.

- Draw (using points and arrows) the category corresponding to $\mathcal{P}(\{0,1,2\})$.
- Show that the image and preimage functions defined below are functors between these categories.[102]

$$f : \mathcal{P}(A) \to \mathcal{P}(B) = S \mapsto \{f(a) \mid a \in S\}$$
$$f^{-1} : \mathcal{P}(B) \to \mathcal{P}(A) = S \mapsto \{a \in A \mid f(a) \in S\}$$

[102] i.e., they are order-preserving functions.

6. Let G and H be groups and $\mathbf{B}G$ and $\mathbf{B}H$ be their respective deloopings, then the functors $F : \mathbf{B}G \rightsquigarrow \mathbf{B}H$ are exactly the group homomorphisms from G to H.[103] Let $F : \mathbf{B}G \rightsquigarrow \mathbf{B}H$ be a functor, the action of F on objects is trivial since there is only one object in both categories. On morphisms, F_1 is a function from G to H which preserves composition and the identity morphism which, by definition, are the group multiplication and identity respectively. Thus, F_1 is a group homomorphism.

   Given a homomorphism $f : G \to H$, the reverse reasoning shows we obtain a functor $\mathbf{B}G \rightsquigarrow \mathbf{B}H$ by acting trivially on objects and with f on morphisms.

[103] Similarly for the deloopings of monoids.

7. For any group G, the functors $F : \mathbf{B}G \rightsquigarrow \mathbf{Set}$ are in correspondence with left actions of G. Indeed, if $S = F(*)$, then

$$F_1 : G = \mathrm{Hom}_{\mathbf{B}G}(*,*) \to \mathrm{Hom}_{\mathbf{Set}}(S,S)$$

   is such that $F(gh) = F(g) \circ F(h)$ for any $g, h \in G$ and $F(1_G) = \mathrm{id}_S$.[104] Moreover, since for any $g \in G$,

$$F(g^{-1}) \circ F(g) = F(g^{-1}g) = F(1_G) = \mathrm{id}_S = F(1_G) = F(gg^{-1}) = F(g) \circ F(g^{-1}),$$

   the function $F(g)$ is a bijection (its inverse is $F(g^{-1})$) and we conclude F_1 is the permutation representation of the group action defined by $g \star s = F(g)(s)$ for all $g \in G$ and $s \in S$.

   Given a group action on a set S, we leave you to show that letting $F_0 = * \mapsto S$ and F_1 be the permutation representation of the action yields a functor $F : \mathbf{B}G \rightsquigarrow \mathbf{Set}$.

[104] This is because gh is the composite of g and h in $\mathbf{B}G$ and 1_G is the identity morphism in $\mathbf{B}G$.

8. In the previous example, replacing $\mathbf{Set}$ with $\mathbf{Vect}_k$, one obtains k-linear representations of G instead of actions of G.[105]

[105] You might not know about linear representations, we just mention them in passing.

Remark B.37 (Non-examples). From this long (and yet hardly exhaustive) list, one might get the feeling that every important mathematical transformation is a functor. This is not the case, so I wanted to show where functoriality can fail and hopefully give you a bit of intuition about why they fail. Here are two instances showcasing the two most common ways (in my experience) you can decide that a mapping is not functorial.

Let us define $F : \mathbf{FDVect}_k \rightsquigarrow \mathbf{Set}$ which assigns to any vector space over k a choice of basis. There is no non-trivializing way to define an action of F on linear maps which make F into a functor. One informal reason for this failure is that we cannot choose bases globally, so F is defined locally and its parts cannot be glued together.[106]

Another non-example is given by the center[107] of a group in **Grp**. A homomorphism $H \to G$ does not necessarily send the center of H in the center of G (take for instance $S_2 \hookrightarrow S_3$), thus, we cannot easily define the function $Z(H) \to Z(G)$ induced by the homomorphism (unless we send everything to $1_G \in Z(G)$). This time, Z is not a functor because it does not interact well with the morphisms of the category. Actually, if you decided to only keep group isomorphisms in the category, you could define the functor Z because isomorphisms preserve the center of groups.

[106] If you feel like you are making a non-canonical choice for every object, there is a good chance you are not dealing with a functor.

[107] The **center** of a group G, often denoted $Z(G)$, is the subset of G containing elements that commute with all other elements, i.e.,

$$Z(G) = \{x \in G \mid \forall g \in G, xg = gx\}.$$

In this chapter, we introduced a novel structure, namely categories, that functors preserve.[108] Since we also introduced several categories where objects had some structure that morphisms preserve, it is reasonable to wonder whether categories and functors are also part of a category. In fact, the only missing ingredient is the composition of functors (we already know what the source and target of a functor is and every category has an identity functor). After proving the following proposition, we end up with the category **Cat** where objects are small categories and morphisms are functors.[109]

[108] We defined functors precisely so that they preserve the structure of categories.

[109] In order to avoid paradoxes of the Russel kind, it is essential to restrict **Cat** to contain only small categories.

Proposition B.38. *Let $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ and $G : \mathbf{D} \rightsquigarrow \mathbf{E}$ be functors and $G \circ F : \mathbf{C} \rightsquigarrow \mathbf{E}$ be their* ***composition*** *defined by $G_0 \circ F_0$ on objects and $G_1 \circ F_1$ on morphisms. Then, $G \circ F$ is a functor.*

Proof. One could proceed with a really hands-on proof and show that $G \circ F$ satisfies the three necessary properties in a manner not unlike when proving the group homomorphisms compose. This should not be too hard, but you will have to deal with notation for objects, morphisms and the composition from all three different categories. This can easily lead to confusion or worse: boredom!

Instead, we will use the diagrams we introduced in the first definition of a functor. From the functoriality of F and G, we get two sets of three diagrams and combining them yields the diagrams for $G \circ F$.[110]

[110] Since F is a functor, the top two squares of (17) and the left squares of (18) and (19) commute. Since G is a functor, the bottom two squares (17) and the right squares of (18) and (19) commute.

$$\begin{tikzcd}
\mathbf{C}_0 \arrow[d, "F_0"'] & \mathbf{C}_1 \arrow[l, "s"'] \arrow[r, "t"] \arrow[d, "F_1"] & \mathbf{C}_0 \arrow[d, "F_0"] \\
\mathbf{D}_0 \arrow[d, "G_0"'] & \mathbf{D}_1 \arrow[l, "s"'] \arrow[r, "t"] \arrow[d, "G_1"] & \mathbf{D}_0 \arrow[d, "G_0"] \\
\mathbf{E}_0 & \mathbf{E}_1 \arrow[l, "s"'] \arrow[r, "t"] & \mathbf{E}_0
\end{tikzcd} \tag{17}$$

$$\begin{tikzcd}
\mathbf{C}_2 \arrow[r, "F_2"] \arrow[d, "\circ_{\mathbf{C}}"'] & \mathbf{D}_2 \arrow[r, "G_2"] \arrow[d, "\circ_{\mathbf{D}}"] & \mathbf{E}_2 \arrow[d, "\circ_{\mathbf{E}}"] \\
\mathbf{C}_1 \arrow[r, "F_1"'] & \mathbf{D}_1 \arrow[r, "G_1"'] & \mathbf{E}_1
\end{tikzcd} \tag{18}
\qquad
\begin{tikzcd}
\mathbf{C}_0 \arrow[r, "F_0"] \arrow[d, "u_{\mathbf{C}}"'] & \mathbf{D}_0 \arrow[r, "G_0"] \arrow[d, "u_{\mathbf{D}}"] & \mathbf{E}_0 \arrow[d, "u_{\mathbf{E}}"] \\
\mathbf{C}_1 \arrow[r, "F_1"'] & \mathbf{D}_1 \arrow[r, "G_1"'] & \mathbf{E}_1
\end{tikzcd} \tag{19}$$

To finish the proof, you need to convince yourself that combining commutative diagrams in this way yields commutative diagrams. We proceed with a proof by example. Take diagram (19), we know the left and right square are commutative because F and G are functors. To show that the rectangle also commutes, we need to show the top path and bottom path from $\mathbf{C}_0$ to $\mathbf{E}_1$ compose to the same function. Here is the derivation:[111]

$$\begin{aligned} G_1 \circ F_1 \circ u_{\mathbf{C}} &= G_1 \circ u_{\mathbf{D}} \circ F_0 && \text{left square commutes} \\ &= u_{\mathbf{E}} \circ G_0 \circ F_0 && \text{right square commutes} \qquad \square \end{aligned}$$

[111] In this case, both the diagram and the derivation are fairly simple. This will not stay true in the rest of the book, but the complexity of diagrams will grow way slower than the complexity of derivations, and we will mostly omit the latter for this reason.

The category **Cat** is a concrete category. Intuitively, it is because categories are sets with extra sturcture that functors preserve. Rigorously, there is a forgetful functor $\mathbf{Cat} \rightsquigarrow \mathbf{Set}$.

OL **Exercise B.39** (NOW!)**.** Show that both assignments $\mathbf{C} \mapsto \mathbf{C}_0$ and $\mathbf{C} \mapsto \mathbf{C}_1$ yield functors $\mathbf{Cat} \rightsquigarrow \mathbf{Set}$.[112] Their action on morphism of categories (functors) is straightforward: the first sends F to F_0 and the second sends F to F_1. Show that the functor $(-)_0$ is not faithful, but $(-)_1$ is.

[112] Recall that we assumed $\mathbf{C} \in \mathbf{Cat}$ is small, meaning both $\mathbf{C}_0$ and $\mathbf{C}_1$ are sets.

We will often use $-$ as a **placeholder** for an input so the latter remains nameless. For instance, $f(-,-)$ means f takes two inputs. The type of the inputs and outputs will be made clear in the context.

This last exercise suggests we should view a category as a set of morphisms with extra structure. However, Definition B.4 reveals we can also see a category as a directed graph with extra structure. We can make this formal by first defining the category **DGph** whose objects are small directed graphs and morphisms are functors without the requirement of (15) and (16).[113] There is a functor $\mathbf{Cat} \rightsquigarrow \mathbf{DGph}$ that simply forgets about composition and identities.

[113] Explicitly, a morphism $G \to G'$ is a pair of functions $F_0 : G_0 \to G'_0$ and $F_1 : G_1 \to G'_1$ satisfying for any $f \in G_1$, $F_0(s(f)) = s(F_1(f))$ and $F_0(t(f)) = t(F_1(f))$. Less cryptically, it is a mapping from objects to objects and arrows to arrows such that an arrow $A \to B$ is mapped to an arrow $F_0A \to F_0B$.

Since functors are also a new structure, one might expect that there are transformations between functors that preserve it. It is indeed the case, they are called natural transformations and they are the main subject of Chapter **??**. Moreover, although we will not cover it, there is a whole tower of abstraction that one could build in this way, and it is the subject of study of higher category theory.

B.3 Diagram Paving

If you are in awe at how wonderful the diagrammatic proof of Proposition B.38, this section is for you. We introduce the proof technique called **diagram paving**[114] and set up some exercises for practice.

[114] Usually, diagram paving refers to a more general version of what I will show you. That technique is used in higher category theory.

The key idea in that proof is that combining commutative diagrams yields commutative diagrams.[115] In general, paving a diagram that we want to show commutes is the process of progressively adding more objects and morphisms to obtain multiple diagrams we know (by hypothesis or previous lemmas) commute that combine into the original one.

[115] The term "combining" is not precisely defined, our intuition of what it means should be enough.

Let us clarify by example. In the setting of Proposition B.38, to show that $G \circ F$ is a functor, we need to prove (14) instantiated with $G \circ F$ is commutative.[116] It is drawn in (20).

[116] We only do the first diagram.

$$\begin{tikzcd} \mathbf{C}_0 \arrow[d, "(G\circ F)_0"'] & \mathbf{C}_1 \arrow[l, "s"'] \arrow[r, "t"] \arrow[d, "(G\circ F)_1"] & \mathbf{C}_0 \arrow[d, "(G\circ F)_0"] \\ \mathbf{E}_0 & \mathbf{E}_0 \arrow[l, "s"] \arrow[r, "t"'] & \mathbf{E}_0 \end{tikzcd} \tag{20}$$

We can factor the action of $G \circ F$ and draw (21). We indicated with $\circlearrowright$ that some parts of the diagram are known to commute (by definition of $G \circ F$).[117]

$$\begin{tikzcd}
\mathbf{C}_0 \arrow[d, "F_0"'] \arrow[dd, bend right=60, "(G\circ F)_0"'] & \mathbf{C}_1 \arrow[l, "s"'] \arrow[r, "t"] \arrow[d, "F_1"] & \mathbf{C}_0 \arrow[d, "F_0"] \arrow[dd, bend left=60, "(G\circ F)_0"] \\
\mathbf{D}_0 \arrow[d, "G_0"'] & \mathbf{D}_1 \arrow[d, "G_1"] & \mathbf{D}_0 \arrow[d, "G_0"] \\
\mathbf{E}_0 & \mathbf{E}_0 \arrow[l, "s"] \arrow[r, "t"'] & \mathbf{E}_0
\end{tikzcd} \tag{21}$$

Then we can decompose the two rectangles into four squares that all commute by hypothesis that F and G are functors.

$$\begin{tikzcd}
\mathbf{C}_0 \arrow[d, "F_0"'] \arrow[dd, bend right=60, "(G\circ F)_0"'] & \mathbf{C}_1 \arrow[l, "s"'] \arrow[r, "t"] \arrow[d, "F_1"] \arrow[dl, phantom, "\circlearrowright"] \arrow[dr, phantom, "\circlearrowright"] & \mathbf{C}_0 \arrow[d, "F_0"] \arrow[dd, bend left=60, "(G\circ F)_0"] \\
\mathbf{D}_0 \arrow[d, "G_0"'] & \mathbf{D}_1 \arrow[l, "s"'] \arrow[r, "t"] \arrow[d, "G_1"] \arrow[dl, phantom, "\circlearrowright"] \arrow[dr, phantom, "\circlearrowright"] & \mathbf{D}_0 \arrow[d, "G_0"] \\
\mathbf{E}_0 & \mathbf{E}_0 \arrow[l, "s"] \arrow[r, "t"'] & \mathbf{E}_0
\end{tikzcd} \tag{22}$$

Finally, we recognize that all the commutative diagrams in (22) combine into (20), so the latter is commutative.

From now on, when doing proofs by paving a diagram, we will only show the last paved diagram. Instead of $\circlearrowright$, we will use letters to indicate regions that commute so we can refer to each region in the text and explain why they commute.

There is one last thing we want to mention to end this chapter. We gave two central definitions, categories and functors, and we presented several examples of each. By defining products, we give you access to an unlimited amount of new categories and functors you can construct from known ones.[118]

Definition B.40 (Product category)**.** Let $\mathbf{C}$ and $\mathbf{D}$ be two categories, the **product** of $\mathbf{C}$ and $\mathbf{D}$, denoted by $\mathbf{C} \times \mathbf{D}$, is the category whose objects are pairs of objects in $\mathbf{C}_0 \times \mathbf{D}_0$ and for any two pairs $(X, Y), (X', Y') \in (\mathbf{C} \times \mathbf{D})_0$,[119]

$$\mathrm{Hom}_{\mathbf{C}\times\mathbf{D}}((X,Y),(X',Y')) := \mathrm{Hom}_{\mathbf{C}}(X,X') \times \mathrm{Hom}_{\mathbf{D}}(Y,Y').$$

The identity morphisms and the composition are defined componentwise. Explicitly, for all $X \in \mathbf{C}_0$ and $Y \in \mathbf{D}_0$, $\mathrm{id}_{(X,Y)} = (\mathrm{id}_X, \mathrm{id}_Y)$, and for all $(f, f') \in \mathbf{C}_2$ and $(g, g') \in \mathbf{D}_2$, $(f, g) \circ (f', g') = (f \circ f', g \circ g')$.[120]

OL **Exercise B.41** (NOW!)**.** Verify that the category depicted in (9) is appropriately denoted by $\mathbf{2} \times \mathbf{2}$, i.e., that it is the product category formed with $\mathbf{C} = \mathbf{D} = \mathbf{2}$.

OL **Exercise B.42.** Show that the assignment $\Delta_{\mathbf{C}} : \mathbf{C} \rightsquigarrow \mathbf{C} \times \mathbf{C} = X \mapsto (X, X)$ is functorial, i.e., give its action on morphisms and show it satisfies the relevant axioms. We call $\Delta_{\mathbf{C}}$ the **diagonal functor**.

[117] We did not leave the arrow $(G \circ F)_1$ because it would make the diagram messy.

[118] This is akin to products of groups, direct sums of vector spaces, etc. In Chapter D, we will see how all of these constructions are instances of a more general construction called (categorical) product.

[119] Explicitly, a morphism $(X, Y) \to (X', Y')$ is a pair of morphisms $X \to X'$ and $Y \to Y'$.

[120] We leave you to check that this defines the composition for all of $(\mathbf{C} \times \mathbf{D})_2$. Namely, if (f, g) and (f', g') are composable, then (f, f') and (g, g') are composable.

Definition B.43 (Product functor)**.** Let $F : \mathbf{C} \rightsquigarrow \mathbf{C}'$ and $G : \mathbf{D} \rightsquigarrow \mathbf{D}'$ be two functors, the **product** of F and G, denoted $F \times G : \mathbf{C} \times \mathbf{D} \rightsquigarrow \mathbf{C}' \times \mathbf{D}'$, is defined componentwise on objects and morphisms, i.e., for any $(X, Y) \in (\mathbf{C} \times \mathbf{D})_0$ and $(f, g) \in (\mathbf{C} \times \mathbf{D})_1$,

$$(F \times G)(X, Y) = (FX, GY) \text{ and } (F \times G)(f, g) = (Ff, Gg).$$

Let us check this defines a functor.

i. By definition of $\mathbf{C}' \times \mathbf{D}'$, (Ff, Gg) is a morphism from (FX, GY) to (FX', GY').

ii. For $(f, f') \in \mathbf{C}_2$ and $(g, g') \in \mathbf{D}_2$, we have

$$\begin{aligned}(F \times G)((f, g) \circ (f', g')) &= (F \times G)(f \circ f', g \circ g')\\ &= (F(f \circ f'), G(g \circ g'))\\ &= (Ff \circ Ff', Gg \circ Gg')\\ &= (Ff, Gg) \circ (Ff', Gg')\\ &= (F \times G)(f, g) \circ (F \times G)(f', g').\end{aligned}$$

iii. Since F and G preserve identity morphisms, we have

$$(F \times G)(\mathrm{id}_{(X,Y)}) = (F \times G)(\mathrm{id}_X, \mathrm{id}_Y) = (F\mathrm{id}_X, G\mathrm{id}_Y) = (\mathrm{id}_{FX}, \mathrm{id}_{GY}) = \mathrm{id}_{(FX,GY)}.$$

OL **Exercise B.44** (NOW!)**.** Let $F : \mathbf{C} \times \mathbf{C}' \rightsquigarrow \mathbf{D}$ be a functor. For $X \in \mathbf{C}_0$, we define $F(X, -) : \mathbf{C}' \rightsquigarrow \mathbf{D}$ on objects by $Y \mapsto F(X, Y)$ and on morphisms by $g \mapsto F(\mathrm{id}_X, g)$. Show that $F(X, -)$ is a functor. Define $F(-, Y)$ similarly.

OL **Exercise B.45.** Let $F : \mathbf{C} \times \mathbf{C}' \to \mathbf{D}$ be an action defined on objects and morphisms satisfying

$$F(f, g) = F(f, \mathrm{id}_{t(g)}) \circ F(\mathrm{id}_{s(f)}, g) = F(\mathrm{id}_{t(f)}, g) \circ F(f, \mathrm{id}_{s(g)}).$$

Show that if for any $X \in \mathbf{C}_0$ and $Y \in \mathbf{C}'_0$, $F(X, -)$ and $F(-, Y)$ as defined above are functors, then F is a functor. In other words, the functoriality of F can be proven componentwise.

In the next chapters, we will present other interesting constructions of categories, but we can stop here for now.

C Duality

The concept of duality is ubiquitous throughout mathematics. It can relate two perspectives of the same object as for dual vector spaces, two complementary optimization problems such as a maximization and a minimization linear program, and even two seemingly unrelated subjects like topology and logic (Stone duality). While this vague principle of duality is behind many groundbreaking results, the duality in question here is categorical duality and it is a bit more precise.

Informally, there is nothing more to say than "Take all the diagrams in a definition/theorem, reverse the arrows and reap the benefits of the dual concept/result."[121] The more formal version will follow after we first exhibit the principle in action.

[121] In my opinion, this is already a very good reason to learn category theory because we can basically get twice as much math as before by framing things with a categorical language.

Recall that, intuitively, a functor is a structure-preserving transformation between categories. A simple example we have seen is functors between posets that are order-preserving functions. However, as a consequence, one might conclude that order-reversing functions impair the structure of a poset, which feels arbitrary. The same happens between deloopings of groups because anti-homomorphisms[122] do not arise as functors between such categories.

[122] An **anti-homomorphism** $f : G \to H$ is a function satisfying $f(gg') = f(g')f(g)$ and $f(1_G) = f(1_H)$.

For a more concrete situation, recall the powerset functor $\mathcal{P}$ described in Example B.34.1. It assigns to any set X the powerset $\mathcal{P}X$, and to any function $f : X \to Y$ the image function $\mathcal{P}(f) : \mathcal{P}X \to \mathcal{P}Y$. There is another important function associated to f between powersets: the inverse image f^{-1} that assigns to $S \subset Y$ the set of points in X whose images are in S. Unfortunately, f^{-1} goes in the "wrong" direction $\mathcal{P}Y \to \mathcal{P}X$.

This is quite unsatisfactory because the assignment $f \mapsto f^{-1}$ is well-behaved, e.g. we have $\mathrm{id}_X^{-1} = \mathrm{id}_{\mathcal{P}X}$ for any set X and, for any functions $f : X \to Y$ and $g : Y \to X$, $(f \circ g)^{-1} = g^{-1} \circ f^{-1}$. This second equation looks just like the second condition on functors reversed. In words, taking the inverse image *preserves* composition but in reverse.

It seems arbitrary to distinguish between both options. There are two ways to remedy this discrepancy between intuition and formalism; both have duality as an underlying principle. In this chapter, we will describe both ways, dismiss one of them, and showcase the strength of duality while exploring more basic category theory.

C.1 Contravariant Functors

By modifying Definition B.25 to require that $F(f)$ goes in the opposite direction, we obtain a contravariant functor. Incidentally, what we defined as a functor before is also called a **covariant** functor.

Definition C.1 (Contravariant functor)**.** Let $\mathbf{C}$ and $\mathbf{D}$ be categories, a **contravariant functor** $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ is a pair of maps $F_0 : \mathbf{C}_0 \to \mathbf{D}_0$ and $F_1 : \mathbf{C}_1 \to \mathbf{D}_1$ making diagrams (23), (24) and (25) commute.[123]

[123] Where F_2' is now induced by the definition of F_1 with $(f,g) \mapsto (F_1(g), F_1(f))$.

$$\begin{tikzcd} \mathbf{C}_0 \arrow[d, "F_0"'] & \mathbf{C}_1 \arrow[l, "s"'] \arrow[r, "t"] \arrow[d, "F_1"] & \mathbf{C}_0 \arrow[d, "F_0"] \\ \mathbf{D}_0 & \mathbf{D}_1 \arrow[l, "t"] \arrow[r, "s"'] & \mathbf{D}_0 \end{tikzcd} \tag{23}$$

$$\begin{tikzcd} \mathbf{C}_2 \arrow[r, "F_2'"] \arrow[d, "\circ_{\mathbf{C}}"'] & \mathbf{D}_2 \arrow[d, "\circ_{\mathbf{D}}"] \\ \mathbf{C}_1 \arrow[r, "F_1"'] & \mathbf{D}_1 \end{tikzcd} \tag{24}$$

$$\begin{tikzcd} \mathbf{C}_0 \arrow[r, "F_0"] \arrow[d, "u_{\mathbf{C}}"'] & \mathbf{D}_0 \arrow[d, "u_{\mathbf{D}}"] \\ \mathbf{C}_1 \arrow[r, "F_1"'] & \mathbf{D}_1 \end{tikzcd} \tag{25}$$

In words, F must satisfy the following properties.

i. For any $A, B \in \mathbf{C}_0$, if $f \in \mathrm{Hom}_{\mathbf{C}}(A,B)$ then $F(f) \in \mathrm{Hom}_{\mathbf{D}}(F(B), F(A))$.

ii. If $f, g \in \mathbf{C}_1$ are composable, then $F(f \circ g) = F(g) \circ F(f)$.

iii. If $A \in \mathbf{C}_0$, then $u_{\mathbf{D}}(F(A)) = F(u_{\mathbf{C}}(A))$.

Examples C.2. Just like their covariant counterparts, contravariant functors are quite numerous. Here are a couple of simple ones.

1. Contravariant functors $F : (X, \leq) \rightsquigarrow (Y, \sqsubseteq)$ correspond to order-reversing functions between the posets X and Y and contravariant functors $F : \mathbf{B}G \rightsquigarrow \mathbf{B}H$ correspond to anti-homomorphisms between the groups G and H.

2. The **contravariant powerset functor** $2^{-} : \mathbf{Set} \rightsquigarrow \mathbf{Set}$ sends a set X to its powerset 2^X,[124] and a function $f : X \to Y$ to the preimage map $2^f : 2^Y \to 2^X$, the latter sends a subset $S \subseteq Y$ to
$$2^f(S) = f^{-1}(S) := \{x \in X \mid f(x) \in S\} \subseteq X.$$

[124] We use a different notation for the powerset to emphasize the difference between $\mathcal{P}$ and 2^{-}.

Next, there is a couple of functors that are key to understand the philosophy put forward by category theory.[125]

[125] We will talk more about it when covering the Yoneda lemma in Chapter **??**.

Example C.3 (Hom functors)**.** Let $\mathbf{C}$ be a locally small category and $A \in \mathbf{C}_0$ one of its objects.[126] We define the covariant and contravariant **Hom functors** from $\mathbf{C}$ to **Set**.

[126] We need local smallness so that each $\mathrm{Hom}_{\mathbf{C}}(A,B)$ is a set and the functors land in **Set**.

1. The covariant Hom functor $\mathrm{Hom}_{\mathbf{C}}(A, -) : \mathbf{C} \rightsquigarrow \mathbf{Set}$ sends an object $B \in \mathbf{C}_0$ to the hom-set $\mathrm{Hom}_{\mathbf{C}}(A,B)$ and a morphism $f : B \to B'$ to the function
$$\mathrm{Hom}_{\mathbf{C}}(A,f) : \mathrm{Hom}_{\mathbf{C}}(A,B) \to \mathrm{Hom}_{\mathbf{C}}(A,B') = g \mapsto f \circ g.$$

This function is called **post-composition by** f and is denoted $f \circ (-)$.[127] Let us show $\mathrm{Hom}_{\mathbf{C}}(A, -)$ is a covariant functor.

[127] Some authors denote $f \circ (-)$ as f^*, we prefer to keep this notation for later (see pullbacks).

i. For any $f \in \mathbf{C}_1$, it is clear from the definition that

$$\mathrm{Hom}_{\mathbf{C}}(A, s(f)) = s(f \circ (-)) \text{ and } \mathrm{Hom}_{\mathbf{C}}(A, t(f)) = t(f \circ (-)).$$

ii. For any $(f_1, f_2) \in \mathbf{C}_2$, we claim that

$$\mathrm{Hom}_{\mathbf{C}}(A, f_1 \circ f_2) = \mathrm{Hom}_{\mathbf{C}}(A, f_1) \circ \mathrm{Hom}_{\mathbf{C}}(A, f_2).$$

In the L.H.S., an element $g \in \mathrm{Hom}_{\mathbf{C}}(A, s(f_1 \circ f_2))$ is mapped to $(f_1 \circ f_2) \circ g$ and in the R.H.S., an element $g \in \mathrm{Hom}_{\mathbf{C}}(A, s(f_2))$ is mapped to $f_1 \circ (f_2 \circ g)$. Since $s(f_1 \circ f_2) = s(f_2)$ and composition is associative, we conclude that the two maps are the same.

iii. For any $B \in \mathbf{C}_0$, the post-composition by $u_{\mathbf{C}}(B)$ is defined to be the identity,[128] hence (16) also commutes.

[128] Namely, for any $f : A \to B$, $u_{\mathbf{C}}(B) \circ f = f$.

2. The contravariant Hom functor $\mathrm{Hom}_{\mathbf{C}}(-, A) : \mathbf{C} \rightsquigarrow \mathbf{Set}$ sends an object $B \in \mathbf{C}_0$ to the hom-set $\mathrm{Hom}_{\mathbf{C}}(B, A)$ and a morphism $f : B \to B'$ to the function

$$\mathrm{Hom}_{\mathbf{C}}(f, A) : \mathrm{Hom}_{\mathbf{C}}(B', A) \to \mathrm{Hom}_{\mathbf{C}}(B, A) = g \mapsto g \circ f.$$

This function is called **pre-composition by** f and is denoted $(-) \circ f$.[129] Let us show $\mathrm{Hom}_{\mathbf{C}}(-, A)$ is a contravariant functor.

[129] Some authors denote $(-) \circ f$ as f_*, we prefer to keep this notation for later (see pushouts).

i. For any $f \in \mathbf{C}_1$, it is clear from the definition that

$$\mathrm{Hom}_{\mathbf{C}}(s(f), A) = t((-) \circ f)) \text{ and } \mathrm{Hom}_{\mathbf{C}}(t(f), A) = s((-) \circ f).$$

ii. For any $(f_1, f_2) \in \mathbf{C}_2$, we claim that

$$\mathrm{Hom}_{\mathbf{C}}(f_1 \circ f_2, A) = \mathrm{Hom}_{\mathbf{C}}(f_2, A) \circ \mathrm{Hom}_{\mathbf{C}}(f_1, A).$$

In the L.H.S., an element $g \in \mathrm{Hom}_{\mathbf{C}}(t(f_1 \circ f_2), A)$ is mapped to $g \circ (f_1 \circ f_2)$ and in the R.H.S., an element $g \in \mathrm{Hom}_{\mathbf{C}}(t(f_1), A)$ is mapped to $(g \circ f_1) \circ f_2$. Since $t(f_1 \circ f_2) = t(f_1)$ and composition is associative, we conclude that the two maps are the same.

iii. For any $B \in \mathbf{C}_0$, pre-composition by $u_{\mathbf{C}}(B)$ is defined to be the identity,[130] hence (25) also commutes.

[130] Namely, for any $f : B \to A$, $f \circ u_{\mathbf{C}}(B) = f$.

It can take a bit of time to get comfortable with Hom functors. For now, we will give only one example of each kind (covariant and contravariant), but we will take more time to play with them later in the book.

Example C.4 (Ring of functions)**.**

Example C.5 (Dual vector space)**.** In the category $\mathbf{Vect}_k$, there is a special object k,[131] let us see what the contravariant functor $\mathrm{Hom}_{\mathbf{Vect}_k}(-,k)$ does. It assigns to any vector space V the set of linear maps $V \to k$, that is, the carrier set of the dual space V^*. It assigns to linear maps $T : V \to W$, the function

$$\mathrm{Hom}_{\mathbf{Vect}_k}(W,k) \to \mathrm{Hom}_{\mathbf{Vect}_k}(V,k) = \phi \mapsto \phi \circ T.$$

We know that $\mathrm{Hom}_{\mathbf{Vect}_k}(V,k) = V^*$ can be seen as a vector space and it is easy to check that pre-composition by T is a linear map $W^* \to V^*$. Therefore, we find that the assignment $V \mapsto V^* = \mathrm{Hom}_{\mathbf{Vect}_k}(-,k)$ is a contravariant functor $\mathbf{Vect}_k \rightsquigarrow \mathbf{Vect}_k$.

We will not dwell too long on contravariant functors as we will see right away how they can be avoided, but first, let us give a reason why we want to avoid them.

OL **Exercise C.6.** Let $F : \mathbf{C} \rightsquigarrow \mathbf{D}$, $G : \mathbf{D} \rightsquigarrow \mathbf{E}$ be contravariant functors, and $G \circ F : \mathbf{C} \rightsquigarrow \mathbf{E}$ be their composition defined by $G_0 \circ F_0$ on objects and $G_1 \circ F_1$ on morphisms. Show that $G \circ F$ is a *covariant* functor.[132] Using diagrams will be easier.

C.2 Opposite Category

Another way to deal with order-reversing maps $(X, \leq) \to (Y, \subseteq)$ is to consider the reverse order on X and a covariant functor $(X, \geq) \rightsquigarrow (Y, \subseteq)$. This also works for anti-homomorphisms by constructing the opposite group G^{op} in which the operation is reversed, namely $g \cdot^{\mathrm{op}} h = hg$. The opposite category is a generalization of these constructions.

Definition C.7 (Opposite category)**.** Let $\mathbf{C}$ be a category, we denote the **opposite** category with $\mathbf{C}^{\mathrm{op}}$ and define it by[133]

$$\mathbf{C}_0^{\mathrm{op}} = \mathbf{C}_0,\ \mathbf{C}_1^{\mathrm{op}} = \mathbf{C}_1,\ s^{\mathrm{op}} = t,\ t^{\mathrm{op}} = s,\ u_{\mathbf{C}^{\mathrm{op}}} = u_{\mathbf{C}}$$

with the composition defined by $f^{\mathrm{op}} \circ^{\mathrm{op}} g^{\mathrm{op}} = (g \circ f)^{\mathrm{op}}$.[134] This naturally leads to the following contravariant functor $(-)^{\mathrm{op}}_{\mathbf{C}} : \mathbf{C} \rightsquigarrow \mathbf{C}^{\mathrm{op}}$ which sends an object A to A^{op} and a morphism f to f^{op}. It is called the **opposite functor**.

With this definition, one can see contravariant functors as covariant functors. Formally, let $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ be a contravariant functor, we can view F as covariant functor from $\mathbf{C}^{\mathrm{op}}$ to $\mathbf{D}$ or from $\mathbf{C}$ to $\mathbf{D}^{\mathrm{op}}$ via the compositions $F \circ (-)^{\mathrm{op}}_{\mathbf{C}^{\mathrm{op}}}$ and $(-)^{\mathrm{op}}_{D} \circ F$ respectively.[135]

In the rest of this book, we choose to work with covariant functors of type $\mathbf{C}^{\mathrm{op}} \rightsquigarrow \mathbf{D}$ instead of contravariant functors $\mathbf{C} \rightsquigarrow \mathbf{D}$,[136] and functors will be covariant by default.

Examples C.8. 1. As hinted at before, the category corresponding to $(X, \geq)$ is the opposite category of $(X, \leq)$ and $(\mathbf{B}G)^{\mathrm{op}}$ is the category corresponding to the opposite group of G, i.e.: $(\mathbf{B}G)^{\mathrm{op}} = \mathbf{B}(G^{\mathrm{op}})$.

2. We have seen that functors $\mathbf{B}G \rightsquigarrow \mathbf{Set}$ correspond to left actions of a group G. You can check that functors $\mathbf{B}G^{\mathrm{op}} \rightsquigarrow \mathbf{Set}$ correspond to right actions of G.

[131] We know it is special because we know some linear algebra, but k also has some interesting categorical properties (see Exercise C.40).

[132] We conclude that we cannot straightforwardly compose contravariant functors. This alone makes the following alternative more desirable because we want functors to be morphisms in a category, hence they must be composable.

[133] Intuitively, we reverse the direction of all morphisms in $\mathbf{C}$ and reverse the order of composition as well.

[134] Note that the $-^{\mathrm{op}}$ notation here is just used to distinguish elements in $\mathbf{C}$ and $\mathbf{C}^{\mathrm{op}}$ but the collection of objects and morphisms are the same.

[135] Recall from Exercise C.6 that these compositions are covariant.

[136] We still had to learn about contravariant functors because you might encounter them in the wild.

3. The two Hom functors defined in Example C.3 are now written

$$\mathrm{Hom}_{\mathbf{C}}(A,-) : \mathbf{C} \rightsquigarrow \mathbf{Set} \text{ and } \mathrm{Hom}_{\mathbf{C}}(-,A) : \mathbf{C}^{\mathrm{op}} \rightsquigarrow \mathbf{Set}.$$

By Exercise B.45, they can be combined into a functor

$$\mathrm{Hom}_{\mathbf{C}}(-,-) : \mathbf{C}^{\mathrm{op}} \times \mathbf{C} \rightsquigarrow \mathbf{Set}$$

acting on objects as $(A,B) \mapsto \mathrm{Hom}_{\mathbf{C}}(A,B)$ and on morphisms as $(f,g) \mapsto (g \circ - \circ f)$. The condition in Exercise B.45 is satisfied because[137]

$$\begin{aligned}\mathrm{Hom}_{\mathbf{C}}(f,g) &= g \circ - \circ f \\ &= \mathrm{id}_{t(g)} \circ (g \circ - \circ \mathrm{id}_{t(f)}) \circ f = \mathrm{Hom}_{\mathbf{C}}(f, \mathrm{id}_{t(g)}) \circ \mathrm{Hom}_{\mathbf{C}}(\mathrm{id}_{t(f)}, g) \\ &= g \circ (\mathrm{id}_{s(g)} \circ - \circ f) \circ \mathrm{id}_{s(f)} = \mathrm{Hom}_{\mathbf{C}}(\mathrm{id}_{s(f)}, g) \circ \mathrm{Hom}_{\mathbf{C}}(f, \mathrm{id}_{s(g)}).\end{aligned}$$

This will be called the Hom **bifunctor**.

OL **Exercise C.9.** Let $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ be a functor, show that its dual F^{op} defined by $A^{\mathrm{op}} \mapsto (FA)^{\mathrm{op}}$ on objects and $f^{\mathrm{op}} \mapsto (Ff)^{\mathrm{op}}$ on morphisms is a functor $\mathbf{C}^{\mathrm{op}} \rightsquigarrow \mathbf{D}^{\mathrm{op}}$.

Remark C.10. It is sometimes useful to compose the Hom bifunctor with other functors as follows. Given two functors $F, G : \mathbf{C} \rightsquigarrow \mathbf{D}$, there is a functor $\mathrm{Hom}_{\mathbf{D}}(F-, G-) : \mathbf{C}^{\mathrm{op}} \times \mathbf{C} \rightsquigarrow \mathbf{D}$ acting on objects by $(X,Y) \mapsto \mathrm{Hom}_{\mathbf{D}}(FX, GY)$ and on morphisms by $(f,g) \mapsto Gg \circ (-) \circ Ff$. One can check functoriality by showing

$$\mathrm{Hom}_{\mathbf{D}}(F-, G-) = \mathrm{Hom}_{\mathbf{D}}(-,-) \circ (F^{\mathrm{op}} \times G).$$

C.3 Duality in Action

Let us start to illustrate how duality can be useful while covering important definitions and results.

Definition C.11 (Monomorphism)**.** Let $\mathbf{C}$ be a category, a morphism $f \in \mathbf{C}_1$ is said to be **monic** (or a **monomorphism**) if for any parallel morphisms g and h such that $(f,g), (f,h) \in \mathbf{C}_2$, $f \circ g = f \circ h$ implies $g = h$. Equivalently, f is monic if $g = h$ whenever the following diagram commutes.[138]

$$\begin{tikzcd} \bullet \arrow[r, bend left, "g"] \arrow[r, bend right, "h"'] & \bullet \arrow[r, "f"] & \bullet \end{tikzcd} \tag{26}$$

Standard notation for a monomorphism is $\bullet \rightarrowtail \bullet$ (`\rightarrowtail`).[139]

Proposition C.12. *Let* $\mathbf{C}$ *be a category and* $f : A \to B$ *a morphism, if there exists* $f' : B \to A$ *such that* $f' \circ f = \mathrm{id}_A$,[140] *then* f *is a monomorphism.*

Proof. If $f \circ g = f \circ h$, then $f' \circ f \circ g = f' \circ f \circ h$ implying $g = h$. ☐

Not all monomorphisms have a left inverse, those that do are called **split monomorphisms**.

[137] Looking at where the source and target functions are applied, these equalities do not match exactly what is in Exercise B.45 since $\mathrm{Hom}_{\mathbf{C}}(-,-)$ is contravariant in the first component.

[138] According to Definition B.7, this diagram commutes if and only if $f \circ g = f \circ h$ because the paths (f,g) and (f,h) are the only paths of length bigger than one.

[139] Another widespread notation is $\bullet \hookrightarrow \bullet$. I prefer to use the latter when we understand the morphism as an "inclusion" of the first object in the second. These are often monic.

[140] We say that f' is a **left inverse** of f.

Proposition C.13. *Let* $\mathbf{C}$ *be a category and* $(f_1, f_2) \in \mathbf{C}_2$*, if* $f_1 \circ f_2$ *is a monomorphism, then* f_2 *is a monomorphism.*

Proof. Let $g, h \in \mathbf{C}_1$ be such that $f_2 \circ g = f_2 \circ h$, we readily get that $(f_1 \circ f_2) \circ g = (f_1 \circ f_2) \circ h$. Since $f_1 \circ f_2$ is a monomorphism, this implies $g = h$. □

The last two results hint at the fact that monomorphisms are analogues to injective functions and we will see that they are exactly the same in the category **Set**, but first let us introduce the dual concept after the formal definition of duality.

Definition C.14 (Duality)**.** Given a definition or statement in an arbitrary category $\mathbf{C}$, one could view this concept inside the category $\mathbf{C}^{\mathrm{op}}$ and obtain a similar definition or statement where all morphisms and the order of composition are reversed, this is called the **dual** concept. Since $\mathbf{C}^{\mathrm{op}^{\mathrm{op}}} = \mathbf{C}$, taking the dual is an involution, namely, the dual of the dual of a thing is that thing. For a definition or result where multiple *arbitrary* categories are involved, the dual version is obtained by taking the opposite of all categories.[141] It is common but not systematic to refer to a dual notion with the prefix "co" (e.g.: presheaf and copresheaf).

[141] Note the emphasis on the word "arbitrary". For instance, a **presheaf** is a functor $F : \mathbf{C}^{\mathrm{op}} \rightsquigarrow \mathbf{Set}$ and the dual concept is a **copresheaf**, a functor $F : \mathbf{C} \rightsquigarrow \mathbf{Set}$; we did not take the opposite of **Set**.

Dualizing the definition of a monomorphism yields an epimorphism.

Definition C.15 (Epimorphism)**.** Let $\mathbf{C}$ be a category, a morphism $f \in \mathbf{C}_1$ is said to be **epic** (or an **epimorphism**) if for any two parallel morphisms g and h such that $(g, f), (h, f) \in \mathbf{C}_2$, $g \circ f = h \circ f$ implies $g = h$. Equivalently, f is epic if $g = h$ whenever the following diagram commutes.[142]

$$\begin{tikzcd} \bullet \arrow[r, "f"] & \bullet \arrow[r, bend left, "g"] \arrow[r, bend right, "h"'] & \bullet \end{tikzcd} \tag{27}$$

[142] Seeing the diagrams make it clearer that the concepts are dual. Reversing the arrows in (26) yields (27) and vice-versa.

Standard notation for an epimorphism is $\bullet \twoheadrightarrow \bullet$ (`\twoheadrightarrow`).

The dual versions of Propositions C.12 and C.13 also hold. Although translating our previous proofs to the dual case is straightforward, we will do the two next proofs relying on duality to convey the general sketch that works anytime a dual result needs to be proven.

Proposition C.16. *Let* $\mathbf{C}$ *be a category and* $f : A \to B$ *a morphism, if there exists* $f' : B \to A$ *such that* $f \circ f' = \mathrm{id}_B$*, then* f *is epic.*[143]

[143] We say that f' is a **right inverse** of f.

Proof. Observe that f is epic in $\mathbf{C}$ if and only if f^{op} is monic in $\mathbf{C}^{\mathrm{op}}$ (reverse the arrows in the definition).[144] Moreover, by definition,

[144] This is another way to see that two concepts are dual.

$$f'^{\mathrm{op}} \circ f^{\mathrm{op}} = (f \circ f')^{\mathrm{op}} = \mathrm{id}_B{}^{\mathrm{op}} = \mathrm{id}_{B^{\mathrm{op}}},$$

so by the result for monomorphisms, f^{op} is monic and hence f is epic. □

Not all epimorphisms have a right inverse, those that do are called **split epimorphisms**.

Proposition C.17. *Let* $\mathbf{C}$ *be a category and* $(f_1, f_2) \in \mathbf{C}_2$*, if* $f_1 \circ f_2$ *is epic, then* f_1 *is epic.*

Proof. Since ${f_2}^{\mathrm{op}} \circ {f_1}^{\mathrm{op}} = (f_1 \circ f_2)^{\mathrm{op}}$ is monic, the result for monomorphisms implies ${f_1}^{\mathrm{op}}$ is monic and hence f_1 is epic. □

Example C.18 (Set). We mentioned that monomorphisms are like generalizations of injective functions, and you may have guessed that epimorphisms are, in the same sense, generalizations of surjective functions. Let us make this precise.

- A function $f : A \to B$ is a monomorphism in **Set** if and only if it is injective:[145]

  ($\Leftarrow$) Since f is injective, it has a left inverse, so it is monic by Proposition C.12.

  ($\Rightarrow$) Given $a \in A$, let $g_a : \{*\} \to A$ be the function sending $*$ to a. For any $a_1 \neq a_2 \in A$, the functions g_{a_1} and g_{a_2} are different, hence $f \circ g_{a_1} \neq f \circ g_{a_2}$. Therefore, $f(a_1) \neq f(a_2)$ implying f is injective.

- A function $f : A \to B$ is an epimorphism if and only if it is surjective:[146]

  ($\Leftarrow$) Since f is surjective, it has a right inverse, so it is epic by Proposition C.16.

  ($\Rightarrow$) Let $h : B \to \{0, 1\}$ be the constant function at 1 and $g : B \to \{0, 1\}$ be the indicator function of $\mathrm{Im}(f) \subseteq B$, namely,

  $$g(x) = \begin{cases} 1 & \exists a \in A, x = f(a) \\ 0 & \text{otherwise} \end{cases}.$$

  We see that $g \circ f = h \circ f$ are both constant at 1, and f being epic implies $g = h$. Thus, any element of B is in the image of f, that is, f is surjective.

[145] As a consequence, since all injective functions have a left inverse, all the monomorphisms in **Set** are split.

[146] If you assume the axiom of choice, all surjective functions have a right inverse and thus all epimorphisms in **Set** are split.

Example C.19 (Mon). Inside the category **Mon**, the monomorphisms are precisely the injective homomorphisms.

($\Rightarrow$) Let $f : M \to M'$ be an injective homomorphisms and $g_1, g_2 : N \to M$ be two parallel homomorphisms. Suppose that $f \circ g_1 = f \circ g_2$, then for all $x \in N$, $f(g_1(x)) = f(g_2(x))$, so by injectivity of f, $g_1(x) = g_2(x)$. Therefore, $g_1 = g_2$ and since g_1 and g_2 were arbitrary, f is a monomorphism.

($\Leftarrow$) Let $f : M \to M'$ be a monomorphism. Let $x, y \in M$ and define $p_x : (\mathbb{N}, +) \to M$ by $k \mapsto x^k$ and similarly for p_y. It is easy to show that p_x and p_y are homomorphisms.[147] If $f(x) = f(y)$, then, by the homomorphism property, for all $k \in \mathbb{N}$

$$f(p_x(k)) = f(x^k) = f(x)^k = f(y)^k = f(y^k) = f(p_y(k)).$$

In other words, we get $f \circ p_x = f \circ p_y$, so $p_x = p_y$ and $x = y$. This direction follows.

Conversely, an epimorphism is not necessarily surjective. For example, the inclusion homomorphism $i : (\mathbb{N}, +) \to (\mathbb{Z}, +)$ is clearly not surjective, but it is an epimorphism. Indeed, let $g, h : (\mathbb{Z}, +) \to M$ be two monoid homomorphisms satisfying $g \circ i = h \circ i$. In particular, $g(n) = h(n)$ for any $n \in \mathbb{N} \subset \mathbb{Z}$. It remains to show that also $g(-n) = h(-n)$: we have

$$\begin{aligned} h(n)g(-n) = g(n)g(-n) = g(n-n) = g(0) = 1_M \\ h(-n)h(n) = h(-n+n) = h(0) = 1_M, \end{aligned}$$

but then $g(-n) = h(-n)h(n)g(-n) = h(-n)$.

[147] It follows from the definition of x^k which is $x \overset{k}{\cdots} x$.

OL **Exercise C.20.** Show that a monomorphism in **Cat** is a functor that is faithful and injective on objects, it is called an **embedding**.[148]

OL **Exercise C.21.** Show that a morphism $f \in \mathbf{C}_1$ is monic if and only if the function $\mathrm{Hom}_{\mathbf{C}}(A, f) = f \circ -$ is injective for all $A \in \mathbf{C}_0$. Dually, show that f is epic if and only if the function $\mathrm{Hom}_{\mathbf{C}^{\mathrm{op}}}(A^{\mathrm{op}}, f^{\mathrm{op}}) = \mathrm{Hom}_{\mathbf{C}}(f, A) = - \circ f$ is injective for all $A \in \mathbf{C}_0$.

Remark C.22. These alternative definitions of monomorphisms and epimorphisms are more categorical in nature. In fact, in the setting of enriched category theory they are preferable because they generalize easily.

Definition C.23 (Isomorphism)**.** Let **C** be a category, a morphism $f : A \to B$ is said to be an **isomorphism** if there exists a morphism $f^{-1} : B \to A$ such that $f \circ f^{-1} = \mathrm{id}_B$ and $f^{-1} \circ f = \mathrm{id}_A$.[149]

OL **Exercise C.24.** Show that the property of being monic/epic/an isomorphism is invariant under composition, i.e., if f and g are composable monomorphisms, then $f \circ g$ is monic and similarly for epimorphisms and isomorphisms.

Remark C.25. The results shown about monic and epic morphisms[150] imply that any isomorphism is monic and epic. However, the converse is not true as witnessed by the inclusion morphism i described in Example C.19.[151] A category where all monic and epic morphisms are isomorphisms (e.g.: **Set**) is called **balanced**. If there exists an isomorphism between two objects A and B, then they are **isomorphic**, denoted $A \cong B$. Isomorphic objects are also isomorphic in the opposite category,[152] that is, the concept of **isomorphism** is *self-dual*.

For most intents and purposes, we will not distinguish between isomorphic objects in a category because all the properties we care about will hold for one if and only if they hold for the other. This attitude should be somewhat familiar if you have done a bit of abstract algebra because it is natural to substitute the group $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$ for $\mathbb{Z}/6\mathbb{Z}$ or k^n for an n–dimensional vector space over k. It is less natural in **Set** because, for instance, it equates the sets $\{0, 1\}$ and $\{a, b\}$ which may be too coarse-grained for our intuition.

Example C.26 (**Set**)**.** A function $f : X \to Y$ in $\mathbf{Set}_1$ has an inverse f^{-1} if and only if f is bijective, thus isomorphisms in **Set** are bijections. As a consequence, we have $A \cong B$ if and only if $|A| = |B|$.[153]

Example C.27 (**Cat**)**.** An isomorphism in **Cat** is a functor $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ with an inverse $F^{-1} : \mathbf{D} \rightsquigarrow \mathbf{C}$. This implies that F_0 and F_1 are bijections[154] because $(F^{-1})_0$ is the inverse of F_0 and $(F^{-1})_1$ is the inverse of F_1.

Conversely, if $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ is a functor whose component on objects and morphisms are bijective, we check that defining $F^{-1} : \mathbf{D} \rightsquigarrow \mathbf{C}$ with $(F^{-1})_0 := (F_0)^{-1}$ and $(F^{-1})_1 := (F_1)^{-1}$ yields a functor.

i. Let $f \in \mathrm{Hom}_{\mathbf{D}}(A, B)$, by bijectivity of F_0 and F_1, there are $X, Y \in \mathbf{C}_0$ and $g : X \to Y$ such that $FX = A$, $FY = B$ and $Fg = f$. Then, by definition,

$$s(F^{-1}f) = s(g) = X = F^{-1}FX = F^{-1}A, \text{ and}$$

[148] Finding a nice characterization of epimorphisms in **Cat** is an open question as far as I know.

[149] Then f^{-1} is called the **inverse** of f. One can check that if f' is a left inverse of f and f'' is a right inverse, then $f' = f'' = f^{-1}$. Hence, the inverse is unique.

[150] Proposition C.12 and C.16.

[151] This is not akin to the situation in **Set** because, there, all monomorphisms and epimorphisms are split (assuming the axiom of choice).

[152] Because the left inverse becomes the right inverse and vice-versa.

[153] This is in fact the definition of cardinality.

[154] When F_0 is a bijection, F_1 is a bijection if and only if F is fully faithful. Indeed, ...

$$t(F^{-1}f) = t(g) = Y = F^{-1}FY = F^{-1}B.$$

ii. For any $(f, f') \in \mathbf{D}_2$ with $f = Fg$ and $f' = Fg'$, we find

$$F^{-1}(f \circ f') = F^{-1}(Fg \circ Fg') = F^{-1}F(g \circ g') = g \circ g' = F^{-1}Fg \circ F^{-1}Fg' = Ff \circ Ff'.$$

iii. For any $A \in \mathbf{D}_0$ with $A = FX$, we find

$$F^{-1}\mathrm{id}_A = F^{-1}\mathrm{id}_{FX} = F^{-1}F\mathrm{id}_X = \mathrm{id}_X = \mathrm{id}_{F^{-1}FX} = \mathrm{id}_{F^{-1}A}.$$

We can conclude that isomorphisms in **Cat** are precisely the functors which are bijective on objects and morphisms. Furthermore, Footnote 154 implies they are precisely fully faithful functors that are bijective on objects.

Examples C.28 (Concrete categories)**.** In a concrete category **C** with forgetful functor U, the underlying function of an isomorphism f must bijective because $U(f^{-1})$ is the inverse of Uf. This condition may be sufficient or not.

1. It is a simple exercise in an algebra class to show that isomorphisms in the categories **Mon**, **Grp**, **Ring**, **Field** and $\mathbf{Vect}_k$ are simply bijective homomorphisms.[155]

2. In **Poset**, an isomorphism between $(A, \leq_A)$ and $(B, \leq_B)$ is a bijective function $f : A \to B$ satisfying $a \leq_A a' \Leftrightarrow f(a) \leq_B f(a')$. Such a function is clearly monotone, but its inverse is also monotone as for any $b \leq_B b'$, we have $ff^{-1}(b) \leq_B ff^{-1}(b') \implies f^{-1}(b) \leq_A f^{-1}(b')$.

3. In **Top**, it is not enough to have a bijective continuous function, we need to require that it has a continuous inverse.[156] Such functions are called **homeomorphisms**.

Definition C.29 (Initial object)**.** Let **C** be a category, an object $A \in \mathbf{C}_0$ is said to be **initial** if for any $B \in \mathbf{C}_0$, $|\mathrm{Hom}_{\mathbf{C}}(A, B)| = 1$, namely there are no two parallel morphisms with source A and every object has a morphism coming from A. The[157] initial object of a category, if it exists, is denoted $\varnothing$ and the *unique* morphism from $\varnothing$ to $X \in \mathbf{C}_0$ is denoted $[] : \varnothing \to X$.

Definition C.30 (Terminal object)**.** Let **C** be a category, an object $A \in \mathbf{C}_0$ is said to be **terminal**[158] if for any $B \in \mathbf{C}_0$, $|\mathrm{Hom}_{\mathbf{C}}(B, A)| = 1$, namely there are no two parallel morphisms with target A and every object has a morphism going to A. The terminal object of a category, if it exists, is denoted $\mathbf{1}$ and the *unique* morphism from $X \in \mathbf{C}_0$ into $\mathbf{1}$ is denoted $\langle\rangle : X \to \mathbf{1}$.

Remark C.31 (Notation)*.* The motivation behind the notations $\varnothing$ and $\mathbf{1}$ is given shortly, but the notations for the morphisms will be explained in Chapter D.

An object is initial in a category **C** if and only if it is terminal in $\mathbf{C}^{\mathrm{op}}$, so these two concepts are dual. Also, if an object is initial and terminal, we say it is a **zero** object and usually denote it $\mathbf{0}$.[159]

[155] In fact, isomorphisms are commonly defined as bijective homomorphisms in said algebra class.

[156] Consider $X = \{0, 1\}$ with the two extreme topologies $\tau_\perp = \{\varnothing, X\}$ and $\tau_\top = \mathcal{P}X$. The identity map $\mathrm{id}_X : (X, \tau_\top) \to (X, \tau_\perp)$ is clearly bijective and continuous, but its inverse is not continuous. A similar example shows that a bijective monotone function is not necessarily a poset isomorphism.

[157] We will soon see why we can use *the* instead of *an*.

[158] The terminology final is also common.

[159] Clearly, the concept of zero object is self-dual.

Example C.32 (Set). Let X be a set, there is a unique function from the empty set into X, it is the empty function.[160] We deduce that the empty set is the initial object in **Set**, hence the notation $\varnothing$. For the terminal object, we observe that there is a unique function $X \to \{*\}$ sending all elements of X to $*$, thus $\mathbf{1} = \{*\}$ is terminal in **Set**.

[160] Recall (or learn here) that a function $f : A \to B$ is defined via subset of $f \subseteq A \times B$ that satisfies $\forall a \in A, \exists! b \in B, (a, b) \in f$. When A is empty, $A \times B$ is empty and the only subset $\varnothing \subseteq A \times B$ satisfies the condition vacuously. In passing, when B is empty but A is not, the unique subset of $A \times B$ does not satisfy the condition, so there is no function $A \to \varnothing$.

In this example, we could have chosen any singleton to show it is terminal. However, that choice is irrelevant to a good category theorist because just as any two singletons are isomorphic (because they have the same cardinality), any two terminal objects are isomorphic.

Proposition C.33. *Let* $\mathbf{C}$ *be a category and* $A, B \in \mathbf{C}_0$ *be initial, then* $A \cong B$.

Proof. Let f be the single element in $\mathrm{Hom}_{\mathbf{C}}(A, B)$ and f' be the single element in $\mathrm{Hom}_{\mathbf{C}}(B, A)$. Both the identity morphism id_A and $f' \circ f$ belong to $\mathrm{Hom}_{\mathbf{C}}(A, A)$ which must have cardinality 1 because A is initial. Similarly id_B and $f \circ f'$ belong to $\mathrm{Hom}_{\mathbf{C}}(B, B)$ which has cardinalty 1 because B is initial. We conclude that $f' \circ f = \mathrm{id}_A$ and $f \circ f' = \mathrm{id}_B$. In words, f and f' are inverses, thus $A \cong B$. $\square$

Corollary C.34 (Dual)**.** *Let* $\mathbf{C}$ *be a category and* $A, B \in \mathbf{C}_0$ *be terminal, then* $A \cong B$.[161]

[161] From now on, I let you prove many dual results on your own — I will try to continue doing the complicated ones. They are not necessarily great exercises, but you can certainly do them if you want to follow this book at a slower pace.

Rewording the last two results, we can say that initial (resp. terminal) objects are unique up to isomorphisms. However, the situation is quite nicer. Initial (resp. terminal) objects are unique up to *unique* isomorphisms. Indeed, if there is an isomorphism $f : A \to B$ and A and B are initial (resp. terminal), then, by definition, f is the unique morphism in $\mathrm{Hom}_{\mathbf{C}}(A, B)$.

OL **Exercise C.35.** Show that in **Cat**, the initial object is the empty category (no objects and no morphisms) and the terminal object is the category with one object $\mathbf{1}$ (hence the agreeing notation).[162]

[162] **Hint**: the unique functor $\langle\rangle : \mathbf{C} \to \mathbf{1}$ is the constant functor at the object $\bullet \in \mathbf{1}_0$.

Example C.36 (Grp). Similarly to **Set**, the trivial group with one element is terminal in **Grp**. Moreover, note that there are no empty group (because a group must contain an identity element), but any group homomorphism from the trivial group $\{1\}$ into a group G must send 1 to 1_G, which completely determines the homomorphism. Therefore, the trivial group is also initial in **Grp**, it is the zero object.

Example C.37 (Met). The terminal object in **Met** is the space with only one point $*$. The distance is determined by the axioms on a metric, because $d_{\mathbf{1}}(*, *)$ must be equal to 0.[163] The initial object in **Met** is the empty space, for the same reason that $\varnothing$ is initial in **Set**.

[163] The function sending all of X to $*$ is nonexpansive whatever the distance d on X because $d(x, y) \geq 0 = d_{\mathbf{1}}(*, *)$.

Example C.38 (Rel). The category **Rel** has the empty set $\varnothing$ as both its terminal and initial object. Indeed, for any set A, there is a unique possible relation from A to $\varnothing$ and from $\varnothing$ to A because

$$\mathrm{Hom}_{\mathbf{Rel}}(A, \varnothing) = \{\varnothing\} = \mathrm{Hom}_{\mathbf{Rel}}(\varnothing, A).$$

This is because the only subset of $A \times \varnothing = \varnothing = \varnothing \times A$ is the empty subset. Thus, $\varnothing$ is the zero object of **Rel**.

OL Exercise C.39. Find the initial and terminal objects in $\mathbf{Set}_*$.

OL Exercise C.40. Find the initial and terminal objects in $\mathbf{Vect}_k$.

OL Exercise C.41. Find a category with only two objects X and Y such that

(i) X is initial but not terminal and Y is terminal but not initial.

(ii) X is initial but not terminal and Y neither terminal nor initial.

(iii) X is terminal but not initial and Y is neither terminal nor initial.

(iv) X is initial and terminal and Y is neither terminal nor initial.

Examples C.42. Here are more examples of categories where initial and terminal objects may or may not exist.

1. $\exists$ terminal, $\nexists$ initial: Consider the poset $(\mathbb{N}, \geq)$ represented by diagram (28). It is clear that 0 is terminal and no element can be initial because $0 \geq x$ implies $x = 0$.

$$\overset{0}{\bullet} \longleftarrow \overset{1}{\bullet} \longleftarrow \overset{2}{\bullet} \longleftarrow \cdots \qquad (28)$$

2. $\nexists$ terminal, $\exists$ initial:[164] Recall the category **SetInj** of finite sets and injective functions. The empty set is still initial but the singletons are not terminal because a function from a set S into $\{*\}$ is never injective when $|S| > 1$.

[164] Of course, you could take the opposite of $(\mathbb{N}, \geq)$, that is $(\mathbb{N}, \leq)$, but that is not fun.

3. $\nexists$ terminal, $\nexists$ initial: Let G be a non-trivial group, the delooping of G has no terminal and no initial objects. The category $\mathbf{B}G$ has a single object $*$ with $\mathrm{Hom}_{\mathbf{B}G}(*, *) = G$, so $*$ cannot be initial nor terminal when $|G| > 1$.

   For a more interesting example, consider the category **Field**. Its underlying directed graph is disconnected[165] because there are no field homomorphisms between fields of different characteristic. Therefore, **Field** has no initial nor terminal objects.

[165] There are objects with no morphisms between them.

4. $\exists$ terminal, $\exists$ initial: The empty set is both initial and terminal in the category **Rel** because a relation $\emptyset \to A$ (resp. $A \to \emptyset$) is a subset of $\emptyset \times A$ (resp. $A \times \emptyset$), and the latter has a unique subset for all sets A.

   For an example with no zero object, let X be a non-empty topological space where τ is the collection of open sets.[166] The category of open sets $\mathcal{O}(X)$ satisfies

$$\mathrm{Hom}_{\mathcal{O}(X)}(U, V) = \begin{cases} \{i_{U,V}\} & U \subseteq V \\ \emptyset & U \not\subseteq V \end{cases}$$

   Since the empty set is contained in every open set, it is an initial object. Since the full set X contains every open set, it is a terminal object. Any other set cannot be initial as it cannot be contained in $\emptyset$ nor terminal as it cannot contain X. Moreover, note that the two objects are not isomorphic because $X \not\subseteq \emptyset$.

[166] Recall that it must contain $\emptyset$ and X.

OL Exercise C.43. Let $\mathbf{C}$ be a category with a terminal object $\mathbf{1}$. Show that any morphism $f : \mathbf{1} \to X$ is monic. State and prove the dual statement.

OL **Exercise C.44.** Let $\mathbf{C}$ and $\mathbf{D}$ be categories, and $\mathbf{1_C}$ and $\mathbf{1_D}$ be terminal objects in $\mathbf{C}$ and $\mathbf{D}$ respectively. Show that $(\mathbf{1_C}, \mathbf{1_D})$ is terminal in the $\mathbf{C} \times \mathbf{D}$. State and prove the dual statement.

Example C.45. For our last application of duality in this section,[167] let X be a set and consider the posetal category $(\mathcal{P}X, \subseteq)$. We would like to define the union of two subsets of X in this category. The usual definition $A \cup B = \{x \in X \mid x \in A \text{ or } x \in B\}$ is not suitable because the data in the posetal category $\mathcal{P}X$ never refers to elements of X. In particular, the subsets $A, B \subseteq X$ are simply objects in the category and it is not clear to us how we can determine what elements are in A and B with our categorical tools (objects and morphisms).

[167] Do not worry, we will have plenty of opportunities to use duality later.

We propose another characterization of the union of A and B. First, what is obvious, $A \cup B$ contains A and it contains B. Second, $A \cup B$ is the smallest subset of X containing A and B. Indeed, if $Y \subseteq X$ contains all elements in A and B, then it also contains $A \cup B$. Using the order $\subseteq$ (or equivalently, the morphisms in the category $\mathcal{P}X$), we have[168]

$$A, B \subseteq A \cup B \text{ and } \forall Y \text{ s.t. } A, B \subseteq Y \text{ then } A \cup B \subseteq Y.$$

[168] We leave it as an exercise to show that $A \cup B$ is the only subset of X satisfying this property.

This yields a definition of $\cup$ within the category $\mathcal{P}X$, which means we can look in the opposite of $\mathcal{P}X$ and dualize $\cup$.

The dual of this property (reversing all inclusions) is as follows.[169]

$$A \square B \subseteq A, B \text{ and } \forall Y \text{ s.t. } Y \subseteq A, B \text{ then } Y \subseteq A \square B$$

[169] The symbol $\square$ is a placeholder for the operation which we will find to be dual to union.

Putting this in words, $A \square B$ is the largest subset of X which is contained in A and B. That is, of course, the intersection $A \cap B$. In this sense, union and intersection are dual operations. If you search your memory for properties about union and intersection that you proved when you first learned about sets, you will find that they usually come in pairs, the first property being the dual of the second.[170]

[170] e.g.:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

C.4 More Vocabulary

In the next chapter, we will start heavily using diagrams, and in order to generalize many concepts relying on diagrams, we will need a formal abstract definition of diagrams to work with. We introduce this definition here[171] and throw in a couple of new concepts and their duals to keep practicing with the central idea in this chapter.

[171] In the rest of the book, we use the term *diagram* to refer to both the informal pictures we draw and the formal mathematical object defined below. The context should disambiguate the two usages, but if you are not sure, rememeber that only the latter use will appear with a hyperlink on the word that links to Definition C.46.

Definition C.46 (Diagram)**.** A **diagram** in $\mathbf{C}$ is a functor $F : \mathbf{J} \rightsquigarrow \mathbf{C}$ where $\mathbf{J}$ is usually a small or even finite category. We say that J is the **shape** of the diagram F.

Remark C.47. Diagrams are usually represented by (partially) drawing the image of F. While all the informal diagrams drawn up to this point can correspond to actual formal diagrams, it is not very pertinent to highlight this correspondence in a case-by-case basis. Indeed, the motivation behind Definition C.46 is the need to abstract away from the drawings to work in more generality. For instance, when considering a commutative square in $\mathbf{C}$, it can be helpful to view it as the image of a functor with codomain $\mathbf{C}$ and domain the category $\mathbf{2} \times \mathbf{2}$ represented in (29).

$$\begin{tikzcd} \cdot \arrow[r] \arrow[d] & \cdot \arrow[d] \\ \cdot \arrow[r] & \cdot \end{tikzcd} \qquad (29)$$

Since diagrams are defined as functors, they interact well with other functors. For example, if $F : \mathbf{J} \rightsquigarrow \mathbf{C}$ is a diagram of shape $\mathbf{J}$ in $\mathbf{C}$ and $G : \mathbf{C} \rightsquigarrow \mathbf{D}$ is a functor, then $G \circ F$ is a diagram of shape $\mathbf{J}$ in $\mathbf{D}$. Some functors interact even more nicely with diagrams.

Definition C.48. Let $F : \mathbf{C} \rightsquigarrow \mathbf{C}'$ be a functor and P a property[172] of diagrams.

- We say that F **preserves** diagrams with property P if for any diagram $D : \mathbf{J} \rightsquigarrow \mathbf{C}$, if D has property P, then $F \circ D$ has property P.

- We say that F **reflects** diagrams with property P if for any diagram $D : \mathbf{J} \rightsquigarrow \mathbf{C}$, if $F \circ D$ has property P, then D has property P.

[172] This is intentionally a vague term. In Chapter D, we will have a more formal but less general definition of preserving and reflecting.

Warning C.49. Preserving and reflecting a property P are not dual notions. The dual of preserving (resp. reflecting) P is preserving (resp. reflecting) the dual of P.

Example C.50 (Commutativity)**.** By drawing the objects and morphisms in the image of a diagram $D : \mathbf{J} \rightsquigarrow \mathbf{C}$, we can still use Definition B.7 to say whether D is commutative or not. Since functors preserve composition, if D is commutative and $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ is any functor, $F \circ D$ is also commutative. Indeed, if two paths in $\mathbf{C}$ compose to the same morphism, then the composites of the paths after applying F are still equal. In other words, all functors preserve commutativity. We will use this fact many times in proofs[173] by drawing a commutative diagram and applying F to all objects and morphisms to get another commutative diagram.

[173] Without the rigor of defining the functors represented by the diagrams.

Commutativity is not reflected by all functors. Even if a diagram $D : \mathbf{J} \rightsquigarrow \mathbf{C}$ does not commute, composing D with the unique functor into the terminal category $\mathbf{1}$ yields a (trivially) commutative diagram $\langle\rangle \circ D : \mathbf{J} \rightsquigarrow \mathbf{1}$.

If $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ is faithful, then F reflects commutativity. Let $D : \mathbf{J} \rightsquigarrow \mathbf{C}$ be a diagram and suppose $F \circ D$ is commutative. As in Definition B.7, take a path in the image of D of length greater than one that composes to $p_1 : A \to B$ and another path that composes to $p_2 : A \to B$. After applying F, commutativity of $F \circ D$ ensures the two paths compose to the same morphism $p \in \mathrm{Hom}_{\mathbf{D}}(FA, FB)$. Moreover, p is the image of both p_1 and p_2, and since F is faithful, we conclude that $p_1 = p_2$.

The following two exercises are a quick investigation in preservation and reflection of simple properties we have seen in this chapter.

OL **Exercise C.51.** 1. Find an example of functor which does not preserve monomorphisms.[174]

2. Show that if $f \in \mathbf{C}_1$ is a split monomorphism, then $F(f)$ is also a split monomorphism, i.e.: any functor preserves split monomorphisms.

3. State and prove the dual statement.

4. Infer that all functors preserve isomorphisms, in particular functors send isomorphic objects to isomorphic objects.

[174] We can see a morphism as a diagram of shape $\mathbf{2}$ because a functor $\mathbf{2} \rightsquigarrow \mathbf{C}$ amounts to a choice of a morphism in $\mathbf{C}$. Thus, a functor F preserves monomorphisms if and only if whenever f is monic, $F(f)$ also is.

OL **Exercise C.52.** 1. Find an example of functor which does not reflect monomorphisms.[175]

[175] A functor reflects monomorphisms if whenever Ff is monic, f also is.

2. Show that if F is faithful, then F reflects monomorphisms.
3. State and prove the dual statement.
4. Show that if F is fully faithful, then F reflects isomorphisms.

We have seen how to *categorify*[176] unions and intersections of subsets in Example C.45. The next set-theoretical notion we categorify is subsets. A subset $I \subseteq S$ can be identified with the inclusion function $I \hookrightarrow S$, and since the latter is injective, we may want to consider monomorphisms with target S to be some kind of generalized subset. Observe however that an injection $I \rightarrowtail S$ is not necessarily an inclusion function. This does not matter because, in reality, we are interested in the image of this injection. We run into another obstacle because if two injections into S have the same image, they represent the same subset. We overcome this using the following exercise.

[176] Categorification is an imprecise term referring to the process of casting an idea in a more categorical language. Depending on the original idea and the context where it is used, there can be many ways to describe it with a categorical mind. In the following two chapters, we will spend some time categorifying several set-theoretical notions.

Two monomorphisms related by $\sim$.

$$\begin{tikzcd} \bullet \arrow[rr, "i"] \arrow[dr, "m"'] & & \bullet \arrow[dl, "m'"] \\ & \bullet & \end{tikzcd}$$

OL **Exercise C.53.** Let $\mathbf{C}$ be a category and $X \in \mathbf{C}_0$, we define the relation $\sim$ on monomorphisms with target X by

$$m \sim m' \Leftrightarrow \exists \text{ isomorphism } i, m = m' \circ i.$$

Show that $\sim$ is an equivalence relation.

Definition C.54 (Subobject)**.** Let $\mathbf{C}$ be a category, a **subobject** of $X \in \mathbf{C}_0$ is an equivalence class of the relation $\sim$ defined above. We will often abusively refer to a subobject simply with a monomorphism $Y \rightarrowtail X$ representing the class. The collection of subobjects of X is denoted $\mathrm{Sub}_{\mathbf{C}}(X)$. If for any $X \in \mathbf{C}_0$, $\mathrm{Sub}_{\mathbf{C}}(X)$ is a set, we say that $\mathbf{C}$ is **well-powered**.

Example C.55 (Set). Let $X \in \mathbf{Set}_0$, subobjects of X correspond to subsets of X.[177] Indeed, any subset $I \subseteq X$ has an inclusion function $i : I \hookrightarrow X$ which is injective, hence monic. For the other direction, we can show that $i : I \rightarrowtail X$ and $j : J \rightarrowtail X$ are in the same equivalence class in $\mathrm{Sub}_{\mathbf{Set}}(X)$ if and only if $\mathrm{Im}(i) = \mathrm{Im}(j)$.[178] We conclude that the correspondence between $\mathrm{Sub}_{\mathbf{Set}}(X)$ and $\mathcal{P}(X)$ sends $[i]$ to the image of i and $I \subseteq X$ to the equivalence class of the inclusion $i : I \hookrightarrow X$.

[177] The notation $\mathrm{Sub}_{\mathbf{Set}}(X)$ is perfect!

[178] ($\Rightarrow$) If $i \sim j$, then there exists a bijection f such that $i = j \circ f$. It follows that the image of j is the image of i.
($\Leftarrow$) Suppose $\mathrm{Im}(i) = \mathrm{Im}(j)$, we define $f : I \to J = x \mapsto j^{-1}(i(x))$, where j^{-1} is the left inverse of j. It is clear that $i = j \circ f$ and a quick computation shows f is an isomorphism with inverse $x \mapsto i^{-1}(j(x))$, where $i^{-1}(x)$ is the left inverse of i.

The next exercise generalizes the poset of subsets of X $(\mathcal{P}X, \subseteq)$.

OL **Exercise C.56.** Let $\mathbf{C}$ be a category and $X \in \mathbf{C}_0$, we define the relation $\leq$ on $\mathrm{Sub}_{\mathbf{C}}(X)$:

$$[m] \leq [m'] \Leftrightarrow \exists \text{ morphism } k, m = m' \circ k.$$

Show that $\leq$ is a well-defined partial order.

OL **Exercise C.57.** Show that the correspondence between $\mathcal{P}X$ and $\mathrm{Sub}_{\mathbf{Set}}(X)$ from Example C.55 is an isomorphism of posets $(\mathcal{P}X, \subseteq) \cong (\mathrm{Sub}_{\mathbf{Set}}(X), \leq)$.[179]

[179] We saw what poset isomorphisms were in Example C.28.2.

OL **Exercise C.58.** Show that a subobject in **Cat** is a subcategory.

We can use duality to obtain (for free) the notion of quotient objects.

Definition C.59 (Quotients)**.** Let $\mathbf{C}$ be a category and $X \in \mathbf{C}_0$, there is an equivalence relation $\sim$ on epimorphisms with source X defined by

$$q \sim q' \Leftrightarrow \exists \text{ isomorphism } i, q = i \circ q'.$$

A **quotient object** (or simply quotient) of X is an equivalence class of the relation $\sim$ defined above.[180] The collection of quotients of X is denoted $\mathrm{Quot}_{\mathbf{C}}(X)$. If for any $X \in \mathbf{C}_0$, $\mathrm{Quot}_{\mathbf{C}}(X)$ is a set, we say that $\mathbf{C}$ is **co-well-powered**. There is a partial order $\leq$ on $\mathrm{Quot}_{\mathbf{C}}(X)$ defined by

$$[q] \leq [q'] \Leftrightarrow \exists \text{ morphism } k, q = k \circ q'.$$

[180] We will often abusively refer to a quotient simply with an epimorphism $X \twoheadrightarrow Y$.

The terminology for this dual notion is motivated by the following exercise.

SOL Exercise C.60. Show that a quotient object of $G \in \mathbf{Grp}_0$ is a quotient group of G.

I love finding a categorical definition for something I am used to thikning of in classical terms.[181] It facilitates a better understanding of the essential components of the classical notion, and duality can open the gates to a parallel world where we can have just as much fun.

[181] This feeling led me to study more category theory.

For now, we only played with definitions without discovering anything deep. Some people maintain it is useless to take a categorical point of view if it does not lead to new results. Category theorists (I presume) believe that it helps organize our thoughts regardless of the mathematical outcomes. The rest of the book focuses on practicing categorical thinking without necessarily demonstrating its advantages other than its unifying/orgasitional power.

D Limits and Colimits

The unifying power of categorical abstraction is arguably its biggest benefit. Indeed, it is often the case that many mathematical objects or results from different fields fit under the same categorical definition or fact. In my opinion, category theory is at its peak of elegance when a complex idea becomes close to trivial when viewed categorically, and when this same view helps link together the intuitions behind many ideas throughout mathematics.

The next two chapters concern one particular instance of this power: universal constructions. Along with Chapter G, these three chapters constitute the heart of our investigation into a philosophical idea central to category theory:[182]

> A mathematical object is completely determined by its relations with other objects of the same kind.

[182] We already hinted at it in Chapter B. I am not a good philosopher of mathematics, but I believe this statement is a fundamental belief in structuralism.

This chapter will cover limits and colimits which are special cases of universal constructions. We postpone the rigorous definition of the term "universal", so, for a while, I recommend you try to recognize *universality* as the thing that all definitions of (co)limits given below have in common.[183]

[183] This is also a good practice for reading more literature on category theory since "universal" can also be used informally.

The first section presents several examples. Each of its subsection is dedicated to one kind of limit or colimit of which a detailed example in **Set** is given along with a couple of interesting examples in other categories. It is not straightforward to build intuition about all kinds of (co)limits due to their innumerable applications. For now, I think it is fine if you are comfortable with the intuition in **Set** as it transposes well to concrete categories, but if you persist in learning category theory, you will get to see examples with other flavors. The second section gives a formal framework to talk about all the examples previously explored as well as a few general results. The third section is a training ground to practice a new proof technique called diagram chasing,[184] we will cover important results there too.

[184] It extends diagram paving using the tools seen in the chapter.

In the sequel, **C** denotes a category.

D.1 Examples

Before giving the definition of (co)limits which is very abstract, we present several examples of how they are used. These are very interesting on their own because they show you how a lot of things mathematicians care about in different contexts can be seen as the same abstract construction. Still, keep in mind that, after adding another level of abstraction, we will bring all these examples together as instances of (co)limits.

Products

Given two sets S and T, the most common construction of the Cartesian product $S \times T$ is conceptually easy: you take all pairs of elements S and T, that is,

$$S \times T := \{(s,t) \mid s \in S, t \in T\}.$$

This construction requires to pick out elements in S and T, form pairs of elements, and use the set-builder notation. While these steps are straightforward set-theoretically, it is not so clear how one would translate them into categorical language.[185] You can try to do it for the first step.

[185] Only working with the objects and morphisms of the category **Set**.

OL **Exercise D.1.** Inside the category **Set**, give a categorical definition of an element of a set. Your definition must only refer to objects and morphisms, so it can be generalized to other categories. Does your definition still correspond to an intuitive notion of elements inside **Poset**, **Grp**, **Cat**?

If one hopes to generalize products to other categories, the construction must only involve objects and morphisms.

Question D.2. *What are essential functions (morphisms in* **Set***) to consider when studying $S \times T$?*

Answer. Projection maps. They are functions $\pi_1 : S \times T \to S$ and $\pi_2 : S \times T \to T$,[186] but that is not enough to define the product. Indeed, there are projection maps $\pi_1' : S \times T \times S \to S$ and $\pi_2' : S \times T \times S \to T$, but $S \times T \times S$ is not always isomorphic to $S \times T$. □

[186] The projections are defined by $\pi_1(s,t) = s$ and $\pi_2(s,t) = t$ for all $(s,t) \in S \times T$.

Question D.3. *What is unique*[187] *about $S \times T$ with the projections π_1 and π_2?*

[187] Always up to isomorphism of course.

Answer. For one, π_1 and π_2 are surjective, and while they are not injective, they have an invertible-like property. Namely, given $s \in S$ and $t \in T$, the pair (s,t) is completely determined from $\pi_1^{-1}(s) \cap \pi_2^{-1}(t)$. □

Again, in order to get rid of the references to specific elements, another point of view is needed. Let X be a set of *choices* of pairs, an element $x \in X$ chooses elements in S and T via functions $c_1 : X \to S$ and $c_2 : X \to T$ (similar to the projections). Now, the *almost-inverse* defined above yields a function

$$! : X \to S \times T = x \mapsto \pi^{-1}(c_1(x)) \cap \pi^{-1}(c_2(x)).$$

This function maps $x \in X$ to an element in $S \times T$ that makes the same choice as x, and it is the only one that does so. Categorically, $!$ is the unique morphism in $\mathrm{Hom}_{\mathbf{C}}(X, S \times T)$ satisfying $\pi_i \circ ! = c_i$ for $i = 1, 2$. Later, we will see that this property completely determines $S \times T$. For now, enjoy the power we gain from generalizing this idea.

Definition D.4 (Binary product). Let $A, B \in \mathbf{C}_0$. A (categorical) **binary product** of A and B is an object, denoted $A \times B$, along with two morphisms $\pi_A : A \times B \to A$ and $\pi_B : A \times B \to B$ called **projections** that satisfy the following universal property[188]: for every object $X \in \mathbf{C}_0$ with morphisms $f_A : X \to A$ and $f_B : X \to B$, there is a unique morphism $! : X \to A \times B$ making diagram (30) commute.

$$\begin{tikzcd} & X \arrow[dl, "f_A"'] \arrow[d, dashed, "!"] \arrow[dr, "f_B"] & \\ A & A \times B \arrow[l, "\pi_A"] \arrow[r, "\pi_B"'] & B \end{tikzcd} \tag{30}$$

[188] Remember that the word universal is not yet defined, we are trying to get an idea of what it means with these examples.

We will often denote $! = \langle f_A, f_B \rangle$ and call it the **pairing** of f_A and f_B.

Example D.5 (**Set**). Cleaning up the argument above, we show that the Cartesian product $A \times B$ with the usual projections is a binary product in **Set**. To show that it satisfies the universal property, let X, f_A and f_B be as in the definition. A function $! : X \to A \times B$ that makes (30) commute must satisfy

$$\forall x \in X, \pi_A(!(x)) = f_A(x) \text{ and } \pi_B(!(x)) = f_B(x).$$

Equivalently, $!(x) = (f_A(x), f_B(x))$. Since this uniquely determines $!$, $A \times B$ is indeed the binary product.

Example D.6. Most of the constructions throughout mathematics with the name *product* can also be realized with a categorical product. Examples include the product of groups, rings or vector spaces, the product of topologies, etc. The fact that all these constructions are based on the Cartesian product of the underlying sets is a corollary of a deeper result about the forgetful functors that all these categories have in common.[189]

Let us give the details for **Mon**, they can be easily adapted for the other categories of algebraic objects (groups, rings, vector spaces) — this does not translate so readily for the product of two topological spaces.

[189] We show in Chapter H that these forgetful functors are right adjoints and thus they preserve binary products (Proposition H.20).

Example D.7. In another flavor, let X be a topological space and $\mathcal{O}(X)$ be the category of opens. If $A, B \subseteq X$ are open, what is their product? Following Definition D.4, the existence of π_A and π_B imply that $A \times B$[190] is included in both sets, or equivalently $A \times B \subseteq A \cap B$.

Moreover, for any open set X included in A and B (via f_A and f_B), X should be included in $A \times B$ (via $!$).[191] In particular, X can be $A \cap B$ (it is open by definition of a topology), thus $A \cap B \subseteq A \times B$. In conclusion, the product of two open sets is their intersection. In an arbitrary poset, the same argument is used to show the product is the greatest lower bound/infimum/meet.

[190] Recall that $\times$ denotes the categorical product, not the Cartesian product of sets.

[191] Notice that uniqueness of $!$ is already given in a posetal category.

Remark D.8. Given two objects in an arbitrary category, their product does not necessarily exist. Nevertheless, when it exists, one can (and we will) show that it is unique up to unique isomorphism.[192] Thus, in the sequel, we will speak of *the* product of two objects and similarly for other constructions presented in this chapter. Moreover, we will often refer to the object $A \times B$ alone (without the projections) as the product.

[192] The uniqueness of the isomorphism is under the condition that it preserves the structure of the product. We will clear up this subtlety in Remark D.66.

OL **Exercise D.9.** Let (A, R) and (B, S) be two objects in $2\mathbf{Rel}$.[193] We denote $R \wedge S$ the binary relation on $A \times B$ defined by (we write the relations infix like for orders)

$$(a, b)\, R \wedge S\, (a', b') \Leftrightarrow a\, R\, a' \text{ and } b\, S\, b'.$$

1. Show that $(A \times B, R \wedge S)$ is the product of (A, R) and (B, S) in $2\mathbf{Rel}$.
2. Show that if R and S are reflexive/transitive/antisymmetric, then so is $R \wedge S$.
3. Conclude that, in both **Poset** and **Pre**, the product of any two objects exists.

OL **Exercise D.10.** Let A and B be two sets, find their product in the category **Rel**.

OL **Exercise D.11.** Let $\mathbf{C}$ and $\mathbf{D}$, be two categories, we defined the product category $\mathbf{C} \times \mathbf{D}$ in Definition B.40. Resolve the clash of notations by checking that $\mathbf{C} \times \mathbf{D}$ satisfies the universal property of the categorical product of $\mathbf{C}$ and $\mathbf{D}$.

Before reaching even more generality, it is sane to check that we can prove some properties of the Cartesian product using the categorical definition. This would ensure that we are not venturing in useless abstract nonsense. We prove the harder one and leave you two easier ones as exercises.

Proposition D.12. *Let $A, B, C \in \mathbf{C}_0$ be such that $A \times B$ and $B \times C$ exist. If $A \times (B \times C)$ exists, then $(A \times B) \times C$ exists and both products are isomorphic. In other words, the binary product is associative.*[194]

Proof. We will show that $A \times (B \times C)$ satisfies the definition of the product $(A \times B) \times C$ with projections defined below. This means $(A \times B) \times C$ exists and the fact that $A \times (B \times C) \cong (A \times B) \times C$ follows trivially (we defined them to be the same object).[195]

First, we need two projections $\pi_{A\times B} : A \times (B \times C) \to A \times B$ and $\pi_C : A \times (B \times C) \to C$. In the diagram below, we show how to obtain them.[196]

$$\begin{tikzcd}
 & & A \times (B \times C) \arrow[dll, bend right, "\pi_A"'] \arrow[dl, dotted, "\pi_{A\times B}"] \arrow[dr, "\pi_{B\times C}"'] \arrow[drr, bend left, dotted, "\pi_C"] & & \\
 & A \times B \arrow[dl, "\pi_A"] \arrow[dr, "\pi_B"'] & & B \times C \arrow[dl, "\pi_B"] \arrow[dr, "\pi_C"'] & \\
A & & B & & C
\end{tikzcd} \tag{31}$$

The dotted arrow π_C is simply the composition $\pi_C \circ \pi_{B\times C}$. The dotted arrow $\pi_{A\times B}$ is obtained via the property of the product $A \times B$ and the morphisms $\pi_A : A \times (B \times C) \to A$ and $\pi_B \circ \pi_{B\times C} : A \times (B \times C) \to B$. It is the unique morphism making (31) commute, that is, $\pi_{A\times B} = \langle \pi_A, \pi_B \circ \pi_{B\times C} \rangle$.

Suppose there is an object X and morphisms $p_{A\times B} : X \to A \times B$ and $p_C : X \to C$. We need to find $! : X \to A \times (B \times C)$ that makes (32) commute and is unique with that property. By post-composing with the appropriate projections, we can see how $!$ acts from the point of view of A, B and C:

[193] i.e. A and B are sets and $R \subseteq A \times A$ and $S \subseteq B \times B$.

[194] Just like the Cartesian product is associative (up to isomorphism). The existence hypothesis is not necessary in **Set** because the Cartesian product of any two sets always exists.

[195] In any case, as we will prove in Proposition D.65, if you had another construction for $(A \times B) \times C$, it would be isomorphic to ours.

[196] We overload the notation and rely on the source and target of the morphisms to avoid confusion

$$\begin{tikzcd}
 & & X \arrow[d, dashed, "!"] \arrow[ddl, bend right, "p_{A\times B}"'] \arrow[ddrr, bend left, "p_C"] & & \\
 & & A \times (B \times C) \arrow[ddll, bend right, "\pi_A"'] \arrow[dl, dotted, "\pi_{A\times B}"] \arrow[dr, "\pi_{B\times C}"'] \arrow[ddrr, bend left, dotted, "\pi_C"] & & \\
 & A \times B \arrow[dl, "\pi_A"] \arrow[dr, "\pi_B"'] & & B \times C \arrow[dl, "\pi_B"] \arrow[dr, "\pi_C"'] & \\
A & & B & & C
\end{tikzcd} \tag{32}$$

$$\begin{aligned}
\pi_A \circ ! &= \pi_A \circ \langle \pi_A, \pi_B \circ \pi_{B\times C}\rangle \circ ! = \pi_A \circ p_{A\times B}\\
\pi_B \circ \pi_{B\times C} \circ ! &= \pi_B \circ \langle \pi_A, \pi_B \circ \pi_{B\times C}\rangle \circ ! = \pi_B \circ p_{A\times B}\\
\pi_C \circ \pi_{B\times C} \circ ! &= p_C.
\end{aligned}$$

The last two equations tell us that $\pi_{B\times C} \circ !$ must make (33) commute.

$$\begin{tikzcd}
 & X \arrow[dl, "p_{A\times B}"'] \arrow[d, dashed, "!"'] \arrow[ddr, bend left, "p_C"] & \\
A \times B \arrow[d, "\pi_B"'] & A \times (B \times C) \arrow[d, "\pi_{B\times C}"'] & \\
B & B \times C \arrow[l, "\pi_B"] \arrow[r, "\pi_C"'] & C
\end{tikzcd} \tag{33}$$

Hence, by the universal property of $B \times C$, we must have $\pi_{B\times C} \circ ! = \langle \pi_B \circ p_{A\times B}, p_C\rangle$. This fact combined with the first equation tells us that ! makes (34) commute.

$$\begin{tikzcd}
 & X \arrow[dl, "p_{A\times B}"'] \arrow[dd, dashed, "!"'] \arrow[ddr, bend left, "\langle \pi_B \circ p_{A\times B}, p_C\rangle"] & \\
A \times B \arrow[d, "\pi_A"'] & & \\
A & A \times (B \times C) \arrow[l, "\pi_A"] \arrow[r, "\pi_{B\times C}"'] & B \times C
\end{tikzcd} \tag{34}$$

Hence, by the universal property of $A \times (B \times C)$, we must have $! = \langle \pi_A \circ p_{A\times B}, \langle \pi_B \circ p_{A\times B}, p_C\rangle\rangle$. Notice that the two uses of universal properties ensured that we found the unique possible choice for !. □

Remark D.13. This has been our first proof using **diagram chasing**. It is different from diagram paving because the goal is to construct objects and morphisms that make some diagram commute (often with a proof of uniqueness of your construction).[197] Another unfortunate difference is that diagram chasing proofs are much harder to typeset. On the board, this proof can be done with one big diagram on which we point out the relevant parts at different moments in the proof. Here, we had to draw four diagrams for this proof in order to emphasize different parts of that huge diagram.

[197] In diagram paving, you only use objects and morphisms that are given. One can see diagram paving as part of diagram chasing because the commutativity proofs are done by combining smaller commutative diagrams.

Here are two simpler diagram chasing exercises for you to solve. It should help to highlight the important steps of the proof above. To show $A \times (B \times C)$ is the same thing as $(A \times B) \times C$, we showed the former satisfies the universal property of the latter. We built the appropriate projections, and given another object with maps to $A \times B$ and C, we showed how to construct the pairing of these maps, and finally we showed that pairing was unique.

OL **Exercise D.14.** Let $A, B \in \mathbf{C}_0$. If $A \times B$ exists, then $B \times A$ exists and both products are isomorphic. In other words, the binary product is commutative.[198]

[198] Just like the Cartesian product is commutative (up to isomorphism).

This statement is transparent in the definition of binary products because changing A for B in Definition D.4 has no impact. Still, proving it is more rigorous.

OL **Exercise D.15.** Let $\mathbf{1}$ be the terminal object in $\mathbf{C}$. Show that for any $A \in \mathbf{C}_0$, the product of $\mathbf{1}$ and A is A.[199]

[199] This property is expected because in **Set**, $\mathbf{1} = \{*\}$ and
$$\{*\} \times A = \{(*, a) \mid a \in A\} \cong A.$$

This last exercise can help when you are trying to find the product in a category. For instance, in **Rel**, since the objects are sets, we might expect the binary product of two sets in **Rel** to be the Cartesian product again. However, we saw in Example C.38 that in this category the $\mathbf{1}$ is the empty set, and $\emptyset \times A \neq A$, so the binary product must be something else.[200]

[200] This might help if you got stuck on Exercise D.10.

To generalize the categorical product to more than two objects, one can, for instance, define the product of a finite family of sets recursively with the binary product.[201] This is well-defined thanks to the associativity and commutativity of $\times$, but this is not enough to get the infinite case. In contrast, generalizing the universal property illustrated in (30) yields a simpler definition that works even for arbitrary families. Instead of having only two objects and two projections, we will have a families of objects and projections indexed by an arbitrary set I.

[201] For a family $\{X_1, \dots, X_n\} \subseteq \mathbf{C}_0$:
$$\prod_{i=1}^{n} X_i = \begin{cases} X_1 & n = 1 \\ \left(\prod_{i=1}^{n-1} X_i\right) \times X_n & n > 1 \end{cases}$$

Definition D.16 (Product)**.** Let $\{X\}_{i\in I}$ be an I–indexed family of objects of $\mathbf{C}$. The **product** of this family is an object $\prod_{i\in I} X_i$ along with projections $\pi_j : \prod_{i\in I} X_i \to X_j$ for all $j \in I$ satisfying the following universal property: for any object X with morphisms $\{f_j : X \to X_j\}_{j\in I}$, there is a unique morphism $! : X \to \prod_{i\in I} X_i$ making (35) commute for all $j \in I$.[202]

[202] Analogously to the binary case, we may write $! = \langle f_j\rangle_{j\in I}$ or, in the finite case, $! = \langle f_1, \dots, f_n\rangle$.

$$\begin{tikzcd} X \arrow[d, dashed, "!"'] \arrow[dr, "f_j"] & \\ \prod_{i\in I} X_i \arrow[r, "\pi_j"'] & X_j \end{tikzcd} \tag{35}$$

Warning D.17. In a lot of cases, the arbitary product will be a straightforward generalization of the binary product,[203] but that is not true in all cases. For instance, in the category of open subsets of a topological space, the arbitrary product is not always the intersection. This is because arbitrary intersections of open sets are not necessarily open. To resolve this problem, it suffices to take the interior of the intersection which is open by definition.

[203] e.g. in **Set**, the Cartesian product of an arbitrary family of sets is still the set of ordered tuples (instead of pairs) of elements in the sets.

Commutativity and now associativity of categorical products are true by definition.[204] Here are three more properties of Cartesian products that generalize to categorical products.

[204] We mean the order of the X_is is not taken into account for the universal property. As we did for binary products, we will make this more rigorous in ...

OL **Exercise D.18** (NOW!)**.** Let $\{f_i : X_i \to Y_i\}_{i\in I}$ be a family of morphisms in $\mathbf{C}$, show that there is a unique morphism $\prod_{i\in I} f_i : \prod_{i\in I} X_i \to \prod_{i\in I} Y_i$ making the following square commute for all $j \in I$.

$$\begin{tikzcd} \prod_{i\in I} X_i \arrow[r, "\prod_{i\in I} f_i"] \arrow[d, "\pi_j"'] & \prod_{i\in I} Y_i \arrow[d, "\pi_j"] \\ X_j \arrow[r, "f_j"'] & Y_j \end{tikzcd} \tag{36}$$

We call $\prod_{i\in I} f_i$ the **product** of the f_is. In the finite case, we write $f_1 \times \cdots \times f_n$.

In **Set**, the function $\prod_{i\in I} f_i$ acts on tuples in $\prod_{i\in I} X_i$ by applying f_i to the ith coordinate for every i.

OL Exercise D.19. Let X, Y and $\{X_i\}_{i\in I}$ be objects of $\mathbf{C}$ such that $\prod_{i\in I} X_i$ exists. For any family $f_i : X \to X_i$ and $g : Y \to X$ show that $\langle f_i \rangle_{i\in I} \circ g = \langle f_i \circ g\rangle_{i\in I}$. Conclude that for families $\{f_i : X_i \to Y_i\}_{i\in I}$ and $\{g_i : Z_i \to X_i\}_{i\in I}$, $(\prod f_i) \circ (\prod g_i) = \prod (f_i \circ g_i)$.[205]

A family of objects in $\mathbf{C}$ is also called a **discrete diagram** because it corresponds to a functor from a discrete category (one with no non-identity morphisms) into $\mathbf{C}$.[206] The product of a family of objects is called the limit of the corresponding diagram. The big takeaway from last chapter is that each time we read a new definition, it is worth to dualize it. Thus, we ask: what is the colimit of a discrete diagram?

[205] It may be useful to restate this in the binary case. For any $f : X \to Y$, $f' : X' \to Y'$, $g : Z \to X$ and $g' : Z' \to X'$, we have

$$(f \times f') \circ (g \times g') = (f \circ g) \times (g \circ g').$$

As a corollary, if $\mathbf{C}$ has all binary products, we get a functor $\mathbf{C} \times \mathbf{C} \rightsquigarrow \mathbf{C}$ sending (X, Y) to $X \times Y$ and (f, g) to $f \times g$.

[206] Recall that a diagram is a functor into $\mathbf{C}$ (Definition C.46).

Coproducts

Definition D.20 (Coproduct)**.** Let $\{X\}_{i\in I}$ be an I–indexed family of objects in $\mathbf{C}$, its **coproduct** is an object, denoted $\coprod_{i\in I} X_i$ (or $X_1 + X_2$ in the binary case), along with morphisms $\kappa_j : X_j \to \coprod_{i\in I} X_i$ for all $j \in I$ called **coprojections** satisfying the following universal property: for any object X with morphisms $\{f_j : X_j \to X\}_{j\in I}$, there is a unique morphism $! : \coprod_{i\in I} X_i \to X$ making (37) commute for all $j \in I$.[207]

$$\begin{tikzcd} X_j \arrow[r, "\kappa_j"] \arrow[dr, "f_j"'] & \coprod_{i\in I} X_i \arrow[d, dashed, "!"] \\ & X \end{tikzcd} \tag{37}$$

[207] We may denote $! = [f_j]_{j\in I}$ or, in the finite case, $! = [f_1, \ldots, f_n]$. We call it the **copairing** of $\{f_j\}_{j\in I}$.

Let us find out what coproducts of sets are.

Example D.21 (**Set**)**.** Let $\{X_i\}_{i\in I}$ be a family of sets, first note that if $X_j = \emptyset$ for $j \in I$, then there is only one morphism $X_j \to X$ for any X.[208] In particular, (37) commutes no matter what $\coprod_{i\in I} X_i$ and X are. Therefore, removing X_j from this family does not change how the coproduct behaves, hence no generality is loss from assuming all X_is are non-empty.

Second, for any $j \in I$, let $X = X_j$, $f_j = \mathrm{id}_{X_j}$ and for any $j' \neq j$, let $f_{j'}$ be any function in $\mathrm{Hom}(X_{j'}, X_j)$.[209] Commutativity of (37) implies κ_j has a left inverse because $! \circ \kappa_j = f_j = \mathrm{id}_{X_j}$, so all coprojections are injective.

Third, we claim that for any $j \neq j' \in I$, $\mathrm{Im}(\kappa_j) \cap \mathrm{Im}(\kappa_{j'}) = \emptyset$. Let $X = \{0, 1\}$, f_j and $f_{j'}$ be the constant functions sending everything to 0 and 1 respectively. The universal property implies that

$$\mathrm{Im}\left(! \circ \kappa_j\right) = \{0\} \neq \{1\} = \mathrm{Im}(! \circ \kappa_{j'}),$$

hence for any $x \in X_j$ and $x' \in X_{j'}$, we have $\kappa_j(x) \neq \kappa_{j'}(x')$.

In summary, the previous points say that $\coprod_{i\in I} X_i$ contains distinct copies of the images of all coprojections. Furthermore, the κ_js being injective, their image can be identified with the X_js to obtain[210]

[208] Because $\emptyset$ is initial.

[209] One exists because X_j is non-empty.

[210] The symbol $\sqcup$ denotes the disjoint union of sets.

$$\bigsqcup_{i\in I} X_i \subseteq \coprod_{i\in I} X_i.$$

For the converse inclusion, in (37), let X be the disjoint union and the f_js be the inclusions. Assume there exists x in the R.H.S. that is not in the L.H.S., then we can define $!' : \coprod_{i\in I} X_i \to \bigsqcup_{i\in I} X_i$ that only differs from $!$ at x. Since x is not in the image of any coprojection, the diagrams still commute and this contradicts the uniqueness of $!$.

In conclusion, the coproduct in **Set** is the disjoint union and the coprojections are the inclusions.[211]

[211] We recover the intuition for why empty sets can be ignored. A more general fact is proven in Exercise D.25.

Remark D.22. If this example looks more complicated than the product of sets, it is because we started knowing nothing concrete about coproducts of sets and gradually discovered what properties they had using specific objects and morphisms we know exist in **Set**. In contrast, we knew what products of sets were, and we just had to show they satisfied the universal property.[212]

[212] One might argue that coming up with this universal property was the hard part in that case.

In general, the hard part is to find what construction satisfies a universal property, proving it does is easier.

Examples D.23. In the category of open sets of a space (X, τ), let $\{U_i\}_{i\in I}$ be a family of open sets and suppose $\coprod_i U_i$ exists. The coprojections yield inclusions $U_j \subseteq \coprod_i U_i$ for all $j \in I$, so $\coprod_i U_i$ must contain all U_js and thus $\cup_i U_i$. Moreover, in (37), letting f_j be the inclusion $U_j \hookrightarrow \cup_i U_i$ for all $j \in I$,[213] the existence of $!$ yields an inclusion $\coprod_i U_i \subseteq \cup_i U_i$. We conclude that the coproduct in this category is the union of open sets. In an arbitrary poset, the same argument is used to show the coproduct is the least upper bound/supremum/join.

[213] These morphisms are in $\mathcal{O}(X)$ because $\cup_i U_i$ is open.

In $\mathbf{Vect}_k$, the coproduct, also called the direct sum, is defined by[214]

$$\coprod_{i\in I} V_i = \bigoplus_{i\in I} V_i := \left\{ \vec{v} \in \prod_{i\in I} V_i \mid \vec{v}_i \neq 0 \text{ for finitely many } i\text{'s} \right\},$$

[214] Here, the symbol $\prod$ denotes the Cartesian product of the V_is as sets. The categorical product of vector spaces is also the direct sum, where the projections are the usual ones.

where $\kappa_j : V_j \hookrightarrow \coprod_i V_i$ sends v to $\kappa_j(v) \in \prod_i V_i$ satisfying $\kappa_j(v)_j = v$ and $\kappa_j(v)_{j'} = 0$ whenever $j \neq j'$. To verify this, let $\{f_j : V_j \to X\}_{j\in I}$ be a family of linear maps. We can construct $!$ by defining it on basis elements of the direct sum, which are just the basis elements of all V_js seen as elements of the sum (via the coprojections).[215] Indeed, if b is in the basis of V_j, we let $!(\kappa_j(b)) = f_j(b)$. Extending linearly yields a linear map $! : \coprod_i V_i \to X$. Uniqueness is clear because if $h : \coprod_i V_i \to X$ differs from $!$ on one of the basis elements, it does not make (37) commute.

[215] It is necessary to require finitely many non-zero entries, otherwise the basis of the coproduct would not be the union of all bases of the V_js.

OL **Exercise D.24.** Let A and B be two sets, show that their coproduct exists in the category **Rel** and find what it is.

OL **Exercise D.25.** Show that products are dual to coproducts, namely, if a product of a familiy $\{X_i\}_{i\in I}$ exists in $\mathbf{C}$, then this object and the projections are the coproduct of this family and the coprojections in $\mathbf{C}^{\mathrm{op}}$ and vice-versa. Conclude that you can define the **coproduct of morphisms** dually to Exercise D.18, we denote them $\coprod_{i\in I} f_i$ or $f_1 + \cdots + f_n$ in the finite case.

Applying the duality between products and coproducts to Proposition D.12 and Exercises D.14 and D.15, we get the following results.

Corollary D.26 (Dual)**.** *Taking binary coproducts is commutative and associative, and if $\emptyset$ is initial, then $A + \emptyset \cong A$.*[216]

[216] While in **Set**, we have $A \times \emptyset \cong \emptyset$, this does not generalize to all categories with binary products and an initial object, e.g. $\mathbf{Vect}_k$.

OL **Exercise D.27.** Dually to Exercise D.19, show that if X, Y and $\{X_i\}_{i\in I}$ are objects of $\mathbf{C}$ such that $\coprod_{i\in I} X_i$ exists, then for any family $f_i : X_i \to X$ and $g : X \to Y$ show that $g \circ [f_i]_{i\in I} = [g \circ f_i]_{i\in I}$.

OL **Exercise D.28.** Let $\mathbf{C}$ have a terminal object $\mathbf{1}$. Show that the assignment $X \mapsto X + \mathbf{1}$ is functorial, i.e. define the action of $(- + \mathbf{1})$ on morphisms and show it satisfies the axioms of a functor.[217]

[217] We call $(- + \mathbf{1})$ the **maybe functor**.

In a very similar way to the product and coproduct, we will define various constructions in **Set**.[218]

[218] We will follow more closely the section on coproducts where we started with the definition of the (co)limit and then detailed an example in **Set**.

Equalizers

We briefly mentioned that a product (resp. coproduct) is a limit (resp. colimit) of a discrete diagram. The rest of the examples before generalizing will be (co)limits of small diagrams that contain non-identity morphisms.

Definition D.29 (Fork)**.** A **fork** in $\mathbf{C}$ is a diagram of shape (38) or (39).

$$O \xrightarrow{o} A \underset{g}{\overset{f}{\rightrightarrows}} B \qquad (38) \qquad\qquad A \underset{g}{\overset{f}{\rightrightarrows}} B \xrightarrow{o} O \qquad (39)$$

These are dual notions, so we prefer to call (39) a **cofork**. If (38) commutes then $f \circ o = g \circ o$,[219] and we say that o **equalizes** f and g. If (39) commutes, then $o \circ f = o \circ g$, and we say that o **coequalizes** f and g.

[219] Recall that commutativity does not make parallel morphisms equal.

Definition D.30 (Equalizer)**.** Let $A, B \in \mathbf{C}_0$ and $f, g : A \to B$ be parallel morphisms. The **equalizer** of f and g is an object E and a morphism $e : E \to A$ satisfying $f \circ e = g \circ e$ with the following universal property: for any morphism $o : O \to A$ equalizing f and g, there is a unique $! : O \to E$ making (40) commute.[220]

$$\begin{tikzcd} O \arrow[d, dashed, "!"'] \arrow[dr, "o"] & & \\ E \arrow[r, "e"'] & A \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & B \end{tikzcd} \qquad (40)$$

[220] Try to look for a common pattern in this definition and the definition of a product (both are instances of limits).

In other words, e is a morphism that equalizes f and g, and every other o that equalizes f and g factors through e uniquely. A common notation for e is $\mathrm{eq}(f, g)$. There is also a straightforward generalization to equalizers of more than two morphisms.[221]

[221] If $\{f_i\}_{i\in I}$ is a family of parallel morphisms, their equalizer is a morphism $e \in \mathbf{C}_1$ such that

$$\forall i, j \in I, f_i \circ e = f_j \circ e,$$

and every o with this property factors through e in a unique way.

Example D.31 (Set). Let $f, g : A \to B$ be two functions and suppose their equalizer exists and it is $e : E \to A$. By associativity, for any $h : O \to E$, the composite $e \circ h$ is a candidate for o in diagram (40) because $f \circ (e \circ h) = g \circ (e \circ h)$. What is more, if h' is such that $e \circ h = e \circ h'$, then $h = h'$ or it would contradict the uniqueness of !. We conclude that e is monic/injective.[222]

[222] This argument was independent of the category, hence we can conclude that an equalizer is always a monomorphisms.

This implies E can be identified with its image under e. Since $f \circ e = g \circ e$, the image of e is contained in the subset $\{a \in A \mid f(a) = g(a)\}$. Now, by the universal property of the equalizer, letting O be this subset and o be the inclusion, there is an injection[223] $! : \{a \in A \mid f(a) = g(a)\} \hookrightarrow E$, thus both sets are equal. In conclusion, the equalizer of two parallel functions is the subset E in which they coincide and $e : E \hookrightarrow A$ is the inclusion.

[223] The fact that ! is an injection follows because the inclusion o is an injection and $e \circ ! = o$.

Examples D.32. In a posetal category, hom-sets are singletons, so it must be the case that $f = g$ whenever f and g are parallel. Therefore, any $o : O \to A$ satisfies $f \circ o = g \circ o$. Written using the order notation, the universal property is then equivalent to the fact that $E \leq A$ and $O \leq A$ implies $O \leq E$. In particular, if $O = A$, then $A \leq E$, so $A = E$ by antisymmetry.

In **Ab**, **Ring** or $\mathbf{Vect}_k$, for the same reason that the Cartesian product of the underlying sets is the underlying set of the product,[224] the construction of equalizers is as in **Set**. However, since each of these categories have a notion of additive inverse for morphisms, the equalizer of f and g has a cooler name, that is, $\ker(f - g)$.[225]

[224] We explain this in Chapter H.

[225] The equalizer of f and g is the subgroup/subring/subspace of A where f and g are equal, or equivalently, where $f - g$ is 0 (when $f - g$ and 0 are defined).

Definition D.33 (Idempotents)**.** A morphism $f : A \to A \in \mathbf{C}_1$ is called **idempotent** when $f \circ f = f$. It is called **split idempotent** if there exist morphisms $s : E \to A$ and $r : A \to E$ such that $s \circ r = f$ and $r \circ s = \mathrm{id}_E$.[226]

[226] We can show that split idempotents are idempotent because

$$f \circ f = s \circ r \circ s \circ r = s \circ \mathrm{id}_E \circ r = f.$$

Proposition D.34. *An idempotent morphism $f : A \to A \in \mathbf{C}_1$ is split idempotent if and only if the equalizer of* id_A *and f exists.*

Proof. ($\Rightarrow$) Let $f = s \circ r$ be such that $r \circ s = \mathrm{id}_E$, we claim that s is the equalizer. First, we can check that s equalizes id_A and f because $f \circ s = s \circ r \circ s = s \circ \mathrm{id}_E = s = \mathrm{id}_A \circ s$. Next, given $o : O \to A$ that also equalizes id_A and f, we need to find a morphism ! that makes (41) commute. Its uniqueness is given by s being monic (it has a left inverse). Noticing that $o = f \circ o = s \circ r \circ o$, we find $! = r \circ o$.

$$\begin{tikzcd} O \arrow[d, dashed, "!"'] \arrow[dr, "o"] & & \\ E \arrow[r, "s"'] & A \arrow[r, shift left, "\mathrm{id}_A"] \arrow[r, shift right, "f"'] & A \end{tikzcd} \tag{41}$$

($\Leftarrow$) If $e : E \to A$ is the equalizer of f and id_A, then since f equalizes f and id_A, there exists $! : A \to E$ such that $e \circ ! = f$. By monicity of e, we find that $e \circ (! \circ e) = f \circ e = e$ implies $! \circ e = \mathrm{id}_A$, so f is a split idempotent (let $s = e$ and $r = !$). $\square$

The first two examples had a relatively well-known instantiation in the category **Set**, namely, products are Cartesian products and coproducts are disjoint unions. The notion of equalizer of two functions, while just as intuitive as the others[227], is less common in "classical" set theory. However, it still leads to a nice categorical definition of fiber.

[227] The equalizer of $f, g : A \to B$ is the subset of A where f and g are equal.

OL Exercise D.35. Let $f : A \to B$ be a function and $y \in B$, the *fiber* of y (under f) is $\{x \in A \mid f(x) = y\}$.[228] Give a categorical definition of fibers that does not rely on the special case of **Set**. Just like in Exercise D.1, you should only refer to objects and morphisms. In particular, you can only use the categorical notion of elements (Definition **??**). Does your definition still correspond to an intuitive notion of fibers inside **Poset**, **Grp**, **Cat**?

[228] Fiber is just a synonym for preimage (usually) taken at a single point.

OL **Exercise D.36.** A morphism that is the equalizer of two morphisms is called a **regular monomorphism**. We saw this terminology is justified because, in any category $\mathbf{C}$, equalizers are monic. Show that in **Set**, all monomorphisms are regular monomorphisms (i.e. if m is monic, we can find two functions f and g such that m is their equalizer).[229]

[229] Your intuition from Exercise D.35 may be useful.

OL **Exercise D.37.** Show that in a category where all monomorphisms are regular, if f is monic and epic, then f is an isomorphism (i.e. the category is balanced).

The equalizer of f and g is the limit of the diagram containing only the two parallel morphisms, we define its colimit in the next section.

Coequalizers

Definition D.38 (Coequalizer)**.** Let $A, B \in \mathbf{C}_0$ and $f, g : A \to B$ be parallel morphisms. The **coequalizer** of f and g is an object D and a morphism $d : B \to D$ satisfying $d \circ f = d \circ g$ with the following universal property: for any morphism $o : B \to O$ coequalizing f and g, there is a unique $! : D \to O$ making (42) commute.

$$\begin{tikzcd} A \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & B \arrow[r, "d"] \arrow[dr, "o"'] & D \arrow[d, dashed, "!"] \\ & & O \end{tikzcd} \tag{42}$$

In other words, d coequalizes f and g, and every other o that coequalizes f and g factors through d uniquely. A common notation for d is $\operatorname{coeq}(f, g)$, and there is also a straightforward generalization to more than two morphisms.

Example D.39 (Set). Let $f, g : A \to B$ be two functions and suppose $d : B \to D$ is their coequalizer. Similarly to the dual case, one can show that d is epic/surjective. Since $d \circ f = d \circ g$, for any $b, b' \in B$,

$$\big(\exists a \in A, f(a) = b \text{ and } g(a) = b'\big) \implies d(b) = d(b'). \tag{$*$}$$

Denoting by $\sim$ the relation between two elements of B defined in the L.H.S. of $(*)$, the implication becomes $b \sim b' \implies d(b) = d(b')$. Note that $\sim$ is not necessarily an equivalence relation but $=$ is, thus, the converse implication does not always hold.[230]

Consequently, we consider the equivalence relation generated by $\sim$,[231] denoted by $\simeq$. As noted above, the forward implication $b \simeq b' \implies d(b) = d(b')$ still holds. For the converse, in (42), let $O := B/\simeq$ and $o : B \to B/\simeq$ be the quotient map. Post-composing with ! yields

[230] For instance, when $b \sim b' \sim b''$, $d(b) = d(b'')$, but it might not be the case that $b \sim b''$.

[231] In this case, it is simply the transitive closure.

$$d(b) = d(b') \implies o(b) = o(b') \implies b \simeq b'.$$

The equivalence $b \simeq b' \Leftrightarrow d(b) = d(b')$ and the fact that d is surjective means we can identify D with the quotient $B/\simeq$ and $d : B \to D$ with the quotient map.[232]

[232] You can give the isomorphism $D \cong B/\simeq$.

Examples D.40. In a posetal category, an argument dual to the one for equalizers shows the coequalizer of $f, g : A \to B$ is B.

In **Ab**, **Ring** or $\mathbf{Vect}_k$, let $f, g : A \to B$ be homomorphisms and suppose $d : B \to D$ is their coequalizers. Consider the homomorphism $f - g$, since d coequalizes f and g, $d \circ (f - g) = d \circ f - d \circ g = 0$, or equivalently, $\mathrm{Im}(f - g) \subseteq \ker(d)$. Now, consider diagram (43) as an instance of (42), where q is the quotient map.[233]

$$\begin{tikzcd} A \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & B \arrow[r, "d"] \arrow[dr, "q"'] & D \arrow[d, dashed, "!"] \\ & & B/\mathrm{Im}(f-g) \end{tikzcd} \tag{43}$$

We claim that ! has an inverse, implying that $D \cong B/\mathrm{Im}(f - g)$.[234] Indeed, for $[x] \in B/\mathrm{Im}(f - g)$, we must have

$$!^{-1}([x]) = !^{-1}(q(x)) = !^{-1}(!(d(x))) = d(x),$$

and it is only left to show $!^{-1}$ is well-defined because the inverse of a homomorphism is a homomorphism. This follows because if $[x] = [x']$, then there exists $y \in \mathrm{Im}(f - g)$ such that $x = x' + y$, so

$$!^{-1}(x) = d(x) = d(x' + y) = d(x') + d(y) = d(x') + 0 = !^{-1}(x').$$

In the special case that g is the constant 0 map, $B/\mathrm{Im}(f)$ is called the **cokernel** of f, denoted $\mathrm{coker}(f)$.

OL **Exercise D.41.** Show that an idempotent morphism $f : A \to A \in \mathbf{C}_1$ is split idempotent if and only if the coequalizer of f and id_A exists.

OL **Exercise D.42.** Try to dualize the definition of fibers from Exercise D.35. What goes wrong?

Pullbacks

Definition D.43 (Cospan)**.** A **cospan** in **C** comprises three objects A, B, C and two morphisms f and g as in (44).[235]

$$A \xrightarrow{f} C \xleftarrow{g} B \tag{44}$$

Definition D.44 (Pullback)**.** Let $A \xrightarrow{f} C \xleftarrow{g} B$ be a cospan in **C**. Its **pullback** is an object $A \times_C B$ along with morphisms $p_A : A \times_C B \to A$ and $p_B : A \times_C B \to B$ such that $f \circ p_A = g \circ p_B$ and the following universal property holds: for any object X and morphisms $s : X \to A$ and $t : X \to B$ satisfying $f \circ s = g \circ t$, there is a unique morphism $! : X \to A \times_C B$ making (45) commute.[236]

$$\begin{tikzcd} X \arrow[dr, dashed, "!"] \arrow[drr, bend left, "t"] \arrow[ddr, bend right, "s"'] & & \\ & A \times_C B \arrow[r, "p_B"] \arrow[d, "p_A"'] \arrow[dr, phantom, "\lrcorner", very near start] & B \arrow[d, "g"] \\ & A \arrow[r, "f"'] & C \end{tikzcd} \tag{45}$$

[233] It is commutative because $q \circ (f - g) = 0$ by definition of q.

[234] This is not enough to say that $B/\mathrm{Im}(f - g)$ with the quotient map is the coequalizer, we leave you the task to complete the proof using this isomorphism that crucially satisfies $! \circ d = q$.

[235] Just like forks, coforks and spans that we introduce later, cospan is simply a name that we give to a certain shape of diagram that occurs quite often.

[236] The $\lrcorner$ symbol inside the square is a standard convention to specify that a square is not only commutative, but also a pullback square. Some authors call such a square *cartesian*, but this adjective has too many different meanings in category theory in my opinion, so we will not use it.

We call p_A the pullback of g **along** f and sometimes denote it $f^*(g)$. Symmetrically, p_B is the pullback of f along g, denoted $g^*(f)$.

A drawback of the notation $A \times_C B$ is that it does not refer to the morphisms f and g which are essential in the definition. An alternative notation is $f \times_C g$ (I learned about it here). An argument supporting this notation is in Exercise E.47.

Example D.45 (Set). Let $A \xrightarrow{f} C \xleftarrow{g} B$ be a cospan in **Set** and suppose that its pullback is $A \xleftarrow{p_A} A \times_C B \xrightarrow{p_B} B$. Observe that p_A and p_B look like projections, and in fact, by the universality of the product $A \times B$, there is a map $h : A \times_C B \to A \times B$ such that $h(x) = (p_A(x), p_B(x))$ ((46) commutes). Consider the image of h, if $(a,b) \in \mathrm{Im}(h)$, then there exists $x \in A \times_C B$ such that $p_A(x) = a$ and $p_B(x) = b$. Moreover, the commutativity of the square in (46) implies $f(a) = g(b)$, hence

$$\begin{tikzcd} A \times_C B \arrow[rr, "p_B"] \arrow[dr, dashed, "h"] \arrow[dd, "p_A"'] & & B \arrow[dd, "g"] \\ & A \times B \arrow[ur, "\pi_B"'] \arrow[dl, "\pi_A"] & \\ A \arrow[rr, "f"'] & & C \end{tikzcd} \tag{46}$$

$$\mathrm{Im}(h) \subseteq \{(a,b) \in A \times B \mid f(a) = g(b)\}.$$

Now, let X be the R.H.S., and $s = \pi_A|_X$ and $t = \pi_B|_X$ be the projections to A and B respectively restricted to $X \subseteq A \times B$. Our construction ensures $f \circ s = g \circ t$ hence there is a unique $! : X \to A \times_C B$ satisfying $p_A \circ ! = \pi_A|_X$ and $p_B \circ ! = \pi_B|_X$. Viewing h as going in the opposite direction to $!$,[237] we derive for any $(a,b) \in X$,[238]

$$(h \circ !)(a,b) = (p_A(!(a,b)), p_B(a,b)) = (\pi_A(a,b), \pi_B(a,b)) = (a,b),$$

[237] We just saw that the image of h is contained in X, so we can see h as a function $h : A \times_C B \to X$.

[238] We use the fact that $\pi_A \circ h \circ ! = p_A \circ !$ and similarly for B.

thus $!$ has a left inverse and is injective. Assume towards a contradiction that it is not surjective, then let $y \in A \times_C B$ not be in the image of $!$ and denote $x = !(p_A(y), p_B(y))$. Define $!'$ as acting exactly like $!$ except on $(p_A(y), p_B(y))$ where it goes to y instead of x. This ensure that $!'$ still makes the diagram commute, contradicting the uniqueness of $!$.

As a particular case, when one function in the cospan is an inclusion, say $g : B \hookrightarrow C$, the pullback is the preimage of B under f since[239]

[239] This can be seen as a generalization of the fibers defined in Exercise D.35: seeing an element of C as a function $c : \mathbf{1} \to C$, the fiber $f^{-1}(c)$ is the pullback of c along f.

$$\{(a,b) \in A \times B \mid f(a) = g(b) = b\} \cong \{a \mid f(a) \in B\} = f^{-1}(B) \subseteq A.$$

You can also check that p_A is the inclusion $f^{-1}(B) \hookrightarrow A$ and p_B is f restricted to $f^{-1}(B)$. As a particular case of that, if the cospan consists of two inclusions $A \hookrightarrow C \hookleftarrow B$, then its pullback is the intersection $A \cap B$ with p_A and p_B being the inclusions.

$$\begin{tikzcd} A \cap B \arrow[r, hook] \arrow[d, hook] \arrow[dr, phantom, "\lrcorner", very near start] & B \arrow[d, hook] \\ A \arrow[r, hook] & C \end{tikzcd}$$

Examples D.46. In a posetal category, the commutativity of the square in (45) does not depend on the morphisms, thus the universal property is equivalent to the property of being a product.

The composition of relations R and S can be defined using pullbacks in **Set**. Given relations $R \subseteq X \times Y$ and $S \subseteq Y \times Z$, we can restrict the projections to R and S to obtain (47). Then, taking the pullback of the cospan in the middle and using the characterization of the pullback in **Set** from Example D.45, we obtain

$$\begin{tikzcd} & R \arrow[dl, "\pi_X"'] \arrow[dr, "\pi_Y"] & & S \arrow[dl, "\pi_Y"'] \arrow[dr, "\pi_Z"] & \\ X & & Y & & Z \end{tikzcd} \tag{47}$$

$$R \times_Y S = \left\{((x,y),(y',z)) \in R \times S \mid y = y'\right\}.$$

Observe in (48) that we have functions from $R \times_Y S$ to X and Z: $\pi_X \circ p_R$ and $\pi_Z \circ p_S$. Thus, by the universal property of the product $X \times Z$, there is a function $! : R \times_Y S \to X \times Z$. After a bit of computations, recalling that $p_R((x,y),(y',z)) =$

(x,y) and $p_S((x,y),(y',z)) = (y',z)$, we find that the image of ! is precisely the composite relation[240]

$$S \circ R = \{(x,z) \mid \exists y, (x,y) \in R, (y,z) \in S\}.$$

$$\begin{tikzcd}
& & R \times_Y S \arrow[dl, "p_R"'] \arrow[dr, "p_S"] \arrow[dd, phantom, "\vee"] & & \\
& R \arrow[dl, "\pi_X"'] \arrow[dr, "\pi_Y"] & & S \arrow[dl, "\pi_Y"'] \arrow[dr, "\pi_Z"] & \\
X & & Y & & Z
\end{tikzcd} \tag{48}$$

[240] Our argument here heavily relies on working with sets and functions, but there is a way to generalize relations in other nice enough categories using this idea.

OL **Exercise D.47.** Let $f : X \to Y$ be a morphism in $\mathbf{C}$. Show f is monic if and only if the square in (49) is a pullback.[241]

$$\begin{tikzcd}
X \arrow[r, "\mathrm{id}_X"] \arrow[d, "\mathrm{id}_X"'] \arrow[dr, phantom, very near start, "\lrcorner"] & X \arrow[d, "f"] \\
X \arrow[r, "f"'] & Y
\end{tikzcd} \tag{49}$$

[241] This result and its dual will sometimes be used to treat monomorphisms (resp. epimorphisms) as limits (resp. colimits). See e.g. Exercise D.70 where you will show that monomorphisms are preserved by pullback preserving functors (see Definition D.68).

OL **Exercise D.48.** Supposing (50) commutes, show that if the right square is a pullback and i and j are isomosphisms, then the rectangle is a pullback.[242]

$$\begin{tikzcd}
X \arrow[r, leftrightarrow, "i"] \arrow[d, "h"'] & A \times_C B \arrow[r, "p_B"] \arrow[d, "p_A"'] \arrow[dr, phantom, very near start, "\lrcorner"] & B \arrow[d, "g"] \\
Y \arrow[r, leftrightarrow, "j"'] & A \arrow[r, "f"'] & C
\end{tikzcd} \tag{50}$$

[242] i.e. X along with h and $p_B \circ i$ is a pullback of the cospan
$$Y \xrightarrow{f \circ j} C \xleftarrow{g} B.$$

Supposing (51) commutes, show that if the left square is a pullback and i and j are isomorphisms, then the rectangle is a pullback.

$$\begin{tikzcd}
A \times_C B \arrow[r, "p_B"] \arrow[d, "p_A"'] \arrow[dr, phantom, very near start, "\lrcorner"] & B \arrow[r, leftrightarrow, "i"] \arrow[d, "g"] & X \arrow[d, "h"] \\
A \arrow[r, "f"'] & C \arrow[r, leftrightarrow, "j"'] & Y
\end{tikzcd} \tag{51}$$

When dualizing products and equalizers, the shape of the diagram did not change. Indeed, reversing all morphisms in a discrete diagram gives back a discrete diagram, and reversing two parallel morphisms yields two parallel morphisms. However, the opposite of a cospan is a span.

Pushouts

Definition D.49 (Span)**.** A **span** in $\mathbf{C}$ comprises three objects A, B, C and two morphisms f and g as in (52).

$$A \xleftarrow{f} C \xrightarrow{g} B \tag{52}$$

Definition D.50 (Pushout)**.** Let $A \xleftarrow{f} C \xrightarrow{g} B$ be a span in $\mathbf{C}$. Its **pushout** is an object, denoted $A +_C B$, along with morphisms $k_A : A \to A +_C B$ and $k_B : B \to A +_C B$ such that $k_A \circ f = k_B \circ g$ and the following universal property holds: for any object X and morphisms $s : A \to X$ and $t : B \to X$ satisfying $s \circ f = t \circ g$, there is a unique morphism $! : A +_C B \to X$ making (53) commute.[243]

$$\begin{tikzcd}
C \arrow[r, "g"] \arrow[d, "f"'] & B \arrow[d, "k_B"] \arrow[ddr, bend left, "t"] & \\
A \arrow[r, "k_A"'] \arrow[drr, bend right, "s"'] \arrow[ur, phantom, "\ulcorner"] & A +_C B \arrow[dr, dashed, "!"] & \\
& & X
\end{tikzcd} \tag{53}$$

[243] The $\ulcorner$ symbol is a standard convention to specify that the square is not only commutative, but also a pushout square.

We call k_A the pushout of g **along** f and sometimes denote it $f_*(g)$. Symmetrically, k_B is the pushout of f along g, denoted $g_*(f)$.

Example D.51 (**Set**)**.** Let $A \xleftarrow{f} C \xrightarrow{g} B$ be a span in **Set** and suppose its pushout is $A \xrightarrow{k_A} A +_C B \xleftarrow{k_B} B$. Similarly to above, observe that k_A and k_B are like coprojections, so there is a unique map $! : A + B \to A +_C B$ such that $!(a) = k_A(a)$ and $!(b) = k_B(b)$. Furthermore, for any $c \in C$, $!(f(c)) = !(g(c))$, thus

$$\exists c \in C, f(c) = a \text{ and } g(c) = b \implies !(a) = !(b).$$

This is very similar to what happened for coequalizers and after working everything out, we obtain that $! : A + B \to A +_C B$ is the coequalizer of $\kappa_A \circ f$ and $\kappa_B \circ g$. This is a general fact that does not only apply in **Set** but in every category with binary coproducts and coequalizers.

As a particular case, if $C = A \cap B$ and f and g are simply inclusions, then $A +_C B = A \cup B$ (the *non-disjoint* union).

OL **Exercise D.52.** Show that if (54) is a pushout square, then d is the coequalizer of f and g. State and prove the dual statement.

$$\begin{tikzcd}
A \arrow[r, "g"] \arrow[d, "f"'] & B \arrow[d, "d"] \\
B \arrow[r, "d"'] \arrow[ur, phantom, "\ulcorner"] & D
\end{tikzcd} \tag{54}$$

Example D.53 (Rewriting)**.** The categorical approach to graph rewriting is full of uses of pushouts. In this example, we will try to give a flavor of a particular method called double-pushout rewriting (DPO) in an almost trivial setting using words instead of graphs. $\square$.

Just as we defined products and coproducts for more than two objects, and equalizers and coequalizers for more than two morphisms (Footnote 221), we could define pullbacks (resp. pushouts) of multiple morphisms with the same target (resp. source). However, it starts to get messy at this point, so we will abstract away from specific examples of (co)limits.[244]

[244] There is a slick way of doing arbitrary pullbacks and pushouts (as opposed to the binary ones) that we explore in Exercise E.47.

D.2 Generalization

There exists many other examples of (co)limits but these six examples give quite a good idea of what it is to be a limit or colimit. More precisely, we will see in Theorem D.83 and Exercise D.90 that any limit can be built out of products and equalizers or pullbacks and a terminal object. Dually, we can build colimits out of coproducts and coequalizers or pushouts and an initial object.

Let us try to informally spell out the general pattern in the definitions of each example.

- We start with a shape for a diagram D (e.g. a discrete diagram, two parallel morphisms, a span, a cospan, etc.).
- The limit (resp. colimit) of D is an object L along with morphisms from L to every object in the diagram (resp. in the opposite direction) such that combining D with these morphisms yields a commutative diagram.
- These morphisms satisfy a universal property. For any object L' with morphisms from L' to every object in the diagram (resp. in the opposite direction) that commute with D, there is a unique $! : L' \to L$ (resp. $L \to L'$) such that combining all the morphisms with D yields a commutative diagram.

We have already formalized the first step when we defined diagrams in Definition C.46. For the second and third step, notice that the morphisms given for L and L' have the same conditions, they form what we call a cone (resp. cocone).

Definitions

We start by formalizing limits.

Definition D.54 (Cone)**.** Let $F : \mathbf{J} \rightsquigarrow \mathbf{C}$ be a diagram. A cone from X to F is an object $X \in \mathbf{C}_0$, called the **tip**, along with a family of morphisms $\{\psi_Y : X \to F(Y)\}$ indexed by objects $Y \in \mathbf{J}_0$ such that for any morphism $a : Y \to Z$ in $\mathbf{J}_1$, $F(a) \circ \psi_Y = \psi_Z$, i.e. diagram (55) commutes.

$$\begin{tikzcd} & X \arrow[dl, "\psi_Y"'] \arrow[dr, "\psi_Z"] & \\ F(Y) \arrow[rr, "F(a)"'] & & F(Z) \end{tikzcd} \tag{55}$$

Often, the terminology cone over F is used.

Next, the fact that the morphism ! keeps everything commutative can be generalized. We say that ! is a morphism of cones.

Definition D.55 (Morphism of cones)**.** Let $F : \mathbf{J} \rightsquigarrow \mathbf{C}$ be a diagram and $\{\psi_Y : A \to F(Y)\}_{Y \in \mathbf{J}_0}$ and $\{\phi_Y : B \to F(Y)\}_{Y \in \mathbf{J}_0}$ be two cones over F. A **morphism of cones**

from A to B is a morphism $g : A \to B$ in $\mathbf{C}_1$ such that for any $Y \in \mathbf{J}_0$, $\phi_Y \circ g = \psi_Y$, i.e. (56) commutes.

$$\begin{tikzcd} A \arrow[rr, "g"] \arrow[dr, "\psi_Y"'] & & B \arrow[dl, "\phi_Y"] \\ & F(Y) & \end{tikzcd} \tag{56}$$

After verifying that morphisms can be composed, the last two definitions give rise to the category of cones over a diagram F which we denote $\mathrm{Cone}(F)$. Finally, the universal property can be stated in terms of cones, thus giving the general definition of a limit. Indeed, the limit of a diagram F is a cone L over F such that for every cone L' over F, there is a unique cone morphism $! : L' \to L$ called the **mediating morphism**. Equivalently, L is the terminal object of $\mathrm{Cone}(F)$.

Definition D.56 (Limit)**.** Let $F : \mathbf{J} \rightsquigarrow \mathbf{C}$ be a diagram, the **limit** of F, if it exists, is the terminal object of $\mathrm{Cone}(F)$. It is denoted $\lim_{\mathbf{J}} F$ or $\lim F$.

Remark D.57. Often, $\lim F$ also designates the tip of the cone as an object in $\mathbf{C}$ rather than the whole cone.[245] We may also refer to the whole cone as the **limit cone**.

[245] This can sometimes be a source of confusion because many authors omit parts of the proof involving the rest of the cone, and the reader is expected to reconstruct the missing parts.

Examples D.58. In the previous section, we gave three examples of limits: products are limits of discrete diagrams, equalizers are limits of diagrams with two parallel morphisms, and pullbacks are limits of cospans. We let you verify the details, and we add to this list three examples in increasing order of complexity.

1. Consider an empty diagram in $\mathbf{C}$, that is, the functor $\varnothing$ from the empty category to $\mathbf{C}$. A cone over $\varnothing$ is an object $X \in \mathbf{C}_0$, the tip, and nothing else as there are no objects in the diagram. Consequently, a morphism in $\mathrm{Cone}(\varnothing)$ is simply a morphism in $\mathbf{C}$ between the tips, so $\mathrm{Cone}(\varnothing)$ is the same as the original category $\mathbf{C}$ and $\lim\varnothing$ is the terminal object of $\mathbf{C}$ if it exists.[246]

[246] Equivalently, we can say that the terminal object is the product of an empty family.

2. Given a group G, recall from Example B.34.7 that a G–set can be seen as a diagram in **Set**, i.e. a functor $\mathbf{B}G \rightsquigarrow \mathbf{Set}$. We claim that the limit of this diagram is the set $\mathsf{Fix}(S)$ of fixed points of the action (an element s of a G–set is a **fixed point** if $g \cdot s = s$).[247] Let $F : \mathbf{B}G \rightsquigarrow \mathbf{Set}$ be a G–set with $F(*) = S$, a cone over F is a set P along with a function $p : P \to S$ such that for any $g \in G$, (57) commutes.

[247] Recall that the limit of two parallel morphisms was called an equalizer. In this example, we are taking the limit of several parallel morphisms. Thus, one can also see the limit of F as the generalized equalizer of all the morphisms $g \cdot -$ with $g \in G$.

$$\begin{tikzcd} & P \arrow[dl, "p"'] \arrow[dr, "p"] & \\ S \arrow[rr, "F(g)=g\cdot -"'] & & S \end{tikzcd} \tag{57}$$

We infer from this diagram that the image of p is contained in the set of fixed points.[248] Therefore, p factors uniquely through the inclusion $\mathsf{Fix}(S) \hookrightarrow S$. We conclude that the cone formed by $\mathsf{Fix}(S) \hookrightarrow S$ is the limit cone.

[248] For any $x \in P$, we have $g \cdot p(x) = p(x)$.

3. Let x denote an indeterminate variable and k be a field, $k[x]$ denotes the ring of polynomials over x.[249] We will show that $k[\![x]\!]$, the ring of **formal power series**

[249] In Chapter G, we will describe a nice categorical definition of $k[x]$, but, for now, let us assume you know what polynomials are and how they can be added and multiplied together. You can skip this example if you are not familiar with rings.

over x, can be defined as a limit.

Let $I = \langle x \rangle$ be the ideal generated by x, it contains all the polynomials with no constant terms, and denote $I^n = \langle x^n \rangle$. In the sequel, we view elements of $k[x]/I^n$ as polynomials with degree at most $n-1$.[250] The following three key properties are satisfied (we leave the proofs to the interested readers).

[250] More accurately, $k[x]/I^n$ contains equivalence classes of polynomials, but their representatives are exactly the polynomials of degree at most $n-1$. Since $I^0 = k[x]$, the quotien $k[x]/I^0$ is the trivial ring, i.e. the zero object in **Ring**.

a) For any $n \leq m \in \mathbb{N}$ and $p \in k[x]/I^m$, forgetting about all terms in p of degree at least n yields a ring homomorphism $\pi_{m,n} : k[x]/I^m \to k[x]/I^n$.[251]

[251] Note that $\pi_{m,m}$ is the identity.

b) For any $n \in \mathbb{N}$, we can do the same thing for power series to obtain a homomorphism $\pi_{\infty,n} : k[\![x]\!] \to k[x]/I^n$.

c) Any composition of the homomorphisms above can be seen as a single homomorphism above. Namely, $\forall n \leq m \leq l \in \mathbb{N} \cup \infty$,

$$\pi_{m,n} \circ \pi_{l,m} = \pi_{l,n}.$$

Consider the posetal category $(\mathbb{N}, \geq)$, a) and c) imply that $F(n) := k[x]/I^n$ and $F(m \geq n) := \pi_{m,n}$ defines a functor $F : (\mathbb{N}, \geq) \to \mathbf{Ring}$. This is the diagram represented in (58).

$$\cdots \longrightarrow k[x]/I^n \xrightarrow{\pi_{n,n-1}} \cdots \longrightarrow k[x]/I^2 \xrightarrow{\pi_{2,1}} k[x]/I \xrightarrow{\pi_{1,0}} \mathbf{0} \tag{58}$$

Now, using b) and c), we see that $k[\![x]\!]$ along with $\{\pi_{\infty,n}\}_{n\in\mathbb{N}}$ is a cone over the diagram F. It is in fact the terminal cone. Let $\{p_n : R \to k[x]/I^n\}_{n\in\mathbb{N}}$ be another cone over F and $! : R \to k[\![x]\!]$ a morphism of cones. By commutativity, for any $m \leq n$, the coefficients for x^m of $!(r)$ and $p_n(r)$ must agree. Now, by commutativity of the cone $\{p_n\}_{n\in\mathbb{N}}$, $p_n(r)$ and $p_{n-1}(r)$ have the same coefficients except for x^n, thus we can compactly define ! by

$$!(r) := p_0(r) + \sum_{n>0} (p_n(r) - p_{n-1}(r)).$$

This completely determines !, so it is unique.[252]

[252] Existence follows from the same equation.

The construction of this diagram from quotienting different powers of the same ideal is used in different contexts, it is called the **ring completion** of $k[x]$ with respect to I. For instance, one can define the p–adic integers with base ring $\mathbb{Z}$ and the ideal generated by p for any prime p.

Codefinitions

Put simply, a colimit in $\mathbf{C}$ is a limit in $\mathbf{C}^{\mathrm{op}}$. I suggest you spend a bit of time trying to dualize all of the previous section on your own, but it is done below for completeness.

Definition D.59 (Cocone)**.** Let $F : \mathbf{J} \rightsquigarrow \mathbf{C}$ be a diagram. A **cocone** from F to X is an object $X \in \mathbf{C}_0$ along with a family of morphisms $\{\psi_Y : F(Y) \to X\}$ indexed by objects of $Y \in \mathbf{J}_0$ such that for any morphism $a : Y \to Z$ in $\mathbf{J}$, $\psi_Z \circ F(a) = \psi_Y$, i.e. (59) commutes.

$$\begin{tikzcd} F(Y) \arrow[rr, "F(a)"] \arrow[dr, "\psi_Y"'] & & F(Z) \arrow[dl, "\psi_Z"] \\ & X & \end{tikzcd} \tag{59}$$

Often, the terminology cocone under F is used.

Definition D.60 (Morphism of cocones)**.** Let $F : \mathbf{J} \rightsquigarrow \mathbf{C}$ be a diagram and $\{\psi_Y : F(Y) \to A\}_{Y \in \mathbf{J}_0}$ and $\{\phi_Y : F(Y) \to B\}_{Y \in \mathbf{J}_0}$ be two cocones. A **morphism of cocones** from A to B is a morphism $g : A \to B$ in $\mathbf{C}$ such that for any $Y \in \mathbf{J}_0$, $g \circ \psi_Y = \phi_Y$, i.e. (60) commutes.

$$\begin{tikzcd} & F(Y) \arrow[dl, "\psi_Y"'] \arrow[dr, "\phi_Y"] & \\ A \arrow[rr, "g"'] & & B \end{tikzcd} \tag{60}$$

The category of cocones under F is denoted $\mathrm{Cocone}(F)$.

Definition D.61 (Colimit)**.** Let $F : \mathbf{J} \rightsquigarrow \mathbf{C}$ be a diagram, the **colimit** of F denoted $\mathrm{colim} F$, if it exists, is the initial object of $\mathrm{Cocone}(F)$.

Examples D.62. We dualize two examples from the previous section.

1. Dually to Example D.58.1, $\mathrm{colim}\varnothing$ is the is the initial object of $\mathbf{C}$ if it exists.[253]

[253] Equivalently, the initial object is the coproduct of an empty family.

2. Dually to Example D.58.2, we claim that the colimit of the diagram corresponding to a group action is the set of its orbits. Let $F : \mathbf{B}G \rightsquigarrow \mathbf{Set}$ be a G–set with $F(*) = S$, a cocone from F is a set Q along with a function $q : S \to Q$ such that for any $g \in G$, (61) commutes.

   One can also see the colimit of F as the (generalized) coequalizer of all the morphisms $g \cdot -$ with $g \in G$.

$$\begin{tikzcd} S \arrow[rr, "F(g) = g\cdot -"] \arrow[dr, "q"'] & & S \arrow[dl, "q"] \\ & Q & \end{tikzcd} \tag{61}$$

   We infer that if there exists $g \in G$ such that $g \cdot s = s'$, then $q(s) = q(s')$. Denoting $o(s) := \{g \cdot s \mid g \in G\}$ to be the orbit of $s \in S$, the set of orbits of S

$$O := \{o(s) \mid s \in S\}$$

   along with the map $o : S \to O$ forms a cocone from F since $o(g \cdot -) = o$.[254] This cocone is the colimit since for any $q : S \to Q$ as in (61), any $! : O \to Q$ making (62) commute is completely determined by $!(o(s)) = q(s)$ (which is well-defined since $o(s) = o(s') \implies \exists g \in G, g \cdot s = g \cdot s' \implies q(s) = q(s')$).

[254] Since the orbits are, by definition, stable under the action of G.

$$\begin{tikzcd} S \arrow[rr, "g\cdot -"] \arrow[dr, "o"] \arrow[ddr, bend right, "q"'] & & S \arrow[dl, "o"'] \arrow[ddl, bend left, "q"] \\ & O \arrow[d, dashed, "!"] & \\ & Q & \end{tikzcd} \tag{62}$$

3. Let $X = \{x, y\}$, and for each nonzero $n \in \mathbb{N}$, let (X, d_n) denote the metric space where x and y have distance $\frac{1}{n}$ (all other distances must be 0). Since

morphisms in **Met** are nonexpansive functions, for any $m \leq n$, the identity function $(X, d_m) \to (X, d_n)$ is a morphism in **Met**.[255] We assemble all this data in a diagram of shape $(\mathbb{N}, \leq)$ (the opposite of (58)) depicted in (63).

$$(X, d_1) \longrightarrow (X, d_2) \longrightarrow \cdots \longrightarrow (X, d_n) \longrightarrow \cdots \tag{63}$$

[255] We have
$$d_m(x,y) = \frac{1}{m} \geq \frac{1}{n} = d_n(x,y).$$

$$\begin{tikzcd} (X, d_m) \arrow[rr] \arrow[dr, "\psi_m"'] & & (X, d_n) \arrow[dl, "\psi_n"] \\ & (L, d) & \end{tikzcd} \tag{64}$$

Recall the one point space $(\{*\}, d_{\mathbf{1}})$ is the terminal object $\mathbf{1}$ in **Met** (Example C.37). The family $\{!_n : (X, d_n) \to \mathbf{1}\}$ comprising the unique morphisms to $\mathbf{1}$ is a cocone under (63), and we claim it is the colimit cocone.

Suppose $\psi_n : (X, d_n) \to (L, d)$ is a cocone under (63). Instantiating (59), we find that (64) commutes, hence $\psi_m(x) = \psi_n(x)$ and $\psi_m(y) = \psi_n(y)$ for every $m, n \in \mathbb{N}$. We can give one name ψ to the function $X \to L$ that underlies all ψ_n. For any $n \in \mathbb{N}$, the distance between $\psi(x)$ and $\psi(y)$ is bounded above by $\frac{1}{n}$, otherwise $\psi_n : (X, d_n) \to (L, d)$ would not be nonexpansive. Therefore, the distance can only be 0, and we conclude $\psi(x) = \psi(y)$.

A morphism of cocones f from $\{!_n\}$ to $\{\psi_n\}$ must satisfy $f(!_n(x)) = \psi_n(x) = \psi_n(y)$, so the only possible choice is the function sending $*$ to $\psi(x) = \psi(y)$.

OL Exercise D.63 (Trivial (co)limits)**.** Show the following (co)limits always exist and find what they are.

1. The limit of a diagram with only one morphism.
2. The colimit of a diagram with only one morphism.
3. The limit of a span.
4. The colimit of a cospan.

Instantiating our examples (co)limits in posets was rather simple because they are thin categories, and every diagram in a thin category is commutative. This generalizes to all (co)limits.

OL Exercise D.64. Let **C** be a posetal category. Show that the limit (resp. colimit) of any diagram $F : \mathbf{J} \rightsquigarrow \mathbf{C}$ is the infimum (resp. supremum) of all points in the image of F.

Results

Proposition D.65 (Uniqueness)**.** *Let $F : \mathbf{J} \rightsquigarrow \mathbf{C}$ be a diagram, the limit (resp. colimit) of F, if it exists, is unique up to unique isomorphism.*

Proof. This follows from the uniqueness of terminal (resp. initial) objects.[256] □

[256] Corollary C.34 (resp. Proposition C.33).

Remark D.66. The isomorphism between two limits (also colimits) is unique when viewed as a morphism of cone. There might exists an isomorphism between the tips that is not a morphism of cone. For instance, let A, B and C be finite sets. One can check that both $A \times (B \times C)$ and $(A \times B) \times C$ are products of $\{A, B, C\}$ (with the usual projection maps). Thus, there is an isomorphism between them. One can check that, for it to be a morphism of cones, it must send $(a,(b,c))$ to $((a,b),c)$, but any other bijection between them is an isomorphism in **Set**.

For this reason, the limit really consists of the whole cone, and not just of the object at the tip. Unfortunately, this subtlety is not well cared for in the literature and it can and has led to errors.

Recall the definition of preserve and reflect we gave in Definition C.48. With the framework of (co)limits, we can give more formal related definitions.

SOL Exercise D.67 (NOW!)**.** Let $F : \mathbf{C} \rightsquigarrow \mathbf{C}'$ be a functor and $D : \mathbf{J} \rightsquigarrow \mathbf{C}$ be a diagram. The composition $F \circ D$ is a diagram of shape $\mathbf{J}$ in $\mathbf{C}'$. Show that sending a cone $\{\psi_X : A \to DX\}_{X \in \mathbf{J}_0}$ over F to $\{F\psi_X : FA \to FDX\}_{X \in \mathbf{J}_0}$ is a functor $F_D : \mathrm{Cone}(D) \rightsquigarrow \mathrm{Cone}(F \circ D)$. Dually, construct the functor $F^D : \mathrm{Cocone}(D) \rightsquigarrow \mathrm{Cocone}(F \circ D)$.

In words, $F \circ D$ is the diagram D where we applied F to all objects and morphisms. Then, F_D takes a cone over D and applies F to every object and morphism in it to obtain a cone over $F \circ D$.[257] This allows us to define preservation and reflection of (co)limits, as well as creation.

[257] Similarly for F^D.

Definition D.68. Let $F : \mathbf{C} \rightsquigarrow \mathbf{C}'$ be a functor and $\mathbf{J}$ be a category.

- We say that F **preserves** limits of shape $\mathbf{J}$ if for any diagram $D : \mathbf{J} \rightsquigarrow \mathbf{C}$, if $\{\psi_X\}_{X \in \mathbf{J}_0}$ is the limit cone over D, then $\{F\psi_X\}_{X \in \mathbf{J}_0}$ is the limit cone over $F \circ D$. In other words, for any D, F_D preserves (in the sense of Definition C.48) terminal objects.[258]
- We say that F **reflects** limits of shape $\mathbf{J}$ if for any diagram $D : \mathbf{J} \rightsquigarrow \mathbf{C}$, if $\{\psi_X\}_{X \in \mathbf{J}_0}$ is a cone over D and $\{F\psi_X\}_{X \in \mathbf{J}_0}$ is the limit cone over $F \circ D$, then $\{\psi_X\}_{X \in \mathbf{J}_0}$ is also the limit cone over D. In other words, for any D, F_D reflects (in the sense of Definition C.48) terminal objects.
- We say that F **creates** limits of shape $\mathbf{J}$ if for any diagram $D : \mathbf{J} \rightsquigarrow \mathbf{C}$, if $\{\phi_X\}_{X \in \mathbf{J}_0}$ is a limit cone over $F \circ D$, then there exists a unique cone over D $\{\psi_X\}_{X \in \mathbf{J}_0}$ such that $F\psi_X = \phi_X$ and $\{\psi_X\}_{X \in \mathbf{J}_0}$ is a limit cone.

[258] We will often be less rigorous and write something like $\lim(F \circ D) = F(\lim_{\mathbf{J}} D)$. For instance, we will say that F preserves binary products if $FX \times FY = F(X \times Y)$ or $FX \times FY \cong F(X \times Y)$, but what we actually need to check is that $F\pi_X$ and $F\pi_Y$ satisfy the universal property of the product.

We leave to you the dualization of this definition.[259]

[259] Replace cone by cocone and limit by colimit.

These are more technical and rigorous than our previous notions of preservation and reflection of properties, but the intuition should stay the same. In practice, preservation is used way more often,[260] so let us practice a bit.

[260] In this book, we will not use the other two.

Example D.69. Recall from Exercise B.39 that we have two functors $(-)_0$ and $(-)_1$ from **Cat** to **Set**. It follows from the definition of product categories that both preserve products. Indeed the objects of $\mathbf{C} \times \mathbf{D}$ are pairs of objects in $\mathbf{C}_0 \times \mathbf{D}_0$, and morphisms of $\mathbf{C} \times \mathbf{D}$ are pairs of morphisms in $\mathbf{C}_1 \times \mathbf{D}_1$, so

$$(\mathbf{C} \times \mathbf{D})_0 = \mathbf{C}_0 \times \mathbf{D}_0 \text{ and } (\mathbf{C} \times \mathbf{D})_1 = \mathbf{C}_1 \times \mathbf{D}_1.$$

SOL Exercise D.70. Show that if F preserves pullbacks (i.e.: F preserves limits of cospans), then F preserves monomorphisms. State and prove the dual statement.

OL **Exercise D.71.** Show that if $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ is an isomoprhism, then F preserves and reflects (co)limits of all shape.

As we already hinted at, oftentimes, forgetful functors preserve limits,[261] we let you prove a very specific instance of this.

[261] Due to results in Chapter H.

OL **Exercise D.72.** Let $U : \mathbf{Set}_* \rightsquigarrow \mathbf{Set}$ be the forgetful functor from pointed sets to sets. Show that U preserves products, equalizers and pullbacks.

OL **Exercise D.73.** Fix $A \in \mathbf{C}_0$, show that the functor $\mathrm{Hom}_{\mathbf{C}}(A, -)$ preserves binary products. Namely, if $X, Y \in \mathbf{C}_0$ and $X \times Y$ exists, then

$$\mathrm{Hom}_{\mathbf{C}}(A, X \times Y) \cong \mathrm{Hom}_{\mathbf{C}}(A, X) \times \mathrm{Hom}_{\mathbf{C}}(A, Y).$$

Corollary D.74 (Dual)**.** *Fix $A \in \mathbf{C}_0$, show the functor $\mathrm{Hom}_{\mathbf{C}}(-, A)$ preserves binary coproducts when viewed as a functor $\mathbf{C} \rightsquigarrow \mathbf{Set}^{\mathrm{op}}$, i.e.:*

$$\mathrm{Hom}_{\mathbf{C}}(X + Y, A) \cong \mathrm{Hom}_{\mathbf{C}}(X, A) \times \mathrm{Hom}_{\mathbf{C}}(Y, A).$$

These last two results are strenghtened in Theorem D.88 and Corollary D.89. We are not done proving things about (co)limits, but we move on to the next section where we will do these proofs using diagram chasing.

D.3 Diagram Chasing

We show four results in increasing order of complexity to demonstrate diagram chasing through examples.

Proposition D.75. *Let $\{f_i, g_i : X_i \to Y_i\}_{i \in I}$ be a familiy of parallel morphisms in $\mathbf{C}$ such that for any $i \in I$, (65) is an equalizer, then (66) is an equalizer.*

$$\begin{tikzcd} E_i \arrow[r, "e_i"] & X_i \arrow[r, shift left, "f_i"] \arrow[r, shift right, "g_i"'] & Y_i \end{tikzcd} \qquad (65)$$

$$\begin{tikzcd} \prod_{i \in I} E_i \arrow[r, "\prod_{i \in I} e_i"] & \prod_{i \in I} X_i \arrow[r, shift left, "\prod_{i \in I} f_i"] \arrow[r, shift right, "\prod_{i \in I} g_i"'] & \prod_{i \in I} Y_i \end{tikzcd} \qquad (66)$$

Proof. Suppose $o : O \to \prod_{i \in I} X_i$ also equalizes $\prod f_i$ and $\prod g_i$. We have the following implications.[262]

[262] The second implication uses (36).

$$\begin{aligned} o \circ \prod f_i = o \circ \prod g_i &\implies \pi_i \circ \prod f_i \circ o = \pi_i \circ \prod g_i \circ o \\ &\implies f_i \circ \pi_i \circ o = g_i \circ \pi_i \circ o \end{aligned}$$

Consequently, for each $i \in I$, $\pi_i \circ o$ equalizes f_i and g_i, so it factors uniquely through e_i: $\pi_i \circ o = e_i \circ !_i$ as depicted in (**??**). The universal property of the product allows us to form the pairing $\langle !_i \rangle_{i \in I} : O \to \prod_{i \in I} E_i$, and we have the following derivation.

$$\begin{tikzcd} O \arrow[d, dashed, "!_i"'] \arrow[dr, "\pi_i \circ o"] & & \\ E_i \arrow[r, "e_i"'] & X_i \arrow[r, shift left, "f_i"] \arrow[r, shift right, "g_i"'] & Y_i \end{tikzcd} \qquad (67)$$

$$\begin{aligned} \pi_i \circ \prod e_i \circ \langle !_i \rangle &= e_i \circ \pi_i \circ \langle !_i \rangle \\ &= e_i \circ !_i \\ &= \pi_i \circ o \end{aligned}$$

We conclude from the universal property of $\prod X_i$ that $o = \prod e_i \circ \langle !_i \rangle$ as depicted in (68). It remains to show $\langle !_i \rangle$ is unique with this property.

$$\begin{tikzcd} O \arrow[d, dashed, "\langle !_i \rangle"'] \arrow[dr, "o"] & & \\ \prod_{i\in I} E_i \arrow[r, "\prod_{i\in I} e_i"'] & \prod_{i\in I} X_i \arrow[r, shift left, "\prod_{i\in I} f_i"] \arrow[r, shift right, "\prod_{i\in I} g_i"'] & \prod_{i\in I} Y_i \end{tikzcd} \tag{68}$$

If $m : O \to \prod E_i$ satisfies $\prod e_i \circ f = o$, then

$$e_i \circ \pi_i \circ f = \pi_i \circ \prod e_i \circ f = \pi_i \circ o,$$

but uniqueness of $!_i$ ensures $\pi_i \circ f = !_i$ (they both make (67) commute). This also means $f = \langle !_i \rangle_{i \in I}$, so we are done. □

Corollary D.76 (Dual). *Let $\{f_i, g_i : X_i \to Y_i\}_{i\in I}$ be a familiy of parallel morphisms in* **C** *such that for any $i \in I$, $d_i : Y_i \to D_i$ is the coequalizer of f_i and g_i, then $\coprod d_i$ is the coequalizer of $\coprod f_i$ and $\coprod g_i$.*

One might summarize these results by saying that the product of equalizers is the equalizer of products,[263] and this is telling of a general fact about limits interacting with limits (dually colimits interacting with colimits), see Theorem **??** (Corollary **??**).

[263] Dually, the coproduct of coequalizers is the coequalizer of the coproducts.

Theorem D.77. *Consider the pullback square in* (69).

$$\begin{tikzcd} A \times_C B \arrow[r, "p_B"] \arrow[d, "p_A"'] \arrow[dr, phantom, "\lrcorner", very near start] & B \arrow[d, "g"] \\ A \arrow[r, "f"'] & C \end{tikzcd} \tag{69}$$

If g is monic, then p_A also is. Symmetrically, if f is monic, then p_B also is.[264]

[264] This is commonly stated simply as: "The pullback of a monomorphism is a monomorphism."

Proof. Let $h_1, h_2 : X \to A \times_C B$ be such that $p_A \circ h_1 = p_A \circ h_2$, we need to show that $h_1 = h_2$. First, observe that h_1 and h_2 yield two cones over the cospan $A \xrightarrow{f} C \xleftarrow{g} B$ as depicted in (70).

The two cones are

$$\begin{tikzcd} X \arrow[r, "p_B \circ h_1"] \arrow[d, "p_A \circ h_1"'] & B \\ A & \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} X \arrow[r, "p_B \circ h_2"] \arrow[d, "p_A \circ h_2"'] & B \\ A & \end{tikzcd}$$

They make the squares commute because the original pullback square commutes.

$$\begin{tikzcd} X \arrow[drr, bend left=40, "p_B \circ h_2"] \arrow[drr, bend left=15, "p_B \circ h_1"'] \arrow[dr, shift left, "h_2"] \arrow[dr, shift right, "h_1"'] \arrow[ddr, bend right, "p_A \circ h_1 = p_A \circ h_2"'] & & \\ & A \times_C B \arrow[r, "p_B"] \arrow[d, "p_A"'] \arrow[dr, phantom, "\lrcorner", very near start] & B \arrow[d, "g"] \\ & A \arrow[r, "f"'] & C \end{tikzcd} \tag{70}$$

Furthermore, h_1 and h_2 are cone morphisms between X and $A \times_C B$ and since the pullback is the terminal cone over this cospan, they are unique. Now, we already have that the projections onto A is the same for both new cones, but we claim this is also true for the projections onto B. Indeed, because g is monic and the square commutes, we have the following implications.

$$\begin{aligned} p_A \circ h_1 = p_A \circ h_2 &\implies f \circ p_A \circ h_1 = f \circ p_A \circ h_2 \\ &\implies g \circ p_B \circ h_1 = g \circ p_B \circ h_2 \\ &\implies p_B \circ h_1 = p_B \circ h_2 \end{aligned}$$

In other words, the two new cones are in fact the same cones, hence h_1 and h_2 are the same morphisms by uniqueness, which concludes our proof. □

Corollary D.78 (Dual)**.** *The pushout of an epimorphism is an epimorphism.*

Theorem D.79 (Pasting Lemma)**.** *Consider* (71)*, where the right square is a pullback.*

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "\alpha"'] & B \arrow[r, "g"] \arrow[d, "\beta"'] \arrow[dr, phantom, "\lrcorner", very near start] & C \arrow[d, "\gamma"] \\ A' \arrow[r, "f'"'] & B' \arrow[r, "g'"'] & C' \end{tikzcd} \tag{71}$$

If (71) *commutes, the left square is a pullback if and only if the rectangle is.*[265]

[265] This result is called the **pasting lemma**.

$$\begin{tikzcd} A \arrow[r, "g \circ f"] \arrow[d, "\alpha"'] & C \arrow[d, "\gamma"] \\ A' \arrow[r, "g' \circ f'"'] & C' \end{tikzcd} \tag{72}$$

Proof. ($\Rightarrow$) Explicitly, we have to show that $\alpha : A' \leftarrow A \to C : g \circ f$ is the pullback of $g' \circ f' : A' \to C' \leftarrow C : \gamma$, i.e., that (72) is a pullback square. The commutativity $g' \circ f' \circ \alpha = \gamma \circ g \circ f$ implies this is already a cone over the cospan we just described. Now, suppose there is another cone over this cospan, namely, there exist morphisms $p_{A'} : X \to A'$ and $p_C : X \to C$ satisfying $g' \circ f' \circ p_{A'} = \gamma \circ p_C$ as depicted in (73).

$$\begin{tikzcd} X \arrow[drr, dashed, "!_B"] \arrow[dr, dashed, "!_A"'] \arrow[drrr, bend left, "p_C"] \arrow[ddr, bend right, "p_{A'}"'] & & & \\ & A \arrow[r, "f"] \arrow[d, "\alpha"'] \arrow[dr, phantom, "\lrcorner", very near start] & B \arrow[r, "g"] \arrow[d, "\beta"'] \arrow[dr, phantom, "\lrcorner", very near start] & C \arrow[d, "\gamma"] \\ & A' \arrow[r, "f'"'] & B' \arrow[r, "g'"'] & C' \end{tikzcd} \tag{73}$$

Notice that composing $p_{A'}$ with f', we obtain a cone over the cospan in the right square and by universality of B, this yields a unique morphism $!_B : X \to B$ satisfying $g \circ !_B = p_C$ and $\beta \circ !_B = f' \circ p_{A'}$. This second equality yields cone over the cospan in the left square, thus we get a unique morphism $!_A : X \to A$ satisfying $\alpha \circ !_A = p_{A'}$ and $f \circ !_A = !_B$. Composing the last equality with g, we get

$$g \circ f \circ !_A = g \circ !_B = p_C,$$

showing that $!_A$ is a morphism of cones over the rectangular cospan.

What is more, any other morphism $m : X \to A$ of cones over this cospan must satisfy

$$g \circ f \circ m = p_C \text{ and } \beta \circ f \circ m = f' \circ \alpha \circ m = f' \circ p_{A'},$$

and thus, $f \circ m$ is a morphism of cones over the cospan in the right rectangle. By uniqueness, $f \circ m = !_B$, so m is also a morphism of cones over the cospan in the left square, and by universality of A, $m = !_A$.

($\Leftarrow$) Explicitly, we have to show that $\alpha : A' \leftarrow A \to B : f$ is the pullback of

$f' : A' \to B \leftarrow B : \beta$.

$$\begin{tikzcd}
X \arrow[dr, dashed, "!_A"'] \arrow[drr, bend left, "p_B"] \arrow[ddr, bend right, "p_{A'}"'] & & & \\
& A \arrow[r, "f"] \arrow[d, "\alpha"'] & B \arrow[r, "g"] \arrow[d, "\beta"'] \arrow[dr, phantom, "\lrcorner", very near start] & C \arrow[d, "\gamma"] \\
& A' \arrow[r, "f'"'] & B' \arrow[r, "g'"'] & C'
\end{tikzcd} \tag{74}$$

Let $p_{A'} : A' \leftarrow X \to B : p_B$ be a cone over the cospan of the left square (i.e. $\beta \circ p_B = f' \circ p_{A'}$). The commutativity of (71) implies $p_{A'} : A' \leftarrow X \to C : g \circ p_B$ is a cone over the rectangle cospan, then by universality, there exists a unique $!_A : X \to A$ such that $g \circ f \circ !_A = g \circ p_B$ and $\alpha \circ !_A = p_A$. Moreover, with the commutativity of the left square, we find that $f \circ !_A$ is a morphism of cones over the right cospan satisfying $\beta \circ f \circ !_A = f' \circ \alpha \circ !_A = f' \circ p_{A'} = \beta \circ p_B$ and $g \circ f \circ !_A = g \circ p_B$. But since our hypothesis on $p_{A'}$ and p_B implies p_B is a morphism of cones satisfying the same equations, by universality of B, $p_B = f \circ !_A$. Therefore, $!_A$ is a morphism of cone over the left cospan.

Finally, if $m : X \to A$ also satisfies $\alpha \circ m = p_{A'}$ and $f \circ m = p_B$. We find in particular that m is a morphism of cones over the rectangle cospan, hence by universality, $m = !_A$. □

Corollary D.80 (Dual)**.** *If* (75) *commutes, the right square is a pushout if and only if the rectangle is.*

$$\begin{tikzcd}
A \arrow[r, "f"] \arrow[d, "\alpha"'] & B \arrow[r, "g"] \arrow[d, "\beta"'] & C \arrow[d, "\gamma"] \\
A' \arrow[r, "f'"'] \arrow[ur, phantom, "\ulcorner", very near end] & B' \arrow[r, "g'"'] & C'
\end{tikzcd} \tag{75}$$

OL **Exercise D.81.** Show that (76) is a pullback square. Let $i : A' \to A$ be an isomorphism, show that (77) is a pullback square.[266]

$$\begin{tikzcd}
A \arrow[r, "\mathrm{id}_A"] \arrow[d, "f"'] \arrow[dr, phantom, "\lrcorner", very near start] & A \arrow[d, "f"] \\
B \arrow[r, "\mathrm{id}_B"'] & B
\end{tikzcd} \tag{76} \qquad
\begin{tikzcd}
A' \arrow[r, "i"] \arrow[d, "f \circ i"'] \arrow[dr, phantom, "\lrcorner", very near start] & A \arrow[d, "f"] \\
B \arrow[r, "\mathrm{id}_B"'] & B
\end{tikzcd} \tag{77}$$

[266] We can summarize the first square by saying that the pullback of any morphism along the identity gives back the original morphism. The second square is basically a converse to the statement "pullbacks are unique up to isomorphism" in this very special case.

Definition D.82 ((Co)completeness)**.** A category is said to be **(co)complete** (resp. **finitely** (co)complete) if any small (resp. finite) diagram has a (co)limit.

Theorem D.83. *Suppose that a category* $\mathbf{C}$ *has all products and equalizers then* $\mathbf{C}$ *has all limits, i.e.* $\mathbf{C}$ *is complete.*

Proof. Let $F : \mathbf{J} \rightsquigarrow \mathbf{C}$ be a diagram, we will show that the limit of F is obtained from the equalizer of two morphisms[267]

[267] Recall that s and t denote the sources and targets of morphisms.

$$u_1, u_2 : \prod_{X \in \mathbf{J}_0} F(X) \to \prod_{a \in \mathbf{J}_1} F(t(a)),$$

which are defined below. The equalizer and the products it involves exist by hypothesis.

First, let us try to explain the intuition behind this construction. The limit of F is the terminal cone over F. In particular, it is a cone over F, namely, a family of morphisms $\psi_X : \lim F \to FX$ indexed by $X \in \mathbf{J}_0$ such that for any $a : X \to Y \in \mathbf{J}_1$, $Fa \circ \psi_X = \psi_Y$. Since $\mathbf{C}$ has products, we can also specify the morphisms in the cone by a single morphism $\psi : \lim F \to \prod_{X\in\mathbf{J}_0} FX$.[268]

[268] The family $\{\psi_X\}$ gives rise to ψ by the universal property of the product and ψ gives rise to the family by post-composing with the projections $\pi_X : \prod_{X\in\mathbf{J}_0} FX \to FX$.

$$\psi_X = \pi_X \circ \psi$$

We also could write $\psi = \langle \psi_X \rangle_{X\in\mathbf{J}_0}$.

The additional property of the cone is now $\forall a : X \to Y \in \mathbf{J}_1, Fa \circ \pi_X \circ \psi = \pi_Y \circ \psi$. Replacing the objects X and Y with $s(a)$ and $t(a)$ respectively, we obtain two families of morphisms

$$\{Fa \circ \pi_{s(a)} : \prod_{X\in\mathbf{J}_0} FX \to Ft(a) \mid a \in \mathbf{J}_1\} \quad \text{and} \quad \{\pi_{t(a)} : \prod_{X\in\mathbf{J}_0} FX \to Ft(a) \mid a \in \mathbf{J}_1\}.$$

The universal property of products yields two parallel morphisms $u_1, u_2 : \prod_{X\in\mathbf{J}_0} FX \to \prod_{a\in\mathbf{J}_1} Ft(a)$ making (78) commute.

$$
\begin{tikzcd}
\prod_{X\in\mathbf{J}_0} FX \arrow[dr, dashed, "u_1"] \arrow[drr, bend left, "Fa\circ\pi_{s(a)}"] & & \\
& \prod_{a\in\mathbf{J}_1} Ft(a) \arrow[r, "\pi_a"] & Ft(a) \\
\prod_{X\in\mathbf{J}_0} FX \arrow[ur, dashed, "u_2"] \arrow[urr, bend right, "\pi_{t(a)}"'] & &
\end{tikzcd} \tag{78}
$$

We find that ψ equalizes u_1 and u_2.[269] Since we did not use the fact that ψ is terminal yet, any cone over F yields a morphism from the tip to the product $\prod_{X\in\mathbf{J}_0} FX$ that equalizes u_1 and u_2. Moreover, this process can be reversed, hence any morphism that equalizes u_1 and u_2 corresponds to a cone over F.

[269] We check that $u_1 \circ \psi = u_2 \circ \psi$ by post-composing with π_a for every $a \in \mathbf{J}_1$. Indeed, we have

$$
\begin{aligned}
\pi_a \circ u_1 \circ \psi &= Fa \circ \pi_{s(a)} \circ \psi \\
&= \pi_{t(a)} \circ \psi && \text{(def. of } \psi\text{)} \\
&= \pi_a \circ u_2 \circ \psi,
\end{aligned}
$$

and the universal property of $\prod_{a\in\mathbf{J}_1} Ft(a)$ implies $u_1 \circ \psi = u_2 \circ \psi$.

We are on a good track because we have shown that cones over F are in correspondence with cones over the parallel morphisms u_1 and u_2. If we can show there is also a correspondence between the morphisms of such cones, we will be able to conclude that the terminal cone over u_1 and u_2 (i.e. their equalizer) is the terminal cone over F (i.e. the limit of F).[270]

[270] More abstractly, we show there is an isomorphism between the categories $\mathrm{Cone}(F)$ and $\mathrm{Cone}(U)$, where U is the diagram with only two parallel morphisms sent to u_1 and u_2. One can check that isomorphisms of categories preserve terminal objects (Exercise D.71), so the equalizer of u_1 and u_2 is the limit of F.

$$
\begin{tikzcd}
A \arrow[d, "g"'] \arrow[dr, "\psi"] & & \\
{} & \prod_{X\in\mathbf{J}_0} FX \arrow[r, dashed, shift left, "u_1"] \arrow[r, dashed, shift right, "u_2"'] & \prod_{a\in\mathbf{J}_1} Ft(a) \\
B \arrow[ur, "\phi"'] & &
\end{tikzcd} \tag{79}
$$

Let $\{\psi_X, \phi_X : A \to FX\}_{X\in\mathbf{J}_0}$ be two cones over F, $g : A \to B$ be a morphism of cones, and ψ and ϕ be the corresponding morphism that equalize u_1 and u_2. We will show that (79) commutes. By definition of g, we have $\phi_X \circ g = \psi_X$ for any $X \in \mathbf{J}_0$, which we can rewrite as $\pi_X \circ \phi \circ g = \pi_X \circ \psi$. By the universal property of the product $\prod_{X\in\mathbf{J}_0} FX$, we conclude that $\phi \circ g = \psi$.

Conversely, given g that makes (79) commute, g is a morphism of cones over F because for any $X \in \mathbf{J}_0$, $\phi_X \circ g = \pi_X \circ \phi \circ g = \pi_X \circ \psi = \psi_X$.

In conclusion, let $\psi : L \to \prod_{X\in\mathbf{J}_0}$ be the equalizer of u_1 and u_2, the limit of F is the cone $\{\pi_X \circ \psi_X\}_{X\in\mathbf{J}_0}$.

□

Remark D.84. The same proof yields a more general statement: For any cardinal κ, if a category $\mathbf{C}$ has all products of size less than κ and equalizers, then it has limits of any diagram with less than κ objects and morphisms.

Corollary D.85 (Dual)**.** *If a category $\mathbf{C}$ has all coproducts of size less than κ and coequalizers, then it has colimits of any diagram with less than κ objects and morphisms.*

Definition D.86. A functor $\mathbf{C} \rightsquigarrow \mathbf{D}$ is said to be (**finitely**) **(co)continuous** if it preserves all (finite) (co)limits.

OL **Exercise D.87.** Show that a functor is continuous if and only if it preserves products and equalizers. State and prove the dual statement.

Theorem D.88. *Fix $A \in \mathbf{C}_0$, the functor $\mathrm{Hom}_{\mathbf{C}}(A, -)$ is continuous.*

Proof. We could show that $\mathrm{Hom}_{\mathbf{C}}(A,-)$ preserves equalizers and use Exercises D.73 and D.87, but the direct proof is not very long and it lets us get even more familiar with cones.

Let $D : \mathbf{J} \rightsquigarrow \mathbf{C}$ be a diagram and $\{\psi_X : \lim D \to DX\}_{X\in\mathbf{J}_0}$ be the limit cone, we need to show that $\{\psi_X \circ - : \mathrm{Hom}_{\mathbf{C}}(A, \lim D) \to \mathrm{Hom}_{\mathbf{C}}(A, DX)\}_{X\in\mathbf{J}_0}$ is a limit cone.

$$\begin{tikzcd} & \mathrm{Hom}_{\mathbf{C}}(A, DX) \arrow[dd, "Da\circ-"] \\ \mathrm{Hom}_{\mathbf{C}}(A, \lim D) \arrow[ur, "\psi_X\circ-"] \arrow[dr, "\psi_Y\circ-"'] & \\ & \mathrm{Hom}_{\mathbf{C}}(A, DY) \end{tikzcd} \tag{80}$$

First, for any $a : X \to Y \in \mathbf{J}_1$, we have $Da \circ \psi_X = \psi_Y$, which implies (80) commutes. Hence, $\{\psi_X \circ -\}_{X\in\mathbf{J}_0}$ is a cone over $\mathrm{Hom}_{\mathbf{C}}(A, D-)$.

Next, if $\{\phi_X : T \to \mathrm{Hom}_{\mathbf{C}}(A, DX)\}_{X\in\mathbf{J}_0}$ is another cone over $\mathrm{Hom}_{\mathbf{C}}(A, D-)$, then observe that any $t \in T$ gives rise to a cone over D $\{\phi_X(t) : A \to DX\}_{X\in\mathbf{J}_0}$. Indeed, we have

$$Df \circ \phi_X(t) = ((Df \circ -) \circ \phi_X)(t) = \phi_Y(t).$$

We obtain a unique morphism of cones $g(t) : A \to \lim D$ making (81) commute for all $X \in \mathbf{J}_0$. This yields a function $g : T \to \mathrm{Hom}_{\mathbf{C}}(A, \lim D)$ that is a morphism of cones because combining (81) for every $t \in T$ yields $(\psi_X \circ -) \circ g = \phi_X$.

$$\begin{tikzcd} A \arrow[dd, "g(t)"'] \arrow[dr, "\phi_X(t)"] & \\ & DX \\ \lim D \arrow[ur, "\psi_X"'] & \end{tikzcd} \tag{81}$$

If $g' : T \to \mathrm{Hom}_{\mathbf{C}}(A, \lim D)$ is another morphism of cones, then we must have that $g'(t)$ also makes (81) for all $X \in \mathbf{J}_0$.[271] Therefore, $g'(t) : A \to \lim D$ is a morphism of cones and since $\lim D$ is terminal, we conclude $g'(t) = g(t)$ and $g' = g$. $\square$

[271] We have
$$\psi_X \circ g'(t) = ((\psi \circ -) \circ g')(t) = \phi_X(t).$$

Corollary D.89 (Dual)**.** *Fix $A \in \mathbf{C}_0$, the functor $\mathrm{Hom}_{\mathbf{C}}(-, A)$ is continuous.*[272]

[272] More concisely, the Hom bifunctor is continuous in each argument.

OL **Exercise D.90.** Show that a category with all pullbacks and a terminal object is finitely complete.

Corollary D.91 (Dual)**.** *A category with all pushouts and an initial object is finitely cocomplete.*

Remark D.92. We can conclude[273] that a functor is finitely continuous if and only if it preserves pullbacks and the terminal object and it is finitely coconituous if and only if it preserves pushouts and the initial object.

[273] Similarly to Exercise D.87.

E Universal Properties

We continue our exploration of universal constructions. This chapter is arranged like the previous one, we give lots of examples before abstracting away to define universal properties.[274] This abstracting step involves a new concept: comma categories, which are interesting in their own right.

[274] I estimate we have done enough diagram chasing, so we will not prove as much results as we did in Chapter D.

E.1 Examples

Free Monoid

The construction of a *free* object is common to different fields of mathematics. Informally, when **C** is a category whose objects are objects of another category **D** equipped with extra structure (e.g. **C** is a concrete category and $\mathbf{D} = \mathbf{Set}$), the free **C**–object over a **D**–object X carries the least amount of structure possible to be considered a part of **C** while *containing* X.

The example we will carry out in **Mon** can be carried out in many other categories like **Grp**, **Ab**, **Ring**, etc. We choose **Mon** because the concrete characterization of a free monoid is simple.

Definition E.1 (Classical)**.** The **free monoid** on a set A, denoted by A^*, is the set of finite words with symbols in A with the multiplication being concatenation of words and identity being the **empty word** ε.[275]

[275] Examples of finite words in $\{\mathtt{a},\mathtt{b},\mathtt{c}\}^*$ are `a`, `ab`, `abc`, `accabac`, etc. The concatenation of `abc` and `aacb` is `abcaacb`.

An intuitive way to see A^* is that it is the *smallest* monoid that contains A. We start from single-letter words which are just elements of A, and then generate the rest by concatenating bigger and bigger words together (before finally adding ε).

In order to give a categorical characterization, we need to look at homomorphisms from or into the free monoid. Notice that any homomorphism $h^* : A^* \to M$ is completely determined by where h^* sends single-letter words, i.e., elements of A. Indeed, in order to satisfy the homomorphism property, we must have for any $\mathtt{a},\mathtt{b} \in A$,

$$h^*(\mathtt{ab}) = h^*(\mathtt{a}) \cdot h^*(\mathtt{b}) \text{ and } h^*(\varepsilon) = 1_M.$$

In general, the unique homomorphism sending $\mathtt{a} \in A$ to $h(\mathtt{a})$ can be defined recursively:

$$h^*(w) = \begin{cases} h(\mathtt{a}) \cdot h^*(w') & \mathtt{a} \in A, w \in A^*, w = \mathtt{a}w' \\ 1_M & w = \varepsilon \end{cases}.$$

Concisely, for any function $h : A \to M$, there is a unique homomoprhisms $h^* : A^* \to M$ that sends $\mathtt{a}$ to $h(\mathtt{a})$. We call this fact the universal property of the free monoid.

We repeated several times that universal properties should determine an object up to isomorphism, let us check this. Suppose that a monoid N contains A and satisfies the same property, that is for any (set-theoretic) function $h : A \to M$, there is a unique homomorphism $h_N^* : N \to M$ with $h_N^*(\mathtt{a}) = h(\mathtt{a})$. We claim that N and A^* are isomorphic.

If we take $M = A^*$, and $h : A \to A^* = a \mapsto a$, then we get a homomorphism $h_N^* : N \to A^*$ using the property for N. If we take $M = N$ and the inclusion $i : A \hookrightarrow N$, then the property of A^* yields a homomorphism $i^* : A^* \to N$. By construction, $h_N^* \circ i^* : A^* \to A^*$ and $i^* \circ h_N^* : N \to N$ are both homomorphisms that send $\mathtt{a}$ to $\mathtt{a}$.[276] Note that $\mathrm{id}_{A^*} : A^* \to A^*$ and $\mathrm{id}_N : N \to N$ are also homomorphisms sending $\mathtt{a}$ to $\mathtt{a}$. By the uniqueness in the universal property, we conclude

$$h_N^* \circ i^* = \mathrm{id}_{A^*} \text{ and } i^* \circ h_N^* = \mathrm{id}_N,$$

that is, A^* and N are isomorphic.

[276] Recall that both A^* and N contains all elements in A.

The universal property we gave above determined the free monoid up to isomorphism, so we are happy to make this into a definition. However, this definition cannot take place entirely in the category **Mon**. We had to implicitly rely on the fact that a monoid has an underlying set and homomorphisms are just functions satisfying additional properties. Our categorical definition thus relies on the forgetful functor $U : \mathbf{Mon} \rightsquigarrow \mathbf{Set}$.

Definition E.2 (Categorical)**.** The free monoid of a set A is an object A^* in **Mon** along with a *canonical inclusion* $i : A \to U(A^*)$ that satisfies the following universal property: for any monoid M and function $h : A \to U(M)$, there exists a unique homomorphism $h^* : A^* \to M$ such that $U(h^*) \circ i = h$, namely, $h^*(i(a)) = h(a)$. This is summarized in (82).[277]

$$\begin{tikzcd}
& \text{in } \mathbf{Set} & & \text{in } \mathbf{Mon} \\
A \arrow[r, "i"] \arrow[dr, "h"'] & A^* \arrow[d, dashed, "h^*"] & & A^* \arrow[d, dashed, "h^*"] \arrow[dll, phantom, "\xleftarrow{\text{forgetful}}"] \\
& M & & M
\end{tikzcd} \tag{82}$$

[277] We omit occurences of U as the underlying set (resp. function) of a monoid (resp. homomorphism) is often denoted with the same symbol as the monoid (resp. homomorphism).

We will see in Chapter H that the assignment $A \mapsto A^*$ can be assembled into a functor $-^* : \mathbf{Set} \rightsquigarrow \mathbf{Mon}$. It goes in the opposite direction to the forgetful functor, and in fact can be seen as a weak notion of inverse to U.

Abelianization

Our next example is very similar to the previous one. We add the least amount of structure to a group G to obtain an abelian group G^{ab}.[278]

[278] This assignment assembles into a weak inverse to the intermediate forgetful functor $\mathbf{Ab} \rightsquigarrow \mathbf{Grp}$.

Definition E.3 (Classical)**.** Let G be a group, the **abelianization** of G, denoted by G^{ab}, is the quotient of G by the **commutator subgroup** $G' := \{xyx^{-1}y^{-1} \mid x, y \in G\} \subseteq G$, that is $G^{\mathrm{ab}} := G/G'$.

Let us get more insight into this definition. The abelianization is supposed to be the *biggest* abelian quotient of G. To see why, note that if A is an abelian group, any homomorphism $h : G \to A$ must satisfy $h(xyx^{-1}y^{-1}) = 1_A$ for any $x, y \in G$.[279] Hence, G' is contained in the kernel of h. By the fundamental theorem of homomorphism (ref), there is a unique factorization $h = G \xrightarrow{\pi} G/G' \xrightarrow{h'} A$, where π is the canonical quotient map. We summarize this universal property as follows.

[279] The homomorphism property implies

$$\begin{aligned} h(xyx^{-1}y^{-1}) &= h(x)h(y)h(x)^{-1}h(y)^{-1} \\ &= h(x)h(x)^{-1}h(y)h(y)^{-1} \\ &= 1_A. \end{aligned}$$

Definition E.4 (Categorical)**.** Let G be a group, the abelianization of G is an abelian group G^{ab} with a map $\pi : G \to G^{\mathrm{ab}}$ satisfying the following universal property: for any homomorphism $h : G \to A$ where A is abelian, there is a unique homomorphism $h^* : G^{\mathrm{ab}} \to A$ such that $h^* \circ \pi = h$. This is summarized in (83).

$$\begin{tikzcd} \text{in } \mathbf{Grp} & & & \text{in } \mathbf{Ab} \\ G \arrow[r, "\pi"] \arrow[dr, "h"'] & G^{\mathrm{ab}} \arrow[d, dashed, "h^*"] & & G^{\mathrm{ab}} \arrow[d, dashed, "h^*"] \\ & A & & A \end{tikzcd} \arrow[from=2-4, to=2-2, "\text{forgetful}"] \tag{83}$$

We can verify that this characterizes the abelianization of G up to isomorphism.[280]

[280] Compare with what we proved for free monoids.

OL **Exercise E.5.** Let $p : G \to H$ satisfy the universal property of $\pi : G \to G^{\mathrm{ab}}$. Show that $G^{\mathrm{ab}} \cong H$.

Vector Space Basis

This is the third and last example of the same flavor.[281]

[281] We now work with the forgetful functor $\mathbf{Vect}_k \rightsquigarrow \mathbf{Set}$.

Definition E.6 (Classical)**.** Let V be a vector space over a field k, a **basis** for V is a subset $S \subseteq V$ that is linearly independent and generates V, namely, any $v \in V$ can be expressed as a linear combination of elements in S and any $s \in S$ cannot be expressed as a linear combination of elements in $S \setminus \{s\}$.

Once again, we would like to get rid of the content of this definition talking about elements, so we focus on what this means for linear maps coming out of V. Let S be a basis of V, W be another vector space over k and $T : V \to W$ be a linear map. By linearity, T is completely determined by where it sends the elements of S. Indeed, for any $v \in V$, write v as a linear combination $\sum_{s \in S} \lambda_s s$ with $\lambda_s \in k$ (only finitely many of the coefficients are non-zero), then $T(v) = \sum_{s \in S} \lambda_s T(s)$. We conclude that any (set-theoretic) function $t : S \to W$ extends to a unique linear map $T : V \to W$.[282]

[282] This is completely analogous to how any homomorphism from the free monoid A^* is determined by where it sends the generators (elements of A).

We claim that this property completely characterizes bases of V. Indeed, let $S \subseteq V$ be such that for any $t : S \to W$, there is a unique linear map $T : V \to W$ extending t. We will show that S is generating and linearly independent.

1. Let U be the subspace generated by S.[283] We claim that the quotient space V/U is $\{0\}$ implying $U = V$, i.e., S is generating. Let $t : S \to V/U$ be the function sending everything to 0, both the quotient map $\pi : V \to V/U$ and the 0 map $0 : V \to V/U$ extend t linearly.[284] By the uniqueness in the universal property, π and 0 must coincide, hence V/U must be trivial.

[283] It contains all linear combinations of elements in S.

[284] The former extends t because every linear combination of elements in S is in U which π sends to 0.

2. Fix $v \in S$, we will show that v is not a linear combination of elements in $S \setminus \{v\}$. First, we claim that v is not zero. If it were, then any function $t : S \to k$ sending v to a non-zero element could not be extended. Next, consider the function[285]

$$t : S \to V + V = \begin{cases} (s, 0) & s \neq v \\ (0, v) & s = v \end{cases}.$$

By the universal property, there exists a linear map $T : V \to V + V$ extending t. Notice that applying T to a linear combination of elements in S, we must obtain a vector of $V + V$ whose second coordinate is 0. However, the second coordinate of $T(v)$ is v, not 0. Hence, v is not a linear combination of elements in S. Our choice of v was arbitrary, so we can conclude that S is linearly independent.

[285] Recall that the coproduct of vector spaces is their direct sum, i.e. $V + V = \{(u, w) \mid u, w \in V\}$ and operations are done coordinate-wise.

We have the following alternative definition of a vector space basis.[286]

[286] We are assuming a different point of view than we did for free monoids, but we are doing the same thing. One could start from a set S and say that V is the free vector space over S if there is the inclusion $i : S \to V$ satisfying (84).

This opposite point of view can be misleading. If we try to prove that this characterizes the basis up to isomorphism (i.e. if S and S' are bases of V, then $S \cong S'$), we will have a harder time than before. Comparing with the proofs for free monoids and abelianizations, we find we can easily prove that if V and W have S as a basis, then $V \cong W$.

Definition E.7 (Categorical)**.** Let V be a vector space, a basis of V is a set S along with an inclusion $i : S \to V$ satisfying the following universal property: for any function $t : S \to W$ where W is a vector space, there is a unique linear map $T : V \to W$ such that $T \circ i = t$. This is summarized in (84).

$$\begin{tikzcd} & \text{in } \mathbf{Set} & & & \text{in } \mathbf{Vect}_k \\ S \arrow[r, "i"] \arrow[dr, "t"'] & V \arrow[d, dashed, "T"] & & & V \arrow[d, dashed, "T"] \\ & W & & & W \end{tikzcd} \quad \text{(forgetful: right } T \text{ to left } T) \tag{84}$$

The previous three examples of universal properties are all categorifications of a free construction. Here are two others we leave you to work out on your own.

OL **Exercise E.8.** What is the free partial order over a set S?

Recall that we can see a category as a directed graph with extra structure using the forgetful functor $U : \mathbf{Cat} \rightsquigarrow \mathbf{DGph}$ that forgets about composition and identities. From any directed graph G, we can construct a category of paths of G, denoted by $\mathbf{P}G$. The objects of $\mathbf{P}G$ are those of G, and the morphisms in $\mathrm{Hom}_{\mathbf{P}G}(A, B)$ are paths from A to B in G. The composition of two paths $A \xrightarrow{f_1} \cdots \xrightarrow{f_n} B$ and $B \xrightarrow{g_1} \cdots \xrightarrow{g_m} C$ is the concatenated path $A \xrightarrow{f_1} \cdots \xrightarrow{f_n} B \xrightarrow{g_1} \cdots \xrightarrow{g_m} C$, and the identity on A is the empty path going from A to A.[287]

[287] Of course, concatenating a path with the empty path does nothing.

OL **Exercise E.9.** Show that $\mathbf{P}G$ is the free category over the directed graph G. Moreover, show that when G has a single object, $\mathbf{P}G$ is the delooping of the free monoid ${G_1}^*$.

Exponential Objects

This section and the following two are motivated by important constructions in **Set** that we want to define categorically. Going further in this direction amounts to doing topos theory, namely, studying categories which look a lot like **Set**.

Remark E.10. Let me repeat that there is a choice to make when doing such categorifications. Given a classical construction, we need to decide what is the core idea that we want to keep when we abstract away from concrete details. If this core idea allows you to recover the original construction when instantiating back in **Set**, then your abstraction is appropriate, but it might not be the only one.

OL **Exercise E.11.** Let $\mathbf{C}$ be a category and $X \in \mathbf{C}_0$ be such that for any $Y \in \mathbf{C}_0$, $Y \times X$ exists. Show that $- \times X$ is a functor $\mathbf{C} \rightsquigarrow \mathbf{C}$.

Let A and X be sets, A^X commonly denotes the set of functions $X \to A$, that is, the hom-set $\mathrm{Hom}_{\mathbf{Set}}(A, X)$. This is a somewhat exceptional situation, the hom-set between two objects in **Set** is itself an object of **Set**. There are other categories where hom-sets can actually be viewed as objects of that category.[288] Exponential objects make this parallel formal.

[288] For instance, the set of linear maps $V \to W$ is a vector space where addition and scalar multiplication is done pointwise.

In hope to generalize the construction of A^X to other categories, let us study morphisms into A^X.[289] Given a set B and a morphism $f : B \to A^X$, there is a natural operation called **uncurrying** that takes f to $\lambda^{-1}f : B \times X \to A$ which basically evaluates both f and its output at the same time. Namely, $\lambda^{-1}f(b, x) = f(b)(x)$.

[289] A priori, there is no reason to prefer morphisms into A^X over morphisms out of A^X, but the intuition is cleaner with the former.

As a particular case, we consider the identity function $A^X \to A^X$. Uncurrying yields the **evaluation** function $\mathsf{ev} : A^X \times X \to A$ that evaluates the function in the first coordinate at the second coordinate: $\mathsf{ev}(f, x) = f(x)$.

Now, as the name suggests, uncurrying has an inverse operation called **currying**[290] which takes $g : B \times X \to A$ to $\lambda g : B \to A^X$ defined by $\lambda g(b) = x \mapsto g(b, x)$. Morally, λg delays the evaluation of g on the second input to later.[291] Moreover, notice that the currying of g satisfies $\mathsf{ev}(\lambda g(b), x) = g(b, x) \in A$ for any $b \in B$ and $x \in X$. Intuitively, $\lambda g(b)$ reads the first argument b and waits for the second argument, then $\mathsf{ev}(\lambda g(b), x)$ inputs x, so it is the same thing as doing $g(b, x)$. This along with the fact that currying and uncurrying are bijective operations[292] leads to a universal property that ev satisfies. It is summarized in (85).

[290] Named in honor of Haskell Curry.

[291] For computer scientists, this is also related to the concept of *continuations*.

[292] Check that $\lambda\lambda^{-1}g = g$ and $\lambda^{-1}\lambda g = g$.

$$\begin{tikzcd} & \text{in } \mathbf{Set} & & \text{in } \mathbf{Set} \\ A & A^X \times X \arrow[l, "\mathsf{ev}"'] & & A^X \\ & B \times X \arrow[ul, "g"] \arrow[u, dashed, "\lambda g \times \mathrm{id}_X"'] & & B \arrow[u, dashed, "\lambda g"'] \arrow[ll, phantom, "\xleftarrow{-\times X}"] \end{tikzcd} \tag{85}$$

This is entirely categorical, so we can define exponential objects as follows.

Definition E.12 (Exponential)**.** Let $\mathbf{C}$ be a category and $X \in \mathbf{C}_0$ be such that $- \times X$ is a functor.[293] For $A \in \mathbf{C}_0$, the **exponential** A^X (if it exists) is an object A^X along with a morphism $\mathsf{ev} : A^X \times X \to A$ such that for all $g : B \times X \to A$, there is a unique $\lambda g : B \to A^X$ making (85) commute.

[293] i.e.: all binary products with $X \in \mathbf{C}_0$ exist.

Informally, one can think of A^X as an object which behaves like $\mathrm{Hom}_{\mathbf{C}}(A, X)$. The terminology **internal hom** is often used (sometimes in more general contexts).

OL **Exercise E.13.** Let k be a field, and V and W be vector spaces over k. Show that the vector space $\mathrm{Hom}_{\mathbf{Vect}_k}(V, W)$ equipped with pointwise addition and scalar multiplication of linear maps is the exponential W^V.

OL **Exercise E.14.** Show that if $e : Y \times X \to A$ satisfies the same universal property as ev, then $Y \cong A^X$.[294]

[294] We will stop proving that universal properties determine objects up to isomorphisms, the abstract result (stating that works for all universal properties) is Corollary **??**.

Definition E.15 (Cartesian closed)**.** When a category $\mathbf{C}$ has a terminal object and all exponentials A^X for all $A, X \in \mathbf{C}_0$ (in particular, it has all binary products[295]), we say it is **cartesian closed**.

[295] It also follows that $\mathbf{C}$ has all finite products.

The category of sets is cartesian closed. Here is an exercise calling back to when we showed many familiar properties of Cartesian products generalized to binary products.

OL **Exercise E.16.** Let $\mathbf{C}$ be a category with a terminal object $\mathbf{1}$, and let $X \in \mathbf{C}_0$. Show that X is the exponential $X^{\mathbf{1}}$ and $\mathbf{1}$ is the exponential $\mathbf{1}^X$,[296] i.e. find the evaluation morphisms and prove they satisfy the right universal property.

[296] Other properties about exponentials in **Set** can be generalized (e.g. $(X^Y)^Z \cong X^{Y\times Z}$), but we will wait until we see the Yoneda lemma to give more elegant proofs.

Subobject Classifier

OL **Exercise E.17.** Let $\mathbf{C}$ be a well-powered category with all pullbacks. We define $\mathrm{Sub}_{\mathbf{C}}$ on morphisms: it sends $f : X \to Y$ to $f^*(-) : \mathrm{Sub}_{\mathbf{C}}(Y) \to \mathrm{Sub}_{\mathbf{C}}(X)$ sending $m : I \rightarrowtail Y$ to $f^*(m)$, the pullback of m along f as depicted in (86). Show that this is well-defined (recall that a subobject of Y is an equivalence class of monomorphisms) and makes $\mathrm{Sub}_{\mathbf{C}}$ into a functor $\mathbf{C}^{\mathrm{op}} \rightsquigarrow \mathbf{Set}$.

$$\begin{tikzcd} J \arrow[r] \arrow[d, tail, "f^*(m)"'] \arrow[dr, phantom, "\lrcorner", very near start] & I \arrow[d, tail, "m"] \\ X \arrow[r, "f"'] & Y \end{tikzcd} \tag{86}$$

In **Set**, recall that subobjects are subsets. Hence, letting $\Omega = \{\perp, \top\}$ there is a correspondence between $\mathrm{Sub}_{\mathbf{Set}}(X)$ and $\mathrm{Hom}_{\mathbf{Set}}(X, \Omega)$, it sends $I \subseteq X$ to the characteristic function $\chi_I : X \to \Omega$,[297] and in the other direction $f : X \to \Omega$ is sent to $f^{-1}(\top) \subseteq X$. In particular, we have that $\chi_I^{-1}(\top) = I$, which we can write categorically as the following pullback.[298]

[297] The **characteristic function** χ_I is defined by

$$\chi_I(x) = \begin{cases} \top & x \in I \\ \perp & x \notin I \end{cases}.$$

[298] Recall our discussion on preimages in Example D.45.

$$\begin{tikzcd} I \arrow[r] \arrow[d, hook] \arrow[dr, phantom, "\lrcorner", very near start] & \mathbf{1} \arrow[d, hook, "\top"] \\ X \arrow[r, "\chi_I"'] & \Omega \end{tikzcd} \tag{87}$$

Crucially, this pullback uniquely determines χ_I.[299] The role played by the two element set $\{\perp, \top\}$ can now be generalized to other categories.

[299] If $f : X \to \Omega$ also makes (87) a pullback square, then $f^{-1}(\top) = I$, so f and χ_I must coincide. The preimage of f on $\top$ determines all of f because there is only one other value in the codomain of f.

Definition E.18 (Subobject classifier)**.** Let $\mathbf{C}$ be a category with a terminal object $\mathbf{1}$. The **subobject classifier** (if it exists) is a morphism $\top : \mathbf{1} \to \Omega \in \mathbf{C}_1$ such that for any monomorphism $I \rightarrowtail X$ there is a unique morphism $\chi_m : X \to \Omega$ such that (87) is a pullback square. We call χ_I the **classifying morphism** of $I \rightarrowtail X$.

Example E.19 ($\mathbf{Set}_*$)**.** We find the subobject classifier in $\mathbf{Set}_*$.

Let (X, x) be a pointed set, we first show that a subobject of (X, x) is a subset of X that contains x. An argument like the one in Example C.18 shows that monomorphisms in $\mathbf{Set}_*$ are precisely the injective functions that preserve the point.[300] Hence, for a subset $I \subseteq X$ with $x \in I$, the inclusion $i : (I, x) \hookrightarrow (X, x)$ is a monomorphism. Moreover, we can show (as we did in Example C.55) that two monomorphisms

[300] We can also give a more abstract proof. The forgetful functor $\mathbf{Set}_* \rightsquigarrow \mathbf{Set}$ is faithful so it reflects monomorphisms by Exercise C.52. Also, we saw in Exercise D.72 that it preserves pullbacks, hence it preserves monomorphisms by Exercise D.70.

$(I,i) \rightarrowtail (X,x)$ and $(J,j) \rightarrowtail (X,x)$ are in the same equivalence class of $\mathrm{Sub}_{\mathbf{Set}_*}(X,x)$ if and only if their images coincide (and their image must contain x). We conclude that $\mathrm{Sub}_{\mathbf{Set}_*}(X,x)$ is in correspondence with $\{S \subseteq X \mid x \in S\}$.

The terminal object $\mathbf{1}$ in $\mathbf{Set}_*$ is the singleton $\{*\}$ with distinguished point $*$. Keeping the same notation $\Omega = \{\bot, \top\}$, we claim the subobject classifier is the unique morphism $\top : \mathbf{1} \to (\Omega, \top)$,[301] it sends $*$ to $\top$. For any subset $I \subseteq X$ that contains $x \in X$, we define the classifying morphism $\chi_I : (X,x) \to (\Omega, \top)$ as before (see Footnote 297), noting that it is a morphism in $\mathbf{Set}_*$ because x belongs to I so is mapped to $\top$. It clearly makes the square in (88) commute.[302]

[301] The terminal object $\mathbf{1}$ is also initial in $\mathbf{Set}_*$, see Exercise C.39.

[302] Both paths send everything in I to $\top$.

$$\begin{tikzcd}
(A,a) \arrow[dr, dashed, "f"] \arrow[drr, bend left] \arrow[ddr, bend right, "f"'] & & \\
& (I,x) \arrow[r] \arrow[d, hook] & (\{*\},*) \arrow[d, hook, "\top"] \\
& (X,x) \arrow[r, "\chi_I"'] & (\Omega,\top)
\end{tikzcd} \tag{88}$$

Now, for any morphism $f : (A,a) \to (X,x)$ making (88) commute, we find the image of f must be contained in I.[303] Therefore, we can factor f through the inclusion of I in X (necessarily uniquely). We conclude that the square in (88) is a pullback.

[303] Otherwise some $a \in A$ is mapped to $\bot$ in the bottom path but not the top path.

It remains to show χ_I is the only possible morphism making that possible. If another morphism χ' does, we apply the forgetful functor which preserves pullbacks (Exercise D.72) to get a pullback in $\mathbf{Set}$. Because $\top : \mathbf{1} \to \Omega$ is the subobject classifier in $\mathbf{Set}$, χ' must be the classifying morphism which is the characteristic map χ_I.

Before we can draw a diagram (akin to (82), (83), etc.) summarizing the universal property of the subobject classifier, we need to make sure that the classifying morphisms of two monomorphisms in the same equivalence class in $\mathrm{Sub}_{\mathbf{C}}(X)$ are equal. Let $I' \rightarrowtail X$ and $I \rightarrowtail X$ represent the same subobject, namely, there is an isomorphism $I' \cong I$ making the left square in (89) commute. The right square is a pullback by hypothesis and the left square is a pullback by Exercise D.81. Therefore, the rectangle is a pullback by the pasting lemma, and we see that $\chi_{I'} = \chi_I \circ \mathrm{id}_X$ by uniqueness of the classifying morphism.

$$\begin{tikzcd}
I' \arrow[r, leftrightarrow, "\sim"] \arrow[d, tail] \arrow[dr, phantom, "\lrcorner", very near start] & I \arrow[r] \arrow[d, tail] \arrow[dr, phantom, "\lrcorner", very near start] & \mathbf{1} \arrow[d, tail, "\top"] \\
X \arrow[r, "\mathrm{id}_X"'] \arrow[rr, bend right, "\chi_{I'}"'] & X \arrow[r, "\chi_I"'] & \Omega
\end{tikzcd} \tag{89}$$

Now, in a well-powered category $\mathbf{C}$ that has a terminal object and all pullbacks,[304] the subobject classifier $\top : \mathbf{1} \to \Omega$ is such that for any subobject m of X, there is a unique morphism $\chi_m : X \to \Omega$ satisfying ${\chi_m}^*(\top) = m$. This is summarized in (90) where we identify $\top$ with the function $\mathbf{1} \to \mathrm{Sub}_{\mathbf{C}}(\Omega)$ picking out this equivalence class of $\top : \mathbf{1} \to \Omega$ in $\mathrm{Sub}_{\mathbf{C}}(\Omega)$ (recall that any morphism out of $\mathbf{1}$, in particular $\top : \mathbf{1} \rightarrowtail \Omega$, is monic by Exercise C.43), and similarly for m.

[304] The definition of subobject classifier does not need the well-poweredness and the existence of all pullbacks, but they are necessary to have a universal property because it uses the functor $\mathrm{Sub}_{\mathbf{C}}$. In any case, subobject classifiers are usually used when these conditions are satisfied.

$$\begin{tikzcd}
 & \text{in } \mathbf{Set} & & \text{in } \mathbf{C} \\
\mathbf{1} \arrow[r, "\top"] \arrow[dr, "m"'] & \mathrm{Sub}_{\mathbf{C}}(\Omega) \arrow[d, dashed, "{\chi_m}^*(-)"] & & \Omega \arrow[ll, mapsto, "\mathrm{Sub}_{\mathbf{C}}"'] \\
 & \mathrm{Sub}_{\mathbf{C}}(X) & & X \arrow[u, dashed, "\chi_m"']
\end{tikzcd} \tag{90}$$

Notice that the dashed arrow gets reversed because $\mathrm{Sub}_{\mathbf{C}}$ is contravariant. We could also write "in $\mathbf{C}^{\mathrm{op}}$" and not reverse the arrow.

Power Objects

This is the third and last example that can motivate the study of topos theory.

Let X be a set, $\mathcal{P}X$ commonly denotes the set of all subsets of X. In particular, **Set** is well-powered and $\mathrm{Sub}_{\mathbf{Set}}(X)$ is a set, i.e., an object of **Set**. Again, this is an exceptional situation[305] that we would like to make abstract.

Let us study morphisms into $\mathcal{P}X$. A function $f : Y \to \mathcal{P}X$ assigns to each $y \in Y$ a (possibly empty) set $f(y)$ of values in X. We can also present the data of f as a subset Γ_f of $X \times Y$ containing the pair (x, y) whenever $x \in f(y)$. This yields a bijection between functions $f : Y \to \mathcal{P}X$ and subsets $\Gamma_f \subseteq X \times Y$[306]: given a subset $\Gamma \subseteq X \times Y$, we define $f_\Gamma : Y \to \mathcal{P}X$ by $f(y) = \{x \in X \mid (x, y) \in \Gamma\}$. The trick to rephrase this categorically is to note that Γ_f is the preimage of the "element of" subset $\in_X \subseteq X \times \mathcal{P}X$ under the function $\mathrm{id}_X \times f : X \times Y \to X \times \mathcal{P}X$.[307] Therefore, we have the following pullback (again, see Example D.45).

$$\begin{tikzcd} \Gamma_f \arrow[r] \arrow[d, hook] \arrow[dr, phantom, "\lrcorner", very near start] & \in_X \arrow[d, hook] \\ X \times Y \arrow[r, "\mathrm{id}_X \times f"'] & X \times \mathcal{P}X \end{tikzcd} \tag{91}$$

[305] This is even more exceptional than being cartesian closed. I do not have any simple examples, but we will see a couple of harder examples.

[306] This generalizes the correspondence between elements of $\mathcal{P}X$ and $\mathrm{Sub}_{\mathbf{Set}}(X)$ because

$$\mathcal{P}X \cong \mathrm{Hom}(\mathbf{1}, \mathcal{P}X) \cong \mathrm{Sub}_{\mathbf{Set}}(X \times \mathbf{1}) \cong \mathrm{Sub}_{\mathbf{Set}}(X).$$

[307] We have that $(\mathrm{id}_X \times f)(x, y) = (x, f(y))$ is in $\in_X$ if and only if $x \in f(y)$ if and only if $(x, y) \in \Gamma_f$. Thus, $\Gamma_f = (\mathrm{id}_X \times f)^{-1}(\in_X)$.

We are ready to give the abstract definition.

Definition E.20 (Power object)**.** Let $\mathbf{C}$ be a category and $X \in \mathbf{C}_0$ be such that $X \times -$ is a functor. The **power object** of X (if it exists) is an object $\mathfrak{P}X \in \mathbf{C}_0$ along with a monomorphism $\in_X \rightarrowtail X \times \mathfrak{P}X$ such that for any monomorphism $\gamma : \Gamma \rightarrowtail X \times Y$, there is a unique morphism $f_\gamma : Y \to \mathfrak{P}X$ making (92) a pullback square.

$$\begin{tikzcd} \Gamma \arrow[r] \arrow[d, tail, "\gamma"'] \arrow[dr, phantom, "\lrcorner", very near start] & \in_X \arrow[d, tail] \\ X \times Y \arrow[r, dashed, "\mathrm{id}_X \times f_\gamma"'] & X \times \mathfrak{P}X \end{tikzcd} \tag{92}$$

Note that we obtain f_γ from γ instead of Γ_f from f (like we did in **Set**). In the end, it does not matter because the key property is that there is a correspondence between them. However, in the definition above, the fact that pullbacks are unique up to isomorphisms implies γ is uniquely determined by f_γ up to isomorphism,[308] hence we only need to require f_γ is uniquely determined by γ.

[308] More precisely, the subobject represented by γ is uniquely determined by γ.

Example E.21 ($\mathbf{Set}_*$)**.** Recall that a subobject of (X, x) in $\mathbf{Set}_*$ is a subset of X that contains x. This suggests the power object of X may be the set of subsets of X containing x. However we still need to figure out what would be the distinguished point in that set. It turns out there is no point that works out. In fact, we can show that, in general, (X, x) does not have a power object.

We saw above that the power object $\mathfrak{P}(X, x)$ must satisfy

$$\mathrm{Hom}(\mathbf{1}, \mathfrak{P}(X, x)) \cong \mathrm{Sub}_{\mathbf{Set}_*}((X, x) \times \mathbf{1}).$$

Since $\mathbf{1}$ is initial in $\mathbf{Set}_*$, the L.H.S. is a singleton set. We recall that taking a product with the terminal object does nothing (Exercise D.15), so the R.H.S. is the set of all subsets of X containing x. Hence, this isomorphism cannot be unless $(X, x) = \mathbf{1}$.[309]

[309] In that case, you can check $\mathbf{1}$ has a (uninteresting) power object.

Again, we want to draw a diagram that summarizes this universal property. Just like for subobject classifiers, we have to check f_γ is the same as $f_{\gamma'}$ when γ and γ' are representatives for the same subobject.

OL **Exercise E.22.** Let $\in_X \rightarrowtail X \times \mathfrak{P}X$ be the power object of $X \in \mathbf{C}_0$. Show that if γ and γ' are two monomorphisms equal in $\mathrm{Sub}_{\mathbf{C}}(X \times Y)$, then $f_\gamma = f_{\gamma'}$.

We can conclude that if $\mathbf{C}$ is well-powered and has a terminal object, the power object of $X \in \mathbf{C}_0$ is a monomorphism $\in_X \rightarrowtail X \times \mathfrak{P}X$ such that for any subobject γ of $X \times Y$, there is a unique morphism $f_\gamma : Y \to \mathfrak{P}X$ satisfying $(\mathrm{id}_X \times f_\gamma)^*(\in_X) = \gamma$. This is summarized in (93).

$$\begin{tikzcd}
& \text{in } \mathbf{Set} & & \text{in } \mathbf{C} \\
\mathbf{1} \arrow[r, "\in_X"] \arrow[ddr, "\iota"'] & \mathrm{Sub}_{\mathbf{C}}(X \times \mathfrak{P}X) \arrow[dd, dashed, "{f_\gamma}^*(\mathrm{id}_X \times -)"] & & \mathfrak{P}X \\
& {} & {} \arrow[l, "\mathrm{Sub}_{\mathbf{C}}(X \times -)"'] & {} \\
& \mathrm{Sub}_{\mathbf{C}}(X \times Y) & & Y \arrow[uu, dashed, "f_\gamma"']
\end{tikzcd} \tag{93}$$

In the category **DGph**, any graph has power object.[310] Before proving this, we need to explain what are subobjects and how to take products and pullbacks in **DGph**.

Adapting the solution to Exercise C.58, we find that the subobjects of $G \in \mathbf{DGph}_0$ are graphs H with $H_0 \subseteq G_0$ and $H_1 \subseteq G_1$ such that the source and target maps of H are restrictions of those of G. Similarly to subcategories, we can obtain H from G by deleting arrows and objects, and making sure the sources and targets of remaining arrows also remain.

Again taking inspiration from **Cat**, Definition B.40 (see also Exercise D.11) tells us how to define binary products of graphs if we forget about the composition and identities.[311]

We have not yet defined pullbacks in **Cat**, but we will do it only for **DGph** here because it is easier.

OL **Exercise E.23.** Given two morphisms $f : A \to C$ and $g : B \to C$ in **DGph**, find the pullback $A \times_C B$. Show that the functors $(-)_0 : \mathbf{DGph} \rightsquigarrow \mathbf{Set}$ and $(-)_1 : \mathbf{DGph} \rightsquigarrow \mathbf{Set}$[312] preserve pullbacks.

The second part of this exercise is a hint for the first part, and it is what we will use shortly. Unrolling, it means the objects and arrows of $A \times_C B$ are defined as follows:

$$\begin{aligned}
(A \times_C B)_0 &= \left\{(x, x') \in A_0 \times B_0 \mid f_0(x) = g_0(x')\right\} \\
(A \times_C B)_1 &= \left\{(e, e') \in A_1 \times B_1 \mid f_1(e) = g_1(e')\right\}.
\end{aligned}$$

Example E.24 (**DGph**). [313] Fix a graph X, we will find $\mathfrak{P}X$.

The universal property of $\mathfrak{P}X$ implies that there is a correspondence between morphisms $\mathbf{1} \to \mathfrak{P}X$ and subobjects of X (the terminal object in **DGph** is the graph with one object and one arrow). For **Cat**, we saw that a functor $\mathbf{1} \rightsquigarrow \mathbf{C}$ is just a choice of object in $\mathbf{C}_0$, but this is not the case in **DGph**. A morphism of graphs does not need to preserve identities, thus a morphism $\mathbf{1} \rightsquigarrow X$ is a choice of object plus a choice of loop on it. This means in $\mathfrak{P}X$, we should have one loop for each subgraph of X. Unfortunately, this does not tell us that much at this point.[314]

[310] Recall that **DGph** contains only small directed graphs, those with a set of objects and a set of arrows. The morphisms in **DGph** are like functors, but without the requirements about preserving composition and identities (they are not defined in a directed graph).

[311] In other words, the forgetful functor $\mathbf{Cat} \rightsquigarrow \mathbf{DGph}$ preserves binary products.

[312] These are defined like for **Cat** in Exercise B.39.

[313] I am writing this as if we are figuring it out together, but we will use a couple of clever tricks that come from higher-level arguments that we cannot give yet (seeing **DGph** as a functor category).

[314] We will come back to this later.

To give a complete (and enlightening) description of $\mathfrak{P}X$, we need to know what are its objects, its arrows and the source and target of its arrows. We will make use of a more general consequence of the universal property of $\mathfrak{P}X$: for any graph Y, $\mathrm{Hom}(Y, \mathfrak{P}X) \cong \mathrm{Sub}(X \times Y)$. We can find two graphs O and A such that $\mathrm{Hom}(O, \mathfrak{P}X)$ is in correspondence with the objects of $\mathfrak{P}X$ and $\mathrm{Hom}(A, X)$ with its arrows.

The graph O only contains one object o and no arrow. A morphism $O \to \mathfrak{P}X$ is then just a choice of an object that is the image of o. The product $X \times O$ has the same objects as X but no arrows.[315] Therefore, a subgraph of $X \times O$ is a subset of X_0, and we conclude that we can define $(\mathfrak{P}X)_0 = \mathcal{P}(X_0)$.

[315] By Definition B.40, we have

$$(X \times O)_0 = X_0 \times O_0 = X_0 \times \{o\} \cong X, \text{ and}$$
$$(X \times O)_1 = X_1 \times O_1 = X_1 \times \emptyset \cong \emptyset.$$

The graph A contains two objects and one arrow a between them. It looks like the graph of $\mathbf{2}$, but without the identity morphisms. A morphism $A \to \mathfrak{P}X$ is a choice of an arrow that is the image of a, and a redundant (determined by the first choice) choice for the image of the source and target of a. The product $X \times A$ can be viewed as two copies of the objects of X (one for each object of A), and for each arrow $f : x \to x'$ in X, there is an arrow from the first copy of x to the second copy of x'.[316] Here is a drawing of a small example.

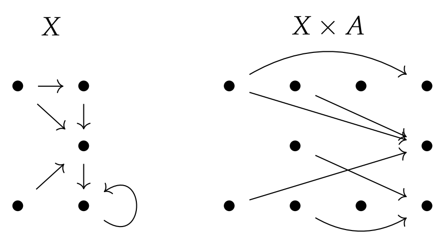
[316] By Definition B.40, we have

$$(X \times A)_0 = X_0 \times A_0 = X_0 \times \{1, 2\} \cong X + X,$$

and all morphisms are of the form $(g, a) : (x, 1) \to (x', 2)$ where $g : x \to x'$. Thus,

$$\mathrm{Hom}_{X \times A}((x, 1), (x', 2)) = \mathrm{Hom}_X(x, x'),$$

and all other hom-sets are empty.



A subgraph $H \rightarrowtail X \times A$ can be seen as two subsets H^1 and H^2 of X_0[317] along with a set of arrows $H^a \subseteq X_1$ whose sources are in H^1 and targets are in H^2. We define $(\mathfrak{P}X)_1$ to be the set of all such triples (H^1, H^2, H^a) to ensure we have $\mathrm{Hom}(A, X) \cong (\mathfrak{P}X)_1 \cong \mathrm{Sub}(X \times A)$.

[317] H^1 contains the objects of H belonging to the first copy of X in $X \times A$ and H^2 contains the objects of H in the second copy.

It seems more than likely that H^1 and H^2, being objects of $\mathfrak{P}X$, are the source and target of the arrow (H^1, H^2, H^a). As a sanity check, let us verify that with this definition of source and target in $\mathfrak{P}X$, the loops are in correspondence with subgraphs of X; that is the first thing we discovered about $\mathfrak{P}X$. If $H^1 = H^2$, then the triple defines the subgraph of X containing all the objects in H^1 (or H^2) and all the arrows in H^a. Conversely, given a subgraph of $H \rightarrowtail X$, we let both H^1 and H^2 be the set of objects of H and H^a be the set of arrows of H.

We seem to be on the right track, and we need one last thing in the definition of power object,[318] the subgraph $\in_X$ of $X \times \mathfrak{P}X$. Since we are almost done, we will totally trust our intuition of what $\in_X$ should be without looking for more justifications. The objects of $\in_X$ are pairs (x, H) where $x \in X_0$ and $H \subseteq X_0$, it makes sense to require that $x \in H$. The arrows of $\in_X$ are pairs $(f, (H^1, H^2, H^a))$ where $f : x \to x'$, $x \in H^1$, $x' \in H^2$, it makes sense to require that $f \in H^a$.[319] We are ready to prove $\in_X \rightarrowtail \mathfrak{P}X$ satisfies the universal property of the power object of X.

[318] Before the proof of the universal property.

[319] Recall that every arrow in H^a has its source in H^1 and its target in H^2 just like f.

Let Γ be a subgraph of $X \times Y$ with inclusion $\gamma : \Gamma \to X \times Y$.[320] We need to define a morphism $f_\gamma : Y \to \mathfrak{P}X$ making (92) a pullback square, and we also need to prove

[320] We assume without loss of generality that γ is an inclusion (not an arbitrary monomorphism) to avoid having different names for stuff in Γ and stuff in $X \times Y$.

it is unique. Let us use Exercise E.23 to compute the pullback for some yet undefined f_γ, and we will then figure out what constraints we obtain on f_γ when requiring that pullback to be Γ. Hopefully, these will uniquely define f_γ.

Call this pullback G. The objects of G are tuples[321]

[321] Recall that $\in_X$ is a subgraph of $X \times \mathfrak{P}X$.

$$((x,y),(x',S)) \subseteq (X \times Y)_0 \times (\in_X)_0$$

that satisfy, by commutativity of (92), $x = x'$ and $f_\gamma(y) = S \subseteq X_0$, and by definition of $\in_X$, $x' \in S$. Since the second pair is determined by the first, we can be equivalently write

$$G_0 = \{(x,y) \in X_0 \times Y_0 \mid x \in f_\gamma(y)\}.$$

Thus, to ensure G has the same objects as Γ, it is enough that f_γ satisfies $x \in f_\gamma(y) \Leftrightarrow (x,y) \in \Gamma$ which means $f_\gamma(y) = \{x \in X_0 \mid (x,y) \in \Gamma_0\}$.

The arrows of G are tuples

$$((g,h),(g',(H^1,H^2,H^a))) \subseteq (X \times Y)_1 \times (\in_X)_1$$

that satisfy, by commutativity of (92), $g = g'$ and $f_\gamma(h) = (H^1,H^2,H^a)$, and by definition of $\in_X$, $s(g) \in H^1$, $t(g) \in H^2$ and $g \in H^a$. Like above, we make things more concise:

$$G_1 = \{(g,h) \in X_1 \times Y_1 \mid s(g) \in f_\gamma(h)^1, t(g) \in f_\gamma(h)^2, g \in f_\gamma(h)^a\}.$$

To ensure G has the same arrows as Γ, we define $f_\gamma(h)$ be the arrow defined by the triple

$$(\{s(g) \mid (g,h) \in \Gamma_1\}, \{t(g) \mid (g,h) \in \Gamma_1\}, \{g \mid (g,h) \in \Gamma_1\}).$$

We leave you two final things to check. First, we only exhibited bijections between the objects and arrows of G and Γ, but in order for (92) to be a pullback, we have to make sure these bijections assemble into an isomorphism making (94) commute. Second, for any other f_γ, the pullback G is another subobject of $X \times Y$ (i.e. there is no isomorphism as in (94)).

$$\begin{tikzcd} \Gamma \arrow[r, phantom, "\cong"] \arrow[dr, hook, "\gamma"'] & G \arrow[r] \arrow[d, hook] \arrow[dr, phantom, "\lrcorner", very near start] & \in_X \arrow[d, hook] \\ & X \times Y \arrow[r, "\mathrm{id}_X \times f_\gamma"'] & X \times \mathfrak{P}X \end{tikzcd} \quad (94)$$

Unlike for exponentials, there is no well-known terminology for a category with all power objects. This is because power objects are usually studied in categories with all finite limits, and when such a category has all power objects, it is called a topos.

Definition E.25 (Topos)**.** A finitely complete category where every object has a power object is called an **(elementary) topos**.

Digression on Toposes

The goal of this section is to give an equivalent definition of a topos using exponentials and subobject classifiers. The proofs will be done in exercises, so it is your chance to do some more diagram chasing.

In **Set**,[322] the power object of the terminal set **1** is the set with two elements,

[322] Recall it is supposed to be the archetypal topos.

$\emptyset \subseteq \mathbf{1}$ and $\mathbf{1} \subseteq \mathbf{1}$. Now, $\gamma = \mathrm{id}_{\mathbf{1}} : \mathbf{1} \to \mathbf{1}$ is a monomorphism, so we can see it as a subobject in $\mathrm{Sub}_{\mathbf{Set}}(\mathbf{1})$ or $\mathrm{Sub}_{\mathbf{Set}}(\mathbf{1} \times \mathbf{1})$ via the isomorphism $\mathbf{1} \cong \mathbf{1} \times \mathbf{1}$. Using the universal property of $\mathcal{P}\mathbf{1}$, we find that $f_\gamma : \mathbf{1} \to \mathcal{P}\mathbf{1}$ sends the single element in $\mathbf{1}$ to $\mathbf{1} \in \mathcal{P}\mathbf{1}$.[323]

[323] More rigorously, it is the universal property of $\in_{\mathbf{1}} \subseteq \mathbf{1} \times \mathcal{P}\mathbf{1}$ which contains the only element in the R.H.S. You can check that this the only f_γ making (95) a pullback.

$$\begin{tikzcd} \mathbf{1} \arrow[r] \arrow[d, hook, "\mathrm{id}_{\mathbf{1}}"'] \arrow[dr, phantom, very near start, "\lrcorner"] & \in_{\mathbf{1}} \arrow[dd, hook] \\ \mathbf{1} \arrow[d, phantom, "\cong" rotate=90] & \\ \mathbf{1} \times \mathbf{1} \arrow[r, "\mathrm{id}_{\mathbf{1}} \times f_\gamma"'] & \mathbf{1} \times \mathcal{P}\mathbf{1} \end{tikzcd} \tag{95}$$

Notice that f_γ is (up to isomorphism) the same function as the subobject classifier $\top : \mathbf{1} \to \{\bot, \top\}$. In fact, in every topos, you can find the subobject classifier this way.

Example E.26 (**DGph**).

Natural Numbers Object

We end this section with a simpler example still related to toposes to some extent. Without going into the details, topos theory is a framework to study mathematical logic and set theory with a categorical point of view.[324] One of the fundamental building blocks of logic and set theory is the set of natural numbers $\mathbb{N} = \{0, 1, 2, \dots\}$ and the principle of induction tied to it. Let us restate the latter categorically.

The set $\mathbb{N}$ comes with a distinguished element 0 that starts off inductive arguments. It corresponds to the function $0 : \mathbf{1} \to \mathbb{N}$ that picks out 0. For the inductive step, we rely on the function $\mathsf{succ} : \mathbb{N} \to \mathbb{N}$ that takes n to $n+1$.[325] The universal property of $\mathbb{N}$ is that for any pair of functions $z : \mathbf{1} \to X$, $s : X \to X$, there exists a unique $f : \mathbb{N} \to X$ making (96) commute.

[324] Ok, just a bit of informal details...
Grothendieck first defined a more constrained version of topos to help his research in algebraic geometry.
Lawvere and Tierney enlarged the notion of topos to the definition we gave, initiating a deep dive into the strong link between logic and toposes.
Later, Caramello launched a research programme on "toposes as bridges" that uses toposes to formally translate results and concepts between mathematical theories.

[325] The name succ refers to $n+1$ being the *successor* of n in $\mathbb{N}$.

$$\begin{tikzcd} \mathbf{1} \arrow[r, "0"] \arrow[dr, "z"'] & \mathbb{N} \arrow[r, "\mathsf{succ}"] \arrow[d, dashed, "f"] & \mathbb{N} \arrow[d, dashed, "f"] \\ & X \arrow[r, "s"'] & X \end{tikzcd} \tag{96}$$

The function f is defined inductively. We let $f(0)$ be the element of X in the image of z so that the triangle commutes, then we let $f(n+1) = s(f(n))$ to ensure the square commutes. This means for any $n \in \mathbb{N}$, $f(n) = s^n(z)$ where s^n denotes the composition $s \circ \overset{n}{\cdots} \circ s$ with $s^0 = \mathrm{id}_X$. We abstract away from **Set**.

Definition E.27 (NNO)**.** In a category $\mathbf{C}$ with a terminal object $\mathbf{1}$, the **natural numbers object** or NNO (if it exists) is an object $\mathfrak{N} \in \mathbf{C}_0$ along with two morphisms $0 : \mathbf{1} \to \mathfrak{N}$ and $\mathsf{succ} : \mathfrak{N} \to \mathfrak{N}$ satisfying the following universal property: for any pair of morphisms $z : \mathbf{1} \to X$ and $s : X \to X$, there exists a unique morphism $! : \mathfrak{N} \to X$ making (97) commute.

$$\begin{tikzcd} \mathbf{1} \arrow[r, "0"] \arrow[dr, "z"'] & \mathfrak{N} \arrow[r, "\mathsf{succ}"] \arrow[d, dashed, "!"] & \mathfrak{N} \arrow[d, dashed, "!"] \\ & X \arrow[r, "s"'] & X \end{tikzcd} \tag{97}$$

OL **Exercise E.28.** Show that the NNO in **Poset** is $(\mathbb{N}, \leq)$ with the same zero and successor functions (now seen as morphisms in **Poset**).

It is not evident how we could summarize the universal property of an NNO using a diagram exactly like the others. Still, the definition really feels like a universal property, so we should not forget this when generalizing what we have seen in all examples above.

E.2 Generalization

Diagrams (82), (83), (84), (85), (90) and (93) look so similar that you can try to infer the following definition unifying all these concepts under one roof.[326]

[326] Although, (85) looks like all arrows have been reversed, so, you guessed it, it will be an instance of the dual notion.

Definition E.29 (Universal morphism)**.** If $F : \mathbf{D} \rightsquigarrow \mathbf{C}$ is a functor and $X \in \mathbf{C}_0$. A **universal morphism** from X to F is a morphism $a : X \to F(A)$ such that for any other morphism $b : X \to F(B)$, there is a unique morphism $f : A \to B$ in $\mathbf{D}$ such that $F(f) \circ a = b$, which is summarized in (98).

$$\begin{tikzcd}
& \text{in } \mathbf{C} & & \text{in } \mathbf{D} \\
X \arrow[r, "a"] \arrow[dr, "b"'] & FA \arrow[d, dashed, "Ff"] & & A \arrow[d, dashed, "f"] \arrow[ll, phantom, "\xleftarrow{\quad F \quad}" description, shift right=3ex] \\
& FB & & B
\end{tikzcd} \tag{98}$$

The dual notion is a universal morphism from F to X.[327] It is a morphism $a : F(A) \to X$ such that for any other morphism $F(B) \to X$, there is a unique morphism $f : B \to A$ in $\mathbf{D}$ satisfying $a \circ F(f) = b$. This is summarized below in (99).

[327] The duality is clear from how (99) is just (98) with all morphisms reversed. More abstractly, we can say that a universal morphism from F to X is a universal morphism from $X \in \mathbf{C}^{\mathrm{op}}$ to $F^{\mathrm{op}} : \mathbf{D}^{\mathrm{op}} \rightsquigarrow \mathbf{C}^{\mathrm{op}}$.

$$\begin{tikzcd}
& \text{in } \mathbf{C} & & \text{in } \mathbf{D} \\
X & FA \arrow[l, "a"'] & & A \\
& FB \arrow[ul, "b"] \arrow[u, dashed, "Ff"'] & & B \arrow[u, dashed, "f"'] \arrow[ll, phantom, "\xleftarrow{\quad F \quad}" description, shift left=3ex]
\end{tikzcd} \tag{99}$$

Examples E.30. In practice and in the literrature, we often say that some construction satisfies a universal property without referring to the actual universal morphism. For example, we say that the free monoid satisfies a universal property, while the less ambiguous thing to say is that the inclusion of a set A into the free monoid A^* is a universal morphism from the set A to the fogetful functor $U : \mathbf{Mon} \rightsquigarrow \mathbf{Set}$.[328] Let us translate the other examples we gave above with this new terminology.

[328] You probably agree that the latter is a mouthful, but the former can feel very vague, especially when you are not familiar with the construction or universal properties in general.

1. The quotient map from a group G to its abelianization G^{ab} is the universal morphism from G to the forgetful functor $\mathbf{Ab} \rightsquigarrow \mathbf{Grp}$.

2. The set $S \subseteq V$ is a basis for the vector space V when the inclusion $S \hookrightarrow V$ is the universal morphism from S to the forgetful functor $\mathbf{Vect}_k \rightsquigarrow \mathbf{Set}$.

3. An exponential object is an object A^X along with the universal morphism ev from the functor $- \times X$ to A.[329]

4. A subobject classifier is a morphism $\top : \mathbf{1} \rightarrowtail \Omega$ such that the corresponding function $\top : \mathbf{1} \to \mathrm{Sub}_{\mathbf{C}}(\Omega)$ is the universal morphism from $\mathbf{1}$ to the functor $\mathrm{Sub}_{\mathbf{C}}$.

5. A power object of X is an object $\mathfrak{P}X$ along with the universal morphism $\in_X$ from $\mathbf{1}$ to $\mathrm{Sub}_{\mathbf{C}}(X \times -)$.

[329] This is an example of a universal morphism from a functor to an object, whereas all the other examples are universal morphisms from an object to a functor.

Another common practice is to use the word free in situations where we have a universal morphism to a forgetful functor (just like the free monoid). For instance, one could say that G^{ab} is the free abelian group over G, or that V is the free vector space over its basis. When you have two categories with an obvious forgetful functor between them, it can be useful to figure out if you can construct free objects. We will get back to this in Chapter H.

A first approximation of the definition of universality is to say that a universal property is the property of being a universal morphism from X to F or from F to X. Unfortunately, this is too constrained. For instance, as we have said, the universal property of NNOs does not correspond to a universal morphism like that. Another example is subobject classifiers in categories that are not well-powered. In such categories, Sub is not a functor into **Set**,[330] so we cannot have a universal morphism from X to Sub.

[330] There might be another suitable codomain for that functor, but let us not think too hard about size issues.

In the next section, we will see that universal morphisms are initial or terminal objects in a comma category. It turns out that in the most general terms, being universal is best defined as being initial or terminal is some category. It may seem vague at first, but this perfectly describes all the universal properties we have used so far that fit the template "for all ... there exists a unique morphism ..."

Definition E.31 (Universal property)**.** A **universal property** is the property of being initial or terminal in a category.[331]

[331] This rather underwhelming definition is also what led me to postpone it to this point, after we have seen many examples and uses of universal properties.

It readily follows (using Proposition C.33 and Corollary C.34) that universal properties determine things up to isomorphism.

OL **Exercise E.32.** Show that in any category $\mathbf{C}$ with a terminal object $\mathbf{1}$ (even if $\mathbf{C}$ is not well-powered), we can define a category whose objects are monomorphisms in $\mathbf{C}$ and $\top : X \rightarrowtail \Omega$ is terminal if and only if it is the subobject classifier in $\mathbf{C}$. In particular, if $\top$ is terminal in that category, then X is terminal in $\mathbf{C}$.

E.3 Comma Categories

Before moving on, we are going to have some fun with new definitions that let us construct new categories out of categories and functors. This section could have appeared in earlier chapters, but those were already dense, and this section ends with a more concise definition of universal morphisms as initial or terminal objects in comma categories.

Definition E.33 (Comma category)**.** Given two functors $\mathbf{D} \overset{F}{\rightsquigarrow} \mathbf{C} \overset{G}{\leftsquigarrow} \mathbf{E}$, there is a category $F \downarrow G$,[332] called the **comma category**, whose objects are triples (X, Y, α) with $X \in \mathbf{D}_0$, $Y \in \mathbf{E}_0$ and $\alpha : F(X) \to G(Y)$ (in $\mathbf{C}_1$), and morphisms between (X_1, Y_1, α) and (X_2, Y_2, β) are pairs of morphisms $f : X_1 \to X_2$ in $\mathbf{D}_1$ and $g : Y_1 \to Y_2$ in $\mathbf{E}_1$ yielding a commutative square as in (100).

[332] Some authors denote this category F/G.

$$\begin{tikzcd} F(X_1) \arrow[r, "F(f)"] \arrow[d, "\alpha"'] & F(X_2) \arrow[d, "\beta"] \\ G(Y_1) \arrow[r, "G(g)"'] & G(Y_2) \end{tikzcd} \tag{100}$$

The identity morphism on (X, Y, α) is the pair $(\mathrm{id}_X, \mathrm{id}_Y)$ making (101) commute. The composition of (f, g) and (f', g') is $(f' \circ f, g' \circ f)$, it makes the following commute by paving with the commutative squares induced by (f, g) and (f', g').

$$\begin{tikzcd} FX \arrow[r, "F\mathrm{id}_X = \mathrm{id}_{FX}"] \arrow[d, "\alpha"'] & FX \arrow[d, "\alpha"] \\ GY \arrow[r, "G\mathrm{id}_Y = \mathrm{id}_{GY}"'] & GY \end{tikzcd} \tag{101}$$

$$\begin{tikzcd} FX_1 \arrow[r, "Ff"'] \arrow[rr, bend left, "F(f' \circ f)"] \arrow[d, "\alpha"'] & FX_2 \arrow[r, "Ff'"'] & FX_3 \arrow[d, "\gamma"] \\ GY_1 \arrow[r, "Gg"] \arrow[rr, bend right, "G(g' \circ g)"'] & GY_2 \arrow[r, "Gg'"] & GY_3 \end{tikzcd} \tag{102}$$

OL **Exercise E.34.** Given two functors $\mathbf{D} \overset{F}{\rightsquigarrow} \mathbf{C} \overset{G}{\leftsquigarrow} \mathbf{E}$ and their comma category $F \downarrow G$, show there are two forgetful functors $U_F : F \downarrow G \rightsquigarrow \mathbf{D}$ and $U_G : F \downarrow G \rightsquigarrow \mathbf{E}$ that send (X, Y, α) to X and to Y respectively.

Example E.35 (NNO)**.** Let $\mathbf{C}$ be a category with a terminal object and a NNO, and let $\mathbf{1} + - : \mathbf{C} \rightsquigarrow \mathbf{C}$ be the maybe functor. The natural numbers object is the initial object in $(\mathbf{1} + -) \downarrow \mathrm{id}_{\mathbf{C}}$. The morphisms $0 : \mathbf{1} \to \mathfrak{N}$ and $\mathsf{succ} : \mathfrak{N} \to \mathfrak{N}$ can be copaired in $[0, \mathsf{succ}] : \mathbf{1} + \mathfrak{N} \to \mathfrak{N}$ that is an object of this comma category. An arbitrary object of $(\mathbf{1} + -) \downarrow \mathrm{id}_{\mathbf{C}}$ is a morphism $f : \mathbf{1} + X \to X$ which we can decompose as $[f \circ \kappa_{\mathbf{1}}, f \circ \kappa_X]$. Writing $z = f \circ \kappa_{\mathbf{1}}$ and $s = f \circ \kappa_X$, by the universal property of the NNO, there is a unique morphism making (97) commute. Equivalently, (103) commutes,[333] which means ! is the unique morphism from $[0, \mathsf{succ}]$ to f in the comma category $(\mathbf{1} + -) \downarrow \mathrm{id}_{\mathbf{C}}$.

$$\begin{tikzcd} \mathbf{1} + \mathfrak{N} \arrow[r, dashed, "\mathrm{id}_{\mathbf{1}} + !"] \arrow[d, "{[0,\mathsf{succ}]}"'] & \mathbf{1} + X \arrow[d, "{f = [z,s]}"] \\ \mathfrak{N} \arrow[r, dashed, "!"'] & X \end{tikzcd} \tag{103}$$

[333] If (97) commutes, we have $z = ! \circ 0$ and $s \circ ! = ! \circ \mathsf{succ}$. Thus, we have

$$\begin{aligned} [z, s] \circ (\mathrm{id}_{\mathbf{1}} + !) &= [z, s \circ !] \\ &= [! \circ 0, ! \circ \mathsf{succ}] \\ &= ! \circ [0, \mathsf{succ}]. \end{aligned}$$

Conversely, if (103) commutes, the same derivation shows $[z, s \circ !] = [! \circ 0, ! \circ \mathsf{succ}]$. By Corollary D.74, we must have $z = ! \circ 0$ and $s \circ ! = ! \circ \mathsf{succ}$.

Definition E.36 (Arrow category)**.** In the setting of Definition E.33, if $F = G = \mathrm{id}_{\mathbf{C}}$, then $\mathrm{id}_{\mathbf{C}} \downarrow \mathrm{id}_{\mathbf{C}}$ is called the **arrow category** of $\mathbf{C}$ and denoted $\mathbf{C}^{\to}$. Its objects are morphisms in $\mathbf{C}$ and its morphisms are commutative squares in $\mathbf{C}$.[334] It may remind you of the category defined in Exercise E.32.

[334] Less concisely, a morphism $\phi : f \to g$ between morphisms $f : X \to Y$ and $g : X' \to Y'$ is a pair of morphisms $\phi_X : X \to X'$ and $\phi_Y : Y \to Y'$ making (**??**) commute.

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "\phi_X"'] & Y \arrow[d, "\phi_Y"] \\ X' \arrow[r, "g"'] & Y' \end{tikzcd} \tag{104}$$

OL **Exercise E.37.** Let $\mathbf{C}$ be a category (note the change of font to distinguish the functors from their action).

1. Show that $\mathsf{id} : \mathbf{C} \rightsquigarrow \mathbf{C}^{\to}$ sending $X \in \mathbf{C}_0$ to id_X is functorial.

2. Show that $\mathsf{s} : \mathbf{C}^{\to} \rightsquigarrow \mathbf{C}$ sending $f \in \mathbf{C}_0^{\to}$ to $s(f)$ is functorial.

3. Show that $\mathsf{t} : \mathbf{C}^{\to} \rightsquigarrow \mathbf{C}$ sending $f \in \mathbf{C}_0^{\to}$ to $t(f)$ is functorial.

OL **Exercise E.38.** Show the assignment $\mathbf{C} \mapsto \mathbf{C}^{\to}$ yields a functor $\mathbf{Cat} \rightsquigarrow \mathbf{Cat}$.

Definition E.39 (Slice category)**.** In the setting of Definition E.33, if $F = \mathrm{id}_{\mathbf{C}}$ and $G = \Delta(X) : \mathbf{1} \rightsquigarrow \mathbf{C}$ is a constant functor selecting one object $G(\bullet) = X \in \mathbf{C}_0$, then

$\mathrm{id}_{\mathbf{C}} \downarrow \Delta(X)$ is called the **slice category** over X and denoted $\mathbf{C}/X$.[335] Its objects are morphisms in $\mathbf{C}$ with target X and its morphisms are commutative triangles with X as a tip as in (105).

[335] Some authors call this category $\mathbf{C}$ over X.

$$\begin{tikzcd} A \arrow[rr] \arrow[dr] & & B \arrow[dl] \\ & X & \end{tikzcd} \tag{105}$$

Identity morphisms are commutative triangles with the top morphism being identity and composition is done by combining triangles as in (106).

$$\begin{tikzcd} A \arrow[r] \arrow[dr] & B \arrow[r] \arrow[d] & C \arrow[dl] \\ & X & \end{tikzcd} \tag{106}$$

OL **Exercise E.40** (NOW!)**.** Suppose $\mathbf{C}$ has a terminal object $\mathbf{1}$, what is $\mathbf{C}/\mathbf{1}$?

Example E.41. Recall that $\Omega = \{\perp, \top\}$ is the subobject classifier in **Set**, that is, a function $A \to \Omega$ can be identified with the subset $f^{-1}(\top) \subseteq A$. Therefore, objects of $\mathbf{Set}/\Omega$ can be seen as sets A equipped with a distinguished subset $P \subseteq A$ that we will call a predicate.[336] Suppose (A, P_A) and (B, P_B) are sets equipped with predicates, what is a morphism $(A, P_A) \to (B, P_B)$ when we see these as objects in $\mathbf{Set}/\Omega$? It is a function $f : A \to B$ making (107) commute.[337]

[336] This terminology comes from the field of logic. You can think of predicates as things that might be satisfied or not by elements of a set. We say that $a \in A$ satisfies P if $a \in P$.

[337] Recall that $\chi_{P_A}(a) = \top \Leftrightarrow a \in P_A$ and similarly for P_B.

$$\begin{tikzcd} A \arrow[rr, "f"] \arrow[dr, "\chi_{P_A}"'] & & B \arrow[dl, "\chi_{P_B}"] \\ & \Omega & \end{tikzcd} \tag{107}$$

Equivalently, f must satisfy $a \in P_B \implies f(a) \in P_B$. Logically-minded people might call $\mathbf{Set}/\Omega$ the category of predicates and predicate-preserving functions. We can also view a predicate as a unary relation on A, and we recognize $\mathbf{Set}/\Omega$ is the category $1\mathbf{Rel}$.

OL **Exercise E.42.** Let $\mathbf{C}$ be a category with all finite products and fix $n \in \mathbb{N}$. Show the assignment $X \mapsto X^n = X \times \overset{n}{\cdots} \times X$ is functorial. Using this functor and intuition from the previous example, define $n\mathbf{Rel}$ as a comma category.

Definition E.43 (Coslice category)**.** In the setting of Definition E.33, if $G = \mathrm{id}_{\mathbf{C}}$ and $F = \Delta(X) : \mathbf{1} \rightsquigarrow \mathbf{C}$ is a constant functor selecting one object $F(\bullet) = X \in \mathbf{C}_0$, then $\Delta(X) \downarrow \mathrm{id}_{\mathbf{C}}$ is called the **coslice category** under X and denoted $X/\mathbf{C}$.[338] Its objects are morphisms in $\mathbf{C}$ with source X and its morphisms are commutative triangles with X as a tip as in (108).[339]

[338] Some authors call this category $\mathbf{C}$ under X.

[339] We leave you to dualize the definition of identities and composition from the definition of slice categories.

$$\begin{tikzcd} & X \arrow[dl] \arrow[dr] & \\ A \arrow[rr] & & B \end{tikzcd} \tag{108}$$

Example E.44. In the solution to Exercise D.1, we saw that a function $\mathbf{1} \to X$ in **Set** can be identified with the element of X it picks out. Therefore, objects of $\mathbf{1}/\mathbf{Set}$ can be seen as sets A equipped with a distinguished element $a \in A$. We already have a name for these things, they are pointed sets. Suppose (A, a) and (B, b) are pointed sets, what is a morphism $(A, a) \to (B, b)$ when we see these as objects of $\mathbf{1}/\mathbf{Set}$? It is a function $f : A \to B$ making (109) commute.

$$\begin{tikzcd} & \mathbf{1} \arrow[dl, "a"'] \arrow[dr, "b"] & \\ A \arrow[rr, "f"'] & & B \end{tikzcd} \tag{109}$$

Equivalently, f must send a to b, i.e., $f(a) = b$. You might now recognize that $\mathbf{1}/\mathbf{Set}$ is really the category $\mathbf{Set}_*$ in disguise.

This example suggests we can define an abstract and general way of defining "pointed" things. However, recall that sometimes, $\mathbf{1}$ is not the right object to talk about elements. For instance, in $\mathbf{Grp}$, $\mathbf{1}$ is also initial so, by the dual to Exercise E.40, $\mathbf{1}/\mathbf{Grp}$ is the same thing as $\mathbf{Grp}$. Still, we can easily define the category $\mathbf{Grp}_*$ of pointed groups: its objects are pairs (G, g) where G is a group and $g \in G$, and morphisms $(G, g) \to (H, h)$ are homomorphisms $f : G \to H$ satisfying $f(g) = h$.

OL **Exercise E.45.** Let $\mathbb{Z}$ be the group of integers equipped with addition. Show that one can define the category $\mathbf{Grp}_*$ as $\mathbb{Z}/\mathbf{Grp}$.

OL **Exercise E.46.** Show that for any category $\mathbf{C}$ and object $X \in \mathbf{C}_0$, the slice category $\mathbf{C}/X$ has a terminal object. State and prove the dual statement.

OL **Exercise E.47.** Show that the product of $f : A \to X$ and $g : B \to X$ in $\mathbf{C}/X$ exists if and only if the pullback of $A \xrightarrow{f} X \xleftarrow{g} B$ exists in $\mathbf{C}$. State and prove the dual statement.

These results can be summarized by saying that pullbacks are products in the slice category, and pushouts are coproducts in the coslice category. This allows us to define arbitrary (not binary) pullbacks and pushouts as arbitrary products and coproducts in the slice and coslice categories.[340]

[340] In the literature, these are called **fibered products** and **fibered sums** respectively.

OL **Exercise E.48.** Given two functors $\mathbf{D} \overset{F}{\rightsquigarrow} \mathbf{C} \overset{G}{\leftsquigarrow} \mathbf{E}$, show that an initial object in $F \downarrow G$ is a terminal object in $G^{\mathrm{op}} \downarrow F^{\mathrm{op}}$.

Back to universal properties. We give a more concise definition.

Proposition E.49. *Let $F : \mathbf{D} \rightsquigarrow \mathbf{C}$ be a functor, $X \in \mathbf{C}_0$ and $\Delta(X) : \mathbf{1} \rightsquigarrow \mathbf{C}$ be the constant functor. A **universal morphism** from X to F is an initial object in $\Delta(X) \downarrow F$.*

Proof. Unrolling the definition of initial object in $\Delta(X) \downarrow F$, we find that it is a morphism $a : X \to F(A)$ such that for any other morphism $b : X \to F(B)$, there is unique morphism $(\bullet, A, a) \to (\bullet, B, b)$, that is, a unique morphism $f : A \to B$ making (110) commute.

$$\begin{tikzcd} X \arrow[r, "\mathrm{id}_X"] \arrow[d, "a"'] & X \arrow[d, "b"] \\ FA \arrow[r, "Ff"'] & FB \end{tikzcd} \tag{110}$$

This is exactly the situation depicted in (98). □

Corollary E.50 (Dual)**.** *A universal morphism from F to X is a terminal object in $F \downarrow \Delta(X)$.*

Proof. We said that a universal morphism from F to X is a universal morphism from $X \in \mathbf{C}^{\mathrm{op}}$ to F^{op}. By the previous result, it is an initial object in $\Delta(X) \downarrow F^{\mathrm{op}}$. By Exercise E.48, it is a terminal object in $F \downarrow \Delta(X)$. □

In case a universal property is realized by a universal morphism, we can formally prove that this property determines an object up to isomorphism.

SOL **Exercise E.51** (NOW!)**.** Show that if there is a universal morphism from X to F and one from Y to F, then $X \cong Y$. State and prove the dual statement.

We have to postpone to Chapter G showing that, as we have claimed, any (co)limit satisfies a universal property. Still, you might have noticed that our definition of universal property also uses a special case of (co)limits, that is, initial and terminal objects. What is more, in the following chapters, we will introduce a couple more concepts which often coincide[341] with the concepts of (co)limits and universal properties.

[341] By *coincide*, we mean that one is a special case of the other or vice-versa or both directions.

F Natural Transformations

In the previous chapters, we saw how to use the framework of categories to do mathematics. While fundamentally the same as "classical" mathematics,[342] doing mathematics with categories can feel different because we study mathematical structures from above rather than from the inside. Now, if we want to study group theory categorically, we have many options:

[342] We rely on rigorous logical arguments.

- We can study single-object categories where every morphism is invertible (deloopings of groups) and functors between them (group homomorphisms).[343]

[343] This amounts to doing "classical" group theory.

- We can go one step higher and study the category **Grp** as a whole. We do not have access to what is inside a group, only how groups relate to each other.[344]

[344] This has been our point of view until now.

- We can climb another step and study **Grp** as an object of a category of categories.[345]

[345] Recall that due to size issues, **Grp** is not an object of **Cat**, but we could carefully define a category of categories that contains **Grp**.

- In between the previous two items, we can study **Grp** as a subcategory of **Cat**. Taking the delooping is a fully faithful functor $\mathbf{B} : \mathbf{Grp} \rightsquigarrow \mathbf{Cat}$, so we identify **Grp** with its image in **Cat**. We still get to study how groups interact with each other, but also how they interact with other categories.

The first and last step are particular to groups, not all mathematical structures can be viewed as a categories. For instance, studying group theory requires to understand group homomorphisms which are functors, not categories. Taking the categorical mindset to the extreme,[346] we should only have to study how homomorphisms relate to each other, but what is a morphism between homomoprhisms? More generally, what is a morphism between functors?

[346] This might seem extreme at this point, but category theorists can go way further.

F.1 Functor Categories

Natural transformations are admittedly what made mathematicians want to study category theory in the first place. In short, they are morphisms between functors.

The abstract structure of a category is very familiar because it resembles what is found in algebraic structures such as groups, rings or vector spaces.[347] That is to say, it consists of the data of one or more sets with one or more operations satisfying one or more properties. The intuition for morphisms of algebraic structures ported well to categories: a functor comrpises functions between the carrier sets (object and morphisms) that preserve the operations (composition, source and target).

[347] In fact, it is technically called an essentially algebraic structure.

Unfortunately, the definition of a functor does not fit this pattern. It is hard to describe what is the "structure" of a functor. A first step towards defining morphisms between functors is to do it in some special cases.

Following the introduction, you can try to find a satisfying definition of morphism between group homomorphisms $f, g : G \to H$,[348] and then figure out its meaning when f and g are seen as functors $\mathbf{B}G \rightsquigarrow \mathbf{B}H$.

[348] Recall that morphisms should compose and there should be an identity morphism.

We will proceed with another special case. Given a functor $F : \mathbf{C} \rightsquigarrow \mathbf{Set}$, we would like to know what is a *subfunctor* of F.[349] To every object $X \in \mathbf{C}_0$, F assigns a set FX. It makes sense that a subfunctor F' sends X to a subset $F'X \subseteq FX$. To every $f \in \mathrm{Hom}_{\mathbf{C}}(X, Y)$, F assigns a function $Ff : FX \to FY$. It makes sense that a subfunctor F' sends f to a restriction of Ff on the domain $F'X$. Moreover, we need to require the image of $F'f$ (Ff restricted to $F'X$) lies in $F'Y$, otherwise the target of $F'f$ cannot be $F'Y$. We can summarize the constraints on F' with the following commutative square.[350]

[349] If we had a notion of morphisms between functors, we could define a subfunctor as a subobject, i.e. an equivalence class of monomorphisms.

$$\begin{tikzcd} F'X \arrow[r, hook] \arrow[d, "F'f"'] & FX \arrow[d, "Ff"] \\ F'Y \arrow[r, hook] & FY \end{tikzcd} \tag{111}$$

[350] (111) commutes if and only if $F'f$ is the restriction of Ff to $F'X$.

It turns out this is enough to ensure that F' is a functor. Indeed, $F'(\mathrm{id}_X)$ is the identity map on FX restricted to $F'X$, which is the identity map on $F'X$. Also, for any $f : X \to Y$ and $g : Y \to X$, $F'f \circ F'g$ is the restriction of $F(g \circ f) = Fg \circ Ff$ to $F'X$.[351]

[351] You can check this manually, or pave the following diagram with the squares showing $F'f$ is Ff restricted to $F'X$ and $F'g$ is Fg restricted to $F'Y$.

$$\begin{tikzcd} F'X \arrow[r, hook] \arrow[dd, "F'(g\circ f)"'] & FX \arrow[d, "Ff"'] \arrow[dd, bend left, "F(g\circ f)"] \\ & FY \arrow[d, "Fg"'] \\ F'Z \arrow[r, hook] & FZ \end{tikzcd} \tag{112}$$

Example F.1. Let F be the maybe functor on **Set** and F' be the identity functor. One can verify that the family of inclusions of X inside $X + \mathbf{1}$ for all sets X yields commutative squares like (111).

We can generalize this to functors with arbitrary codomains.

OL **Exercise F.2.** Let $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ be a functor. Suppose that for every $X \in \mathbf{C}_0$, there is a monomorphism $F'X \rightarrowtail FX$, and for every $f \in \mathrm{Hom}_{\mathbf{C}}(X, Y)$, there is a morphism $F'f$ making (111) commute. Show that F' is a functor $\mathbf{C} \rightsquigarrow \mathbf{D}$.

This does not strictly define a subfunctor because we still need to quotient by some equivalence saying when two functors represent the same subfunctor of F. Informally, if $F'X \rightarrowtail X$ and $F''X \rightarrowtail X$ always represent the same subobject in the same way, then F' and F'' represent the same subfunctor. To make this formal, we define morphisms of functors in full generality.

Definition F.3 (Natural transformation)**.** Let $F, G : \mathbf{C} \rightsquigarrow \mathbf{D}$ be two (covariant) functors, a **natural transformation** $\phi : F \Rightarrow G$ is a map $\phi : \mathbf{C}_0 \to \mathbf{D}_1$ that satisfies $\phi(A) \in \mathrm{Hom}_{\mathbf{D}}(FA, GA)$ for all $A \in \mathbf{C}_0$ and makes (113) commute for any $f \in \mathrm{Hom}_{\mathbf{C}}(A, B)$.[352]

$$\begin{tikzcd} F(A) \arrow[r, "\phi(A)"] \arrow[d, "F(f)"'] & G(A) \arrow[d, "G(f)"] \\ F(B) \arrow[r, "\phi(B)"'] & G(B) \end{tikzcd} \tag{113}$$

[352] When doing proofs relying on naturality (i.e. the property of being natural), we will use (113) where we instantiate ϕ, F, G, A, B and f with the natural transformation, functors, objects and morphism that is needed in the proof. In order to make this instantiation less painful, we will use the shorthand $\mathsf{NAT}(\phi, A, B, f)$ and instantiate the parameters (we can omit F and G because they should be known from the context). I will try to be this precise whenever I use naturality, but it is very common to simply write "by naturality of ϕ" instead of $\mathsf{NAT}(\phi, A, B, f)$.

Each $\phi(A)$ will be called a **component** of ϕ and may also be denoted with ϕ_A.

As usual, there is an **identity transformation** $\mathbb{1}_F : F \Rightarrow F$[353], it sends every object A to the identity map $\mathrm{id}_{F(A)}$. In the setting of Exercise F.2, the monomorphisms $F'X \rightarrowtail FX$ are the components of a natural transformation $F' \Rightarrow F$.[354] Let us go back to our quest to define morphisms of group homomorphisms.

[353] The $\Rightarrow$ (`\Rightarrow`) notation is used more generally for morphisms between morphisms.

[354] To actually define subfunctors, we still need to tell you how to compose natural transformations, but we are not done with examples.

Example F.4. Let $f, g : \mathbf{B}G \rightsquigarrow \mathbf{B}H$ be functors (i.e. group homomorphisms), both send the unique object $*$ in $\mathbf{B}G$ to $*$ in $\mathbf{B}H$. Thus, a natural transformation $\phi : f \Rightarrow g$ has a single component $\phi(*) : * \to *$ in H, which is simply an element $\phi \in H$. The commutativity condition is then exhibited by diagram (114) (which lives in $\mathbf{B}H$) for any $x \in G$.

$$\begin{tikzcd} * \arrow[r, "\phi"] \arrow[d, "f(x)"'] & * \arrow[d, "g(x)"] \\ * \arrow[r, "\phi"'] & * \end{tikzcd} \tag{114}$$

Recall that composition in $\mathbf{B}H$ is just multiplication in H, so naturality of ϕ says that for any $x \in G$, $\phi \cdot f(x) = g(x) \cdot \phi$. Equivalently, $\phi f(x) \phi^{-1} = g(x)$. Therefore, $g = c_\phi \circ f$ where c_ϕ denotes conjugation by ϕ.[355] In short, natural transformations between group homomorphisms correspond to factorizations through conjugations.

[355] In a group $(H, \cdot)$, **conjugation** by an element $h \in H$ is the homomorphism c_h defined $x \mapsto hxh^{-1}$.

Next, a concrete example closer to the general idea of a natural transformation.

Example F.5. Fix some $n \in \mathbb{N}$ and define the functor $\mathrm{GL}_n : \mathbf{CRing} \rightsquigarrow \mathbf{Grp}$ by[356]

$$R \mapsto \mathrm{GL}_n(R) \text{ for any commutative ring } R \text{ and}$$
$$f \mapsto \mathrm{GL}_n(f) \text{ for any ring homomorphism } f.$$

[356] The map $\mathrm{GL}_n(f)$ is just the extension of f on $\mathrm{GL}_n(R)$ by applying f to every element of the matrices.

The second functor is $(-)^\times : \mathbf{CRing} \rightsquigarrow \mathbf{Grp}$ which sends a commutative ring R to its group of units $R^\times$ and a ring homomorphism f to $f^\times$, its restriction on $R^\times$. Checking these mappings define two (covariant) functors is left as an exercise, but one might expect these to be functors as they play nicely with the structure of the objects involved.

A natural transformation between these two functors is $\det : \mathrm{GL}_n \Rightarrow (-)^\times$ which maps a commutative ring R to $\det_R$, the function calculating the determinant of a matrix in $\mathrm{GL}_n(R)$. The first thing to check is that $\det_R \in \mathrm{Hom}_{\mathbf{Grp}}(\mathrm{GL}_n(R), R^\times)$ which is clear because the determinant of an invertible matrix is always a unit, $\det_R(I_n) = 1$ and $\det_R$ is a multiplicative map.[357] The second thing is to verify that diagram (115) commutes for any $f \in \mathrm{Hom}_{\mathbf{CRing}}(R, S)$:

[357] i.e. $\det_R(AB) = \det_R(A)\det_R(B)$.

$$\begin{tikzcd} \mathrm{GL}_n(R) \arrow[r, "\det_R"] \arrow[d, "\mathrm{GL}_n(f)"'] & R^\times \arrow[d, "f^\times = f|_{R^\times}"] \\ \mathrm{GL}_n(S) \arrow[r, "\det_S"'] & S^\times \end{tikzcd} \tag{115}$$

We will check the claim for $n = 2$, but the general proof should only involve more notation to write the bigger expressions, no novel idea. Let $a, b, c, d \in R$, we have

$$\begin{aligned}(\det_S \circ \mathrm{GL}_2(f))\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) &= \det_S\left(\begin{bmatrix} f(a) & f(b) \\ f(c) & f(d) \end{bmatrix}\right) \\ &= f(a)f(d) - f(b)f(c) \\ &= f(ad - bc) \\ &= f^{\times}(ad - bc) \\ &= (f^{\times} \circ \det_R)\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right).\end{aligned}$$

We conclude that the diagram commutes and that det is indeed a natural transformation.[358]

[358] Modulo the cases $n > 2$.

OL **Exercise F.6.** Recall the functors $\mathsf{s}, \mathsf{t} : \mathbf{C}^{\rightarrow} \rightsquigarrow \mathbf{C}$ defined in Exercise E.37. Show that $\phi : \mathsf{s} \Rightarrow \mathsf{t}$ defined by $\phi(f) = f$ for any $f \in \mathbf{C}_0^{\rightarrow} = \mathbf{C}_1$ is a natural transformation.

Because naturality is such a central idea to category theory (just as important as functoriality), we often use it post-rigorously. For instance, when studying a mathematical object X, we might follow some process to obtain another object $F(X)$, and another construction might yield $G(X)$, then we find a process ϕ to go from $F(X)$ to $G(X)$ and we say ϕ is **natural in** X. With these last three words, we implicitly mean a lot of things: that X is an object of some category, that F and G are functors from that category, and that ϕ is the component at X of a natural transformation $F \Rightarrow G$.

It is also possible that F and G take more than one parameter.

OL **Exercise F.7** (NOW!)**.** Let $F, G : \mathbf{C} \times \mathbf{C}' \rightsquigarrow \mathbf{D}$ be two functors. Show that a family

$$\{\phi_{X,Y} : F(X,Y) \to G(X,Y) \mid X \in \mathbf{C}_0, Y \in \mathbf{C}'_0\}$$

is a natural transformation if and only if for any $X \in \mathbf{C}_0$ and $Y \in \mathbf{C}'_0$, both[359]

$$\phi_{X,-} : F(X,-) \Rightarrow G(X,-) \text{ and } \phi_{-,Y} : F(-,Y) \Rightarrow G(-,Y)$$

are natural.

[359] Recall the definition of $F(X,-)$ and $F(-,Y)$ from Exercise B.44. If only one of $\phi_{X,-}$ or $\phi_{-,Y}$ is natural, we say that ϕ is natural in X only, respectively Y only. In words, this exercise says that ϕ is natural in X and Y if and only if it is natural in X and natural in Y.

Examples F.8 (Natural isomorphisms)**.** A **natural isomorphism** is a natural transformation whose components are all isomorphisms. We have already encountered several of them.

1. When defining exponentials, we saw that currying is a bijection $\mathrm{Hom}_{\mathbf{C}}(B \times X, A) \cong \mathrm{Hom}_{\mathbf{C}}(B, A^X)$. It turns out this is a natural isomorphism from the functor $\mathrm{Hom}_{\mathbf{C}}(- \times X, A) : \mathbf{C}^{\mathrm{op}} \rightsquigarrow \mathbf{Set}$ to $\mathrm{Hom}_{\mathbf{C}}(-, A^X) : \mathbf{C}^{\mathrm{op}} \rightsquigarrow \mathbf{Set}$. We simply need to check the square below commutes for any $f : B \to B'$.[360]

$$\begin{tikzcd} \mathrm{Hom}_{\mathbf{C}}(B \times X, A) \arrow[r, "g \mapsto \lambda g"] & \mathrm{Hom}_{\mathbf{C}}(B, A^X) \\ \mathrm{Hom}_{\mathbf{C}}(B' \times X, A) \arrow[u, "-\circ(f \times \mathrm{id}_X)"] \arrow[r, "g \mapsto \lambda g"'] & \mathrm{Hom}_{\mathbf{C}}(B', A^X) \arrow[u, "-\circ f"'] \end{tikzcd} \tag{116}$$

[360] Because these functors have $\mathbf{C}^{\mathrm{op}}$ as a source, note the reversal the arrows

Starting with g in the bottom left, we need to prove $\lambda g \circ f = \lambda\,(g \circ (f \times \mathrm{id}_X))$. The universal property of A^X tells us $\mathsf{ev} \circ (\lambda g \times \mathrm{id}_X) = g$. Pre-composing with $f \times \mathrm{id}_X$, we find

$$g \circ (f \times \mathrm{id}_X) = \mathsf{ev} \circ (\lambda g \times \mathrm{id}_X) \circ (f \times \mathrm{id}_X) = \mathsf{ev} \circ ((\lambda g \circ f) \times \mathrm{id}_X),$$

thus both $\lambda g \circ f$ and $\lambda\,(g \circ (f \times \mathrm{id}_X))$ make (117) commute, and they must be equal by uniqueness.

$$\begin{tikzcd} A & A^X \times X \arrow[l, "\mathsf{ev}"'] \\ & B \times X \arrow[ul, "g\circ(f\times \mathrm{id}_X)"] \arrow[u, dashed, "\lambda g\circ f = \lambda(g\circ(f\times\mathrm{id}_X))"'] \end{tikzcd} \quad (117)$$

2. Without giving all the details, we note that the bijections

$$\mathrm{Hom}_{\mathbf{Set}}(A, M) \cong \mathrm{Hom}_{\mathbf{Mon}}(A^*, M), \text{ and}$$
$$\mathrm{Hom}_{\mathbf{Grp}}(G, A) \cong \mathrm{Hom}_{\mathbf{Ab}}(G^{\mathrm{ab}}, A)$$

are also natural in A and M, and A and G respectively. They are the components of natural isomorphisms[361]

$$\mathrm{Hom}_{\mathbf{Set}}(-, U-) \cong \mathrm{Hom}_{\mathbf{Mon}}(-^*, -), \text{ and}$$
$$\mathrm{Hom}_{\mathbf{Grp}}(-, U-) \cong \mathrm{Hom}_{\mathbf{Ab}}(-^{\mathrm{ab}}, -).$$

[361] Where U denotes the forgetful functors $\mathbf{Mon} \rightsquigarrow \mathbf{Set}$ and $\mathbf{Ab} \rightsquigarrow \mathbf{Grp}$ respectively.

In particular, the assignments $A \mapsto A^*$ and $G \mapsto G^{\mathrm{ab}}$ are functorial, and these natural isomorphisms are witnesses to these functors being left adjoints to the corresponding forgetful functors.[362]

[362] Adjoints are the topic of Chapter H, where we will study more of these kind of natural isomorphisms.

Now, coming back to our idea that natural transformations are morphisms of functors, we shall explain how they compose.

Definition F.9 (Vertical composition)**.** Let $F, G, H : \mathbf{C} \rightsquigarrow \mathbf{D}$ be parallel functors and $\phi : F \Rightarrow G$ and $\eta : G \Rightarrow H$ be two natural transformations. The **vertical composition** of ϕ and η, denoted $\eta \cdot \phi : F \Rightarrow H$ is defined by $(\eta \cdot \phi)(A) = \eta(A) \circ \phi(A)$ for all $A \in \mathbf{C}_0$. If $f : A \to B$ is a morphism in $\mathbf{C}$, then diagram (118) commutes by naturality of ϕ and η, showing that $\eta \cdot \phi$ is a natural transformation from F to H.

The notation $\cdot$ is not widespread, most authors use $\circ$ because vertical composition is the composition in a functor category. I believe the distinction is helpful as you learn this material.

$$\begin{tikzcd} F(A) \arrow[r, "\phi(A)"] \arrow[d, "F(f)"'] & G(A) \arrow[r, "\eta(A)"] \arrow[d, "G(f)"'] & H(A) \arrow[d, "H(f)"'] \\ F(B) \arrow[r, "\phi(B)"'] & G(B) \arrow[r, "\eta(B)"'] & H(B) \end{tikzcd} \quad (118)$$

The meaning of *vertical* will come to light when horizontal composition is introduced in a bit.

Definition F.10 (Functor categories)**.** For any two categories $\mathbf{C}$ and $\mathbf{D}$, there is a **functor category** denoted $[\mathbf{C}, \mathbf{D}]$.[363] Its objects are functors from $\mathbf{C}$ to $\mathbf{D}$, its morphisms are natural transformations between such functors, and the composition is the vertical composition defined above. We leave you to check the associativity of $\cdot$ as it quickly follows from associativity of composition in $\mathbf{D}$. Similarly, you can verify the identity morphism for a functor F is $\mathbb{1}_F$.

[363] Some authors denote it $\mathbf{D}^{\mathbf{C}}$, analogously to the exponential of sets. In fact, $\mathbf{Cat}$ is cartesian closed and $[\mathbf{C}, \mathbf{D}]$ is the exponential. We give most of the proof in Example F.36.5.

OL **Exercise F.11** (NOW!). Show that natural isomorphisms are precisely the isomorphisms in functor categories.

OL **Exercise F.12.** Let $F, G : \mathbf{C} \rightsquigarrow \mathbf{D}$ be two naturally isomorphic functors. Show that if F is full/faithful/(co)continuous, then so is G.

Functors that are naturally isomorphic are essentially the same functor; they send the same object to isomorphic objects and the same morphism to morphisms that are well-behaved under composition with isomorphisms between the source and targets. This suggests that a natural isomorphism between functors transfers all the properties, we check some of them in Exercise F.12.

Example F.13. Recall that a left action of a group G on a set S is just a functor $\mathbf{B}G \rightsquigarrow \mathbf{Set}$. Now, between two such functors $F, F' \in [\mathbf{B}G, \mathbf{Set}]$, a natural transformation is a single map $\sigma : F(*) \to F'(*)$ such that $\sigma \circ F(g) = F'(g) \circ \sigma$ for any $g \in G$. In other words, denoting $\cdot$ for both group actions on $F(*)$ and on $F'(*)$, σ satisfies $\sigma(g \cdot x) = g \cdot (\sigma(x))$ for any $g \in G$ and $x \in F(*)$. In group theory, such a map is called G–**equivariant**.

Therefore, the category $[\mathbf{B}G, \mathbf{Set}]$ can be identified as the category of G–sets (sets equipped with an action of G) with G–equivariant maps as the morphisms.

Examples F.14. We can recover constructions we have seen before by studying categories of functors with a simple domain.

1. The terminal category $\mathbf{1}$ has a single object $\bullet$ and no morphism other than the identity. Recall that for any category $\mathbf{C}$, a functor $F : \mathbf{1} \rightsquigarrow \mathbf{C}$ is a simply a choice of object $F(\bullet) \in \mathbf{C}_0$ because $F(\mathrm{id}_\bullet)$ must be equal to $\mathrm{id}_{F(\bullet)}$. If $F, G \in [\mathbf{1}, \mathbf{C}]$, then a natural transformation $\phi : F \Rightarrow G$ is simply a choice of morphism $\phi : F(\bullet) \to G(\bullet)$ because the naturality square (119) for the only morphism $\mathrm{id}_\bullet$ is trivially commutative. Since vertical composition is just componentwise composition, $[\mathbf{1}, \mathbf{C}]$ can be identified with the category $\mathbf{C}$ itself.

$$\begin{tikzcd} F(\bullet) \arrow[r, "F(\mathrm{id}_\bullet)"] \arrow[d, "\phi"'] & F(\bullet) \arrow[d, "\phi"] \\ G(\bullet) \arrow[r, "G(\mathrm{id}_\bullet)"'] & G(\bullet) \end{tikzcd} \tag{119}$$

2. Similarly, we can see a functor $F : \mathbf{1} + \mathbf{1} \rightsquigarrow \mathbf{C}$[364] as a choice of two objects $F(\bullet_1)$ and $F(\bullet_2)$ (not necessarily distinct), and a natural transformation $\phi : F \Rightarrow G$ between two such functors as a choice of two morphisms $\phi_1 : F(\bullet_1) \to G(\bullet_1)$ and $\phi_2 : F(\bullet_2) \to G(\bullet_2)$. Therefore, we infer that $[\mathbf{1} + \mathbf{1}, \mathbf{C}]$ can be identified with $\mathbf{C} \times \mathbf{C}$.

[364] Recall $\mathbf{1} + \mathbf{1}$ is the category depicted in (5).

3. Let us go one level harder. A functor $F : \mathbf{2} \rightsquigarrow \mathbf{C}$[365] is a choice of two objects FA and FB as well as a morphism $Ff : FA \to FB$. It can also be seen as a single choice of morphism Ff because FA and FB are determined to be the source and target of Ff respectively. A natural transformation $\phi : F \Rightarrow G$ between two such functors is *not* simply a choice of two morphisms $\phi_A : FA \to GA$ and $\phi_B : FB \to GB$ because, while the naturality squares for id_A and id_B trivially commute, the naturality square (120) for f is an additional constraint on ϕ. Namely, it says (ϕ_A, ϕ_B) makes a commutative square with Ff and Gf, hence we can identify $[\mathbf{2}, \mathbf{C}]$ with the arrow category $\mathbf{C}^{\to}$.

[365] Recall $\mathbf{2}$ is the category depicted in (6).

$$\begin{tikzcd} FA \arrow[r, "Ff"] \arrow[d, "\phi_A"'] & FB \arrow[d, "\phi_B"] \\ GA \arrow[r, "Gf"'] & GB \end{tikzcd} \tag{120}$$

OL **Exercise F.15.** Show that the opposite of $[\mathbf{C}, \mathbf{D}]$ is $[\mathbf{C}^{\mathrm{op}}, \mathbf{D}^{\mathrm{op}}]$.

Viewing any category as a functor category as we did in the previous example has one major consequence formalized in the following results. In short, it says you can infer a lot of things from $[\mathbf{C}, \mathbf{D}]$ by studying $\mathbf{D}$. For instance, if $\mathbf{D}$ has all binary products, it follows that the product of functors F and G in $[\mathbf{C}, \mathbf{D}]$ is the functor sending $X \in \mathbf{C}_0$ to $FX \times GX$ and $f \in \mathbf{C}_1$ to $Ff \times Gf$.[366]

[366] Note that this is not the functor $F \times G$, the latter has type $\mathbf{C} \times \mathbf{C} \rightsquigarrow \mathbf{D} \times \mathbf{D}$.

Theorem F.16. *Let $\mathbf{C}$, $\mathbf{D}$ and $\mathbf{J}$ be categories. If all limits of shape $\mathbf{J}$ exist in $\mathbf{D}$, then all such limits also exist in $[\mathbf{C}, \mathbf{D}]$. Moreover, for any diagram $F : \mathbf{J} \rightsquigarrow [\mathbf{C}, \mathbf{D}]$ and for all $X \in \mathbf{C}_0$, we have*[367]

$$(\lim_{\mathbf{J}} F)(X) = \lim_{\mathbf{J}}(F(-)(X)).$$

[367] This equation is commonly referred to as "limits in functor categories are computed pointwise".

Proof. Let us explain why the equation above makes sense (i.e. is well-typed).

On the L.H.S., since F is a diagram in $[\mathbf{C}, \mathbf{D}]$, its limit will be an object of $[\mathbf{C}, \mathbf{D}]$, namely a functor $\lim_{\mathbf{J}} F : \mathbf{C} \rightsquigarrow \mathbf{D}$. Thus if $X \in \mathbf{C}_0$, then $(\lim_{\mathbf{J}} F)(X)$ is an object in $\mathbf{D}$.

On the R.H.S., fix $X \in \mathbf{C}_0$ and observe that $F(-)(X)$ can be seen as a diagram $\mathbf{J} \rightsquigarrow \mathbf{D}$. Indeed, for $A \in \mathbf{J}_0$, $F(A)$ is a functor from $\mathbf{C}$ to $\mathbf{D}$, so $F(A)(X) \in \mathbf{D}_0$, and for $a : A \to B \in \mathbf{J}_1$, $F(a)$ is a natural transformation from $F(A)$ to $F(B)$, so $F(a)(X)$ (the component of $F(a)$ at X) is a morphism $F(A)(X) \to F(B)(X)$ in $\mathbf{D}$. Then, the limit of $F(-)(X)$ is an object in $\mathbf{D}$ (it exists by hypothesis).

We will define a functor L that sends X to $\lim_{\mathbf{J}}(F(-)(X))$, and we will show it is the limit of F, i.e. $L = \lim_{\mathbf{J}} F$.

First, we need to define the action of L on morphisms. Let $f : X \to Y$, by definition, LX and LY are limits of $F(-)(X)$ and $F(-)(Y)$ respectively, the limit cones are depicted in (121). For any $a : A \to B$, the naturality of $F(a)$ means the front square in (122) commutes, so the family $\{F(A)(f) \circ \pi_{A,X} : LX \to F(A)(Y)\}_{A \in \mathbf{J}_0}$ forms a cone over $F(-)(Y)$, and the universal property of LY yields a unique morphism Lf making all of (122) commute.

$$\begin{tikzcd} & LX \arrow[dl, "\pi_{A,X}"'] \arrow[dr, "\pi_{B,X}"] & \\ F(A)(X) \arrow[rr, "F(a)(X)"'] & & F(B)(X) \\ & LY \arrow[dl, "\pi_{A,Y}"'] \arrow[dr, "\pi_{B,Y}"] & \\ F(A)(Y) \arrow[rr, "F(a)(Y)"'] & & F(B)(Y) \end{tikzcd} \tag{121}$$

$$\begin{tikzcd} & LX \arrow[dl, "\pi_{A,X}"'] \arrow[dr, "\pi_{B,X}"] \arrow[dd, dashed, "Lf"'] & \\ F(A)(X) \arrow[rr, "F(a)(X)"'] \arrow[dd, "F(A)(f)"'] & & F(B)(X) \arrow[dd, "F(B)(f)"] \\ & LY \arrow[dl, "\pi_{A,Y}"'] \arrow[dr, "\pi_{B,Y}"] & \\ F(A)(Y) \arrow[rr, "F(a)(Y)"'] & & F(B)(Y) \end{tikzcd} \tag{122}$$

It follows from uniqueness that $L(\mathrm{id}_X) = \mathrm{id}_{LX}$ and $L(g \circ f) = Lg \circ Lf$ (check that these make (123) and (124) commute). Thus, we have our functor $L : \mathbf{C} \rightsquigarrow \mathbf{D}$.

$$\begin{tikzcd} & LX \arrow[dl, "\pi_{A,X}"'] \arrow[dr, "\pi_{B,X}"] \arrow[dd, dashed, "\mathrm{id}_{LX}"'] & \\ F(A)(X) \arrow[rr, "F(a)(X)"'] \arrow[dd, "F(A)(\mathrm{id}_X)"'] & & F(B)(X) \arrow[dd, "F(B)(\mathrm{id}_X)"] \\ & LX \arrow[dl, "\pi_{A,X}"'] \arrow[dr, "\pi_{B,X}"] & \\ F(A)(X) \arrow[rr, "F(a)(X)"'] & & F(B)(X) \end{tikzcd} \tag{123}$$

$$\begin{tikzcd} & LX \arrow[dl, "\pi_{A,X}"'] \arrow[dr, "\pi_{B,X}"] \arrow[dd, dashed, "Lf"'] & \\ F(A)(X) \arrow[rr, "F(a)(X)"'] \arrow[dd, "F(A)(f)"'] & & F(B)(X) \arrow[dd, "F(B)(f)"] \\ & LY \arrow[dl, "\pi_{A,Y}"'] \arrow[dr, "\pi_{B,Y}"] \arrow[dd, dashed, "Lg"'] & \\ F(A)(Y) \arrow[rr, "F(a)(Y)"'] \arrow[dd, "F(A)(g)"'] & & F(B)(Y) \arrow[dd, "F(B)(g)"] \\ & LZ \arrow[dl, "\pi_{A,Z}"'] \arrow[dr, "\pi_{B,Z}"] & \\ F(A)(Z) \arrow[rr, "F(a)(Z)"'] & & F(B)(Z) \end{tikzcd} \tag{124}$$

Next, the back squares in (122) witness the fact that for any $A \in \mathbf{J}_0$, the morphisms $\pi_{A,X}$ are components of a natural transformation $\pi_A : L \Rightarrow F(A)$. Moreover, for any $a : A \to B \in \mathbf{J}_1$, $F(a) \cdot \pi_A = \pi_B$ holds because the commutativity of the triangles in (122) means for every $X \in \mathbf{C}_0$, $F(a)(X) \cdot \pi_{A,X} = \pi_{B,X}$. We conclude that the family $\{\pi_A : L \Rightarrow F(A)\}_{A \in \mathbf{J}_0}$ forms a cone over F. It remains to prove this is the limit cone.

Suppose $\{\phi_A : L' \Rightarrow F(A)\}_{A \in \mathbf{J}_0}$ is another cone over F, that is $F(a) \cdot \phi_A = \phi_B$ for any $a : A \to B \in \mathbf{J}_1$. Looking at the components at X, we find that $\{\phi_A(X) : L'X \to F(A)(X)\}_{A \in \mathbf{J}_0}$ forms a cone over $F(-)(X)$. Thus, the universal property of

LX yields a unique morphism $!_X$ making (125) commute.

$$\begin{tikzcd}
L'X \arrow[r, dashed, "!_X"] \arrow[rrd, "\phi_B(X)"] \arrow[rdd, "\phi_A(X)"'] & LX \arrow[rd, "\pi_{B,X}"] \arrow[dd, "\pi_{A,X}"] & \\
& & F(B)(X) \\
& F(A)(X) \arrow[ru, "F(a)(X)"'] &
\end{tikzcd} \tag{125}$$

To show $!_X$ is natural in X, we need to show $Lf \circ !_X = !_Y \circ L'f$ for all $f : X \to Y$. Notice that the target of both sides is LY, so it might be possible to use the universal property of LY to conclude the equation holds. More precisely, we need to find a cone over $F(-)(Y)$ with tip $L'X$ and show $Lf \circ !_X$ and $!_Y \circ L'f$ are morphisms of cone, then by uniqueness they must be the same morphism.

The process we used to make the cone over $F(-)(Y)$ with tip LX in (122) still works for $L'X$. We get a cone $\{F(A)(f) \circ \phi_A(X) : L'X \to F(A)(Y)\}_{A \in \mathbf{J}_0}$. Now, the following derivations show that $Lf \circ !_X$ and $!_Y \circ L'f$ are morphisms of cone as depicted in (126). We conclude $!$ is natural, so we have a cone morphism $! : L' \Rightarrow L$.

$$\begin{aligned}
\pi_{A,Y} \circ Lf \circ !_X &= F(A)(f) \circ \pi_{A,X} \circ !_X && (122)\\
&= F(A)(f) \circ \phi_A(X) && (125)
\end{aligned}$$

$$\begin{aligned}
\pi_{A,Y} \circ !_Y \circ L'f &= \phi_A(Y) \circ L'f && (125)\\
&= F(A)(f) \circ \phi_A(X) && \mathsf{NAT}(\phi, X, Y, f)
\end{aligned}$$

$$\begin{tikzcd}
L'X \arrow[r, "!_X"] \arrow[rd, "F(A)(f)\circ\phi_A(X)"'] & LX \arrow[r, "Lf"] & LY \arrow[ld, "\pi_{A,Y}"] \\
& F(A)(Y) &
\end{tikzcd}$$
$$\begin{tikzcd}
L'X \arrow[r, "L'f"] \arrow[rd, "F(A)(f)\circ\phi_A(X)"'] & L'Y \arrow[r, "!_Y"] & LY \arrow[ld, "\pi_{A,Y}"] \\
& F(A)(Y) &
\end{tikzcd} \tag{126}$$

Finally, for any other cone morphism $? : L' \Rightarrow L$, the component of $?$ at X make (125) commute, but $!_X$ is unique with this property. Hence $?_X = !_X$ for all $X \in \mathbf{C}_0$, and we conclude $?$ and $!$ coincide. We conclude that $\lim_{\mathbf{J}} F = L$. ☐

Corollary F.17 (Dual)**.** *Let* $\mathbf{C}$*,* $\mathbf{D}$ *and* $\mathbf{J}$ *be categories. If all colimits of shape* $\mathbf{J}$ *exist in* $\mathbf{D}$*, then all such colimits also exist in* $[\mathbf{C}, \mathbf{D}]$*, and they are computed pointwise.*[368]

[368] Uses Exercise F.15.

If you are craving some more diagram chasing or you want to get more familiar with natural transformations and functor categories, you can try doing the following exercises without using Theorem F.16 or Corollary F.17.[369]

[369] You can essentially reproduce the same proof with the shape $\mathbf{J}$ fixed.

OL Exercise F.18. Suppose $\mathbf{D}$ has a terminal object $\mathbf{1}$. Show the constant functor $\Delta(\mathbf{1}) : \mathbf{C} \rightsquigarrow \mathbf{D}$ is terminal in $[\mathbf{C}, \mathbf{D}]$. State and prove the dual statement.

OL Exercise F.19. Suppose $\mathbf{D}$ has all binary products and let $F, G \in [\mathbf{C}, \mathbf{D}]_0$. Show that sending $X \in \mathbf{C}_0$ to $FX \times GX$ and $f \in \mathbf{C}_1$ to $Ff \times Gf$ is a functor and it is the product of F and G in $[\mathbf{C}, \mathbf{D}]$. State and prove the dual statement.

OL Exercise F.20. Suppose $\mathbf{D}$ has all equalizers and let $\phi, \psi : F \Rightarrow G$ be two parallel natural transformations. For $X \in \mathbf{C}_0$, let (127) be the equalizer in $\mathbf{D}$. Find the action of E on morphisms that make E into a functor $\mathbf{C} \rightsquigarrow \mathbf{D}$ and e into a natural transformation $e : E \Rightarrow F$. Finally, show that e is the equalizer of ϕ and ψ in $[\mathbf{C}, \mathbf{D}]$. State and prove the dual statement.

$$\begin{tikzcd}
E(X) \arrow[r, "e_X"] & FX \arrow[r, shift left, "\phi_X"] \arrow[r, shift right, "\psi_X"'] & GX
\end{tikzcd} \tag{127}$$

OL **Exercise F.21.** Suppose $\mathbf{D}$ has all pullbacks and let $\phi : F \Rightarrow G \Leftarrow H : \psi$ be a cospan of natural transformation. For $X \in \mathbf{C}_0$, let (128) be the pullback in $\mathbf{D}$. Find the action of P on morphisms that makes P into a functor $\mathbf{C} \rightsquigarrow \mathbf{D}$ and $\ell : P \Rightarrow F$ and $r : P \Rightarrow G$ into natural transformation. Finally, show that P with ℓ and r is the pullback of that cospan. State and prove the dual statement.

$$\begin{tikzcd} P(X) \arrow[r, "r_X"] \arrow[d, "\ell_X"'] \arrow[dr, phantom, "\lrcorner", very near start] & HX \arrow[d, "\psi_X"] \\ FX \arrow[r, "\phi_X"'] & GX \end{tikzcd} \tag{128}$$

Example F.22 ((Co)limits in $\mathbf{DGph}$)**.**

Another simple application of viewing a category as a functor category is to look at the evaluation functors.

OL **Exercise F.23.** For any object $X \in \mathbf{C}_0$, show that *evaluation* at X is a functor $-X : [\mathbf{C}, \mathbf{D}] \rightsquigarrow \mathbf{D}$. It sends F to FX and ϕ to ϕ_X.

We leave you to check that the source and target functors $\mathsf{s}, \mathsf{t} : \mathbf{C}^{\rightarrow} \rightsquigarrow \mathbf{C}$ are naturally isomorphic to the functors evaluating at $A \in \mathbf{2}_0$ and $B \in \mathbf{2}_0$ respectively.[370] Evaluating at the single object in $\mathbf{B}G$ yields a forgetful functor $[\mathbf{B}G, \mathbf{Set}] \rightsquigarrow \mathbf{Set}$. It sends a group action to the underlying set and an equivariant map to the underlying function.

[370] This offers an alternative way to show s and t are functors in one go.

Using Exercise B.45, we can also conclude there is a functor $\mathsf{Ev} : \mathbf{C} \times [\mathbf{C}, \mathbf{D}] \rightsquigarrow \mathbf{D}$.[371] It sends (X, F) to $F(X)$ and $(f, \phi) : (X, F) \Rightarrow (Y, G)$ to $\phi_Y \circ F(f) = G(f) \circ \phi_X$.

[371] For a fixed $X \in \mathbf{C}_0$, we just saw $\mathsf{Ev}(X, -) = -X$ is a functor. For a fixed $F \in [\mathbf{C}, \mathbf{D}]_0$, $\mathsf{Ev}(-, F)$ is simply the functor F. The equation

$$\begin{aligned} \mathsf{Ev}(Y, \phi) \circ \mathsf{Ev}(f, F) &= \phi_Y \circ F(f) \\ &= G(f) \circ \phi_X \\ &= \mathsf{Ev}(f, G) \circ \mathsf{Ev}(X, \phi) \end{aligned}$$

holds by $\mathsf{NAT}(\phi, X, Y, f)$

We can now restate Theorem F.16 and Corollary F.17 by saying that when $\mathbf{D}$ has all (co)limits of shape $\mathbf{J}$, then Ev preserves (co)limits in its second component, i.e. for any $X \in \mathbf{C}_0$

$$\mathsf{Ev}(X, \lim_{\mathbf{J}} F) = \lim_{\mathbf{J}} \mathsf{Ev}(X, F-).$$

F.2 The 2–category Cat

It is now time to build intuition for the horizontal composition of natural transformations which will ultimately lead to the notion of a 2–category.

Definition F.24 (The left action of functors)**.** Let $F, F' : \mathbf{C} \rightsquigarrow \mathbf{D}$, $G : \mathbf{D} \rightsquigarrow \mathbf{D}'$ be functors and $\phi : F \Rightarrow F'$ a natural transformation as summarized in (129).[372]

[372] Using squiggly arrows for functors in diagrams is very non-standard, but I believe it helps remember what kind of objects we are dealing with. Moreover, since these diagrams are not commutative, it makes a good contrast with the plain arrow notation which was mostly used for commutative diagrams.

$$\begin{tikzcd} \mathbf{C} \arrow[r, squiggly, bend left, "F"{name=U}] \arrow[r, squiggly, bend right, "F'"'{name=L}] \arrow[from=U, to=L, phantom, "\Downarrow\phi"] & \mathbf{D} \arrow[r, squiggly, "G"] & \mathbf{D}' \end{tikzcd} \tag{129}$$

The functor G acts on ϕ by sending it to $G\phi := A \mapsto G(\phi(A)) : \mathbf{C}_0 \to \mathbf{D}'_1$. Showing that (130) commutes for any $f \in \mathrm{Hom}_{\mathbf{C}}(A, B)$ will imply that $G\phi$ is a natural transformation from $G \circ F$ to $G \circ F'$.

$$\begin{tikzcd} (G \circ F)(A) \arrow[r, "G\phi(A)"] \arrow[d, "(G\circ F)(f)"'] & (G \circ F')(A) \arrow[d, "(G\circ F')(f)"] \\ (G \circ F)(B) \arrow[r, "G\phi(B)"'] & (G \circ F')(B) \end{tikzcd} \tag{130}$$

Consider this diagram after removing all applications of G, by naturality of ϕ, it is commutative. Since functors preserve commutativity, the diagram still commutes after applying G, hence $G\phi : G \circ F \Rightarrow G \circ F'$ is indeed natural.[373]

[373] More concisely, we apply G to $\mathsf{NAT}(\phi, A, B, f)$ to obtain (130).

We leave you to check this constitutes a left action, namely, for any $G : \mathbf{D} \rightsquigarrow \mathbf{D}'$, $G' : \mathbf{D}' \rightsquigarrow \mathbf{D}''$ and $\phi : F \Rightarrow F'$,

$$\mathrm{id}_{\mathbf{D}}\phi = \phi \text{ and } G'(G\phi) = (G' \circ G)\phi.$$

Definition F.25 (The right action of functors)**.** Let $F, F' : \mathbf{C} \rightsquigarrow \mathbf{D}$, $H : \mathbf{C}' \rightsquigarrow \mathbf{C}$ be functors and $\phi : F \Rightarrow F'$ a natural transformation as summarized in (131).

$$\begin{tikzcd} \mathbf{C}' \arrow[r, squiggly, "H"] & \mathbf{C} \arrow[r, squiggly, bend left, "F", ""{name=U, below}] \arrow[r, squiggly, bend right, "F'"', ""{name=L, above}] & \mathbf{D} \arrow[Rightarrow, from=U, to=L, "\phi"] \end{tikzcd} \tag{131}$$

The functor H acts on ϕ by sending it to $\phi H := A \mapsto \phi(H(A)) : \mathbf{C}'_0 \to \mathbf{D}_1$. Showing that (132) commutes for any $f \in \mathrm{Hom}_{\mathbf{C}'}(A, B)$ will imply that ϕH is a natural transformation from $F \circ H$ to $F' \circ H$.

$$\begin{tikzcd} (F \circ H)(A) \arrow[r, "\phi H(A)"] \arrow[d, "(F\circ H)(f)"'] & (F' \circ H)(A) \arrow[d, "(F'\circ H)(f)"] \\ (F \circ H)(B) \arrow[r, "\phi H(B)"'] & (F' \circ H)(B) \end{tikzcd} \tag{132}$$

Commutativity of (132) follows by naturality of ϕ: change f in diagram (113) with the morphism $H(f) : H(A) \to H(B)$, i.e. (132) is $\mathsf{NAT}(\phi, HA, HB, Hf)$.

We leave you to check this constitutes a right action, namely, for any $H : \mathbf{C}' \rightsquigarrow \mathbf{C}$, $H' : \mathbf{C}'' \rightsquigarrow \mathbf{C}'$ and $\phi : F \Rightarrow F'$,

$$\phi\mathrm{id}_{\mathbf{C}} = \phi \text{ and } (\phi H)H' = \phi(H \circ H').$$

Proposition F.26. *The two actions commute, i.e. in the setting of* (133), $G(\phi H) = (G\phi)H$.[374]

[374] For this reason, we will drop all the parentheses from such expressions. We will also drop the $\circ$ for composition of functors. All in all, expect to find expressions like $G'G\phi HH'$ and infer the natural transformation $A \mapsto G'(G(\phi(H(H'(A)))))$.

$$\begin{tikzcd} \mathbf{C}' \arrow[r, squiggly, "H"] & \mathbf{C} \arrow[r, squiggly, bend left, "F", ""{name=U, below}] \arrow[r, squiggly, bend right, "F'"', ""{name=L, above}] & \mathbf{D} \arrow[r, squiggly, "G"] & \mathbf{D}' \arrow[Rightarrow, from=U, to=L, "\phi"] \end{tikzcd} \tag{133}$$

Proof. In both the L.H.S. and the R.H.S., an object $A \in \mathbf{C}'_0$ is sent to $G(\phi(H(A)))$. $\square$

OL Exercise F.27 (NOW!)**.** In the setting of (133), show that the assignments $F \mapsto G \circ F \circ H$ and $\phi \mapsto G\phi H$ make a functor $G(-)H : [\mathbf{C}, \mathbf{D}] \rightsquigarrow [\mathbf{C}', \mathbf{D}']$.

A very useful consequence is that for any commutative diagram in $[\mathbf{C}, \mathbf{D}]$, we can pre-compose and post-compose with any functors and still obtain a commutative diagram. For instance, if (134) commutes in $[\mathbf{C}, \mathbf{D}]$, then for any functors $H : \mathbf{C}' \rightsquigarrow \mathbf{C}$ and $G : \mathbf{D} \rightsquigarrow \mathbf{D}'$ (135) commutes.[375]

[375] We will often use this property by writing things like "apply $G(-)H$ to (134)" to use the commutativity of (135) in a proof.

$$\begin{tikzcd} X \arrow[r, "\eta"] \arrow[d, "\phi"'] & Y \arrow[d, "\phi'"] \\ X' \arrow[r, "\eta'"'] & Y' \end{tikzcd} \quad (134) \qquad \begin{tikzcd} G \circ X \circ H \arrow[r, "G\eta H"] \arrow[d, "G\phi H"'] & G \circ Y \circ H \arrow[d, "G\phi' H"] \\ G \circ X' \circ H \arrow[r, "G\eta' H"'] & G \circ Y' \circ H \end{tikzcd} \quad (135)$$

We will refer to these two actions as the **biaction** of functors on natural transformations and they will motivate the definition of another way to compose natural transformations.

Let **C**, **D** and **E** be categories, $H, H' : \mathbf{C} \rightsquigarrow \mathbf{D}$ and $G, G' : \mathbf{D} \rightsquigarrow \mathbf{E}$ be functors and $\phi : H \Rightarrow H'$ and $\eta : G \Rightarrow G'$ be natural transformations. This is summarized in (136).

$$\begin{tikzcd} \mathbf{C} \arrow[r, bend left, squiggly, "H"] \arrow[r, bend right, squiggly, "H'"'] \arrow[r, phantom, "\Downarrow\phi"] & \mathbf{D} \arrow[r, bend left, squiggly, "G"] \arrow[r, bend right, squiggly, "G'"'] \arrow[r, phantom, "\Downarrow\eta"] & \mathbf{E} \end{tikzcd} \quad (136)$$

The ultimate goal is to obtain a composition of ϕ and η that is a natural transformation $G \circ H \Rightarrow G' \circ H'$. Note that the biaction defined above yields four other natural transformations:

$$G\phi : G \circ H \Rightarrow G \circ H' \qquad \eta H : G \circ H \Rightarrow G' \circ H$$
$$G'\phi : G' \circ H \Rightarrow G' \circ H' \qquad \eta H' : G \circ H' \Rightarrow G' \circ H'.$$

All of the functors involved go from **C** to **E**, so all four natural transformations fit in diagram (137) that lives in the functor category $[\mathbf{C}, \mathbf{E}]$.

$$\begin{tikzcd} G \circ H \arrow[r, "G\phi"] \arrow[d, "\eta H"'] & G \circ H' \arrow[d, "\eta H'"] \\ G' \circ H \arrow[r, "G'\phi"'] & G' \circ H' \end{tikzcd} \quad (137)$$

At first glance, this suggests two different definitions for the horizontal composition, that is, the composition of the top path $(\eta H' \cdot G\phi)$ or the composition of the bottom path $(G'\phi \cdot \eta H)$. Surprisingly, both definitions coincide.

Lemma F.28. *Diagram* (137) *commutes, i.e.* $\eta H' \cdot G\phi = G'\phi \cdot \eta H$.[376]

[376] Similarly to NAT, we will refer to the commutativity of (137) with $\mathsf{HOR}(\phi, \eta)$. We use HOR because this lemma is crucial in the definition of HORizontal composition.

Proof. Fix an object $A \in \mathbf{C}_0$. Under $\eta H' \cdot G\phi$, it is sent to $\eta(H'(A)) \circ G(\phi(A))$ and under $G'\phi \cdot \eta H$, it is sent to $G'(\phi(A)) \circ \eta(H(A))$. Thus, the proposition is equivalent to saying diagram (138) is commutative (in **E**) for all $A \in \mathbf{C}_0$.

$$\begin{tikzcd} (G \circ H)(A) \arrow[r, "G(\phi(A))"] \arrow[d, "\eta(H(A))"'] & (G \circ H')(A) \arrow[d, "\eta(H'(A))"] \\ (G' \circ H)(A) \arrow[r, "G'(\phi(A))"'] & (G' \circ H')(A) \end{tikzcd} \quad (138)$$

This follows from $\mathsf{NAT}(\eta, HA, H'A, \phi(A))$. $\square$

Definition F.29 (Horizontal composition)**.** In the setting described in (136), we define the **horizontal composition** of η and ϕ by $\eta \diamond \phi = \eta H' \cdot G\phi = G'\phi \cdot \eta H$.[377]

[377] The $\diamond$ notation is not standard but there are no widespread symbol denoting horizontal composition. I have mostly seen $*$ or plain juxtaposition. Hopefully, you will encounter papers/books clear enough that you can typecheck to find what composition is being used.

One crucial point we have made in earlier chapters is that a notion of composition must satisfy associativity and have identities. We will show the former right after you show the latter.

OL Exercise F.30. Let $H : \mathbf{C}' \rightsquigarrow \mathbf{C}$, $F, F' : \mathbf{C} \rightsquigarrow \mathbf{D}$ and $G : \mathbf{D} \rightsquigarrow \mathbf{D}'$ be functors and $\phi : F \Rightarrow F'$ be a natural transformation (as in (133)). Show that $\phi \diamond \mathbb{1}_H = \phi H$ and $\mathbb{1}_G \diamond \phi = G\phi$. Infer that $\mathbb{1}_{\mathrm{id}_\mathbf{C}}$ is the identity at $\mathbf{C}$ for $\diamond$.

Proposition F.31. *In the setting of* (139)*,* $\psi \diamond (\eta \diamond \phi) = (\psi \diamond \eta) \diamond \phi$*.*

$$\begin{tikzcd}
\mathbf{C} \arrow[r, bend left, squiggly, "H"] \arrow[r, bend right, squiggly, "H'"'] \arrow[r, phantom, "\Downarrow\phi"] & \mathbf{D} \arrow[r, bend left, squiggly, "G"] \arrow[r, bend right, squiggly, "G'"'] \arrow[r, phantom, "\Downarrow\eta"] & \mathbf{E} \arrow[r, bend left, squiggly, "K"] \arrow[r, bend right, squiggly, "K'"'] \arrow[r, phantom, "\Downarrow\psi"] & \mathbf{F}
\end{tikzcd} \tag{139}$$

Proof. Similarly to how we constructed diagram (137) previously, we can use the biaction of functors and composition of functors to obtain the following diagram in $[\mathbf{C}, \mathbf{F}]$.[378]

[378] All $\circ$'s are left out for simplicity.

$$\begin{tikzcd}
& K'GH \arrow[rr, "K'G\phi"] \arrow[dd, "K'\eta H"] & & K'GH' \arrow[dd, "K'\eta H'"] \\
KGH \arrow[ur, "\psi GH"] \arrow[rr, "KG\phi"] \arrow[dd, "K\eta H"'] & & KGH' \arrow[ur, "\psi GH'"'] \arrow[dd, "K\eta H'"] & \\
& K'G'H \arrow[rr, "K'G'\phi"] & & K'G'H' \\
KG'H \arrow[ur, "\psi G'H"] \arrow[rr, "KG'\phi"'] & & KG'H' \arrow[ur, "\psi G'H'"'] &
\end{tikzcd} \tag{140}$$

Here is how each face commutes.

Top: $\mathsf{HOR}(\psi, G\eta)$
Bottom: $\mathsf{HOR}(\psi, G'\eta)$
Left: $\mathsf{HOR}(\psi, \eta H)$
Right: $\mathsf{HOR}(\psi, \eta H')$
Front: $\mathsf{HOR}(K\eta, \phi)$
Back: $\mathsf{HOR}(K'\eta, \phi)$

As detailed in the margin, this commutes because each face of the cube corresponds to a variant of diagram (137) (with some substitutions and application of a functor) and combining commutative diagrams yields commutative diagrams. Then, it follows that $\diamond$ is associative because[379] $\psi \diamond (\eta \diamond \phi)$ is the diagonal of the front face followed by the bottom right arrow, and $(\psi \diamond \eta) \diamond \phi$ is the top front arrow followed by the diagonal of the right face. $\square$

[379] We could have drawn only the front and right face, but the cube is cooler.

There is one last thing to conclude that **Cat** is a 2–category, namely, that the vertical and horizontal compositions interact nicely.

Proposition F.32 (Interchange identity)**.** *In the setting of* (142)*, the* ***interchange identity*** *holds:*

$$(\eta' \cdot \eta) \diamond (\phi' \cdot \phi) = (\eta' \diamond \phi') \cdot (\eta \diamond \phi). \tag{141}$$

It is in the drawing of (142) that the intuition behind the terms vertical and horizontal is taken.

$$\begin{tikzcd}
\mathbf{C} \arrow[r, bend left=50, squiggly, "H"] \arrow[r, squiggly, "H'" description] \arrow[r, bend right=50, squiggly, "H''"'] \arrow[r, phantom, shift left=3, "\Downarrow\phi"] \arrow[r, phantom, shift right=3, "\Downarrow\phi'"] & \mathbf{D} \arrow[r, bend left=50, squiggly, "G"] \arrow[r, squiggly, "G'" description] \arrow[r, bend right=50, squiggly, "G''"'] \arrow[r, phantom, shift left=3, "\Downarrow\eta"] \arrow[r, phantom, shift right=3, "\Downarrow\eta'"] & \mathbf{E}
\end{tikzcd} \tag{142}$$

Proof. Akin to the other proofs, this is a matter of combining the right diagrams. After combining the diagrams in $[\mathbf{C},\mathbf{E}]$ corresponding to $\eta \diamond \phi$ and $\eta' \diamond \phi'$, it is easy to see that the R.H.S. of (141) is the morphism going from $G \circ H$ to $G'' \circ H''$ in (143).

$$\begin{tikzcd}
G \circ H \arrow[r, "G\phi"] \arrow[d, "\eta H"'] & G \circ H' \arrow[d, "\eta H'"] & \\
G' \circ H \arrow[r, "G'\phi"'] & G' \circ H' \arrow[r, "G'\phi'"] \arrow[d, "\eta' H'"'] & G' \circ H'' \arrow[d, "\eta' H''"] \\
& G'' \circ H' \arrow[r, "G''\phi'"'] & G'' \circ H''
\end{tikzcd} \tag{143}$$

Moreover, the diagram corresponding to the L.H.S. can be factored with the following equations (they follow from Exercise F.27) yielding (144).

$$\begin{tikzcd}
G \circ H \arrow[r, "G\phi"] \arrow[d, "\eta H"'] & G \circ H' \arrow[r, "G\phi'"] & G \circ H'' \arrow[d, "\eta H''"] \\
G' \circ H \arrow[d, "\eta' H"'] & & G' \circ H'' \arrow[d, "\eta' H''"] \\
G'' \circ H \arrow[r, "G''\phi"'] & G'' \circ H' \arrow[r, "G''\phi'"'] & G'' \circ H''
\end{tikzcd} \tag{144}$$

$$(\eta' \cdot \eta)H = \eta' H \cdot \eta H \qquad\qquad (\eta' \cdot \eta)H'' = \eta' H'' \cdot \eta H''$$
$$G(\phi' \cdot \phi) = G\phi' \cdot G\phi \qquad\qquad G''(\phi' \cdot \phi) = G''\phi' \cdot G''\phi$$

Combining (143) and (144), we obtain (145) from which the interchange identity readily follows.[380]

$$\begin{tikzcd}
G \circ H \arrow[r, "G\phi"] \arrow[d, "\eta H"'] & G \circ H' \arrow[r, "G\phi'"] \arrow[d, "\eta H'"] & G \circ H'' \arrow[d, "\eta H''"] \\
G' \circ H \arrow[r, "G'\phi"'] \arrow[d, "\eta' H"'] & G' \circ H' \arrow[r, "G'\phi'"] \arrow[d, "\eta' H'"'] & G' \circ H'' \arrow[d, "\eta' H''"] \\
G'' \circ H \arrow[r, "G''\phi"'] & G'' \circ H' \arrow[r, "G''\phi'"'] & G'' \circ H''
\end{tikzcd} \tag{145}$$

□

[380] The top right and bottom left square commute by $\mathsf{HOR}(\eta,\phi')$ and $\mathsf{HOR}(\eta',\phi)$ respectively. This implies all of (145) commutes and we have seen that the path from $G \circ H$ to $G'' \circ H''$ can be seen as the R.H.S. of (141) by looking at (143) or the L.H.S. by looking at (144). Thus, we infer the satisfaction of (141).

All of the structure we have added on top of the category **Cat** can be abstracted away by saying that it is 2–category.

Definition F.33 (Strict 2–cateory)**.** A **strict** 2**–category** consists of

- a category $\mathbf{C}$,
- for every $A, B \in \mathbf{C}_0$ a category $\mathbf{C}(A,B)$ with $\mathrm{Hom}_{\mathbf{C}}(A,B)$ as its objects and morphisms are called 2**–morphisms** (composition is denoted $\cdot$ and identities $\mathbb{1}$),
- a category with $\mathbf{C}_0$ as its objects, where the morphisms are pairs of parallel morphisms of $\mathbf{C}$ along with a 2–morphism between them. A morphism in this category is also called a 2**–cell**. The identity 2–cell at $A \in \mathbf{C}_0$ is the pair $(\mathrm{id}_A, \mathrm{id}_A)$ and the 2–morphism $\mathbb{1}_{\mathrm{id}_A}$ and composition of 2–cells is denoted $\diamond$),

such that the interchange identity (141) holds.[381]

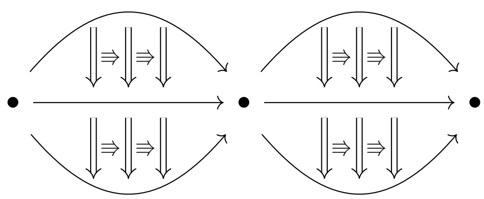
[381] The interchange identity does not come out of nowhere, it is equivalent to the composition $\circ$ being a functor $\mathbf{C}(B,C) \times \mathbf{C}(A,B) \rightsquigarrow \mathbf{C}(A,C)$ that acts on 2–morphisms by $\diamond$ for every $A, B, C \in \mathbf{C}_0$. We leave you to show this in the special case of the 2–category of categories in Exercise F.34.

Exercise F.34 (NOW!)**.** Show that there is a functor $[\mathbf{D},\mathbf{E}] \times [\mathbf{C},\mathbf{D}] \rightsquigarrow [\mathbf{C},\mathbf{E}]$ whose action on objects is $(F,G) \mapsto F \circ G$.

Digression on Higher/Enriched Categories

This book is not the place to further study 2–categories, but we can say a few interesting things about them. There are notions of morphisms between 2–categories (called 2–functors) and morphisms between them (called 2–natural transformations). The latter can be composed in three different ways (analog to vertical and horizontal composition for 2–morphisms) and all possible compositions interact well together. In particular,[382] there is a unique 2–natural transformation that is the composition of all 2–natural transformations in (146) (there are multiple ways to obtain it, depending on what compositions you do in what order, but as in the interchange identity, we require them to lead to the same 2–natural transformation).

[382] There are several so-called coherence axioms that describe how all compositions interact, but we state only one of them.



(146)

The category of 2–categories with 2–functors and 2–natural transformations is now an instance of a 3–category. The field of *higher category theory* studies the generalizations of this to n–categories for any n (even $n = \infty$!). However, most of higher category theory drops the *strict* part of our definition of 2–category because this condition is too strong. Very briefly, they allow the properties of composition, namely associativity, identities and interchange, to hold up to isomorphisms.

There is a relatively simple way to define strict n–categories using *enriched category theory*.[383] The definition of a locally small category can be seen as entirely taking place in the category **Set**. From this point of view, a locally small category is a collection $\mathbf{C}_0$ of objects equipped with

[383] I hope you can indulge this continued digression. While higher and enriched category theory are not as indispensible as basic category theory, they are quite powerful. We will not see how in this book, but I think these two little teasers might inspire some readers to find out by themselves.

- a set $\mathrm{Hom}_{\mathbf{C}}(A, B) \in \mathbf{Set}_0$ for every $A, B \in \mathbf{C}_0$,
- a function $\circ_{A,B,C} \in \mathrm{Hom}_{\mathbf{Set}}\left(\mathrm{Hom}_{\mathbf{C}}(B, C) \times \mathrm{Hom}_{\mathbf{C}}(A, B), \mathrm{Hom}_{\mathbf{C}}(A, C)\right)$ for every $A, B, C \in \mathbf{C}_0$,
- and a function $\mathrm{id}_A \in \mathrm{Hom}_{\mathbf{Set}}\left(\mathbf{1}, \mathrm{Hom}_{\mathbf{C}}(A, A)\right)$,

with conditions that can be stated as commutative diagrams in **Set**. Commutativity of (147) and (148) means that the identity morphisms are neutral with respect to composition and commutativity of (149) means composition is associative.

$$\begin{tikzcd}
\mathrm{Hom}_{\mathbf{C}}(B, C) \times \mathbf{1} \arrow[r, "\mathrm{id} \times \mathrm{id}_B"] \arrow[dr, "\pi_{\mathrm{Hom}_{\mathbf{C}}(B,C)}"'] & \mathrm{Hom}_{\mathbf{C}}(B, C) \times \mathrm{Hom}_{\mathbf{C}}(B, B) \arrow[d, "\circ_{B,B,C}"] \\
& \mathrm{Hom}_{\mathbf{C}}(B, C)
\end{tikzcd} \tag{147}$$

$$\begin{tikzcd}
\mathrm{Hom}_{\mathbf{C}}(B,B) \times \mathrm{Hom}_{\mathbf{C}}(A,B) \arrow[d, "\circ_{A,B,B}"'] & \mathbf{1} \times \mathrm{Hom}_{\mathbf{C}}(A,B) \arrow[l, "\mathrm{id}_B \times \mathrm{id}"'] \arrow[dl, "\pi_{\mathrm{Hom}_{\mathbf{C}}(A,B)}"] \\
\mathrm{Hom}_{\mathbf{C}}(A,B) &
\end{tikzcd} \tag{148}$$

$$\begin{tikzcd}
\mathrm{Hom}_{\mathbf{C}}(C,D) \times \mathrm{Hom}_{\mathbf{C}}(B,C) \times \mathrm{Hom}_{\mathbf{C}}(A,B) \arrow[r, "\circ_{B,C,D} \times \mathrm{id}"] \arrow[d, "\mathrm{id} \times \circ_{A,B,C}"'] & \mathrm{Hom}_{\mathbf{C}}(B,D) \times \mathrm{Hom}_{\mathbf{C}}(A,B) \arrow[d, "\circ_{A,B,D}"] \\
\mathrm{Hom}_{\mathbf{C}}(C,D) \times \mathrm{Hom}_{\mathbf{C}}(A,C) \arrow[r, "\circ_{A,C,D}"'] & \mathrm{Hom}_{\mathbf{C}}(A,D)
\end{tikzcd} \tag{149}$$

It turns out we can abstract the properties of $\mathbf{1}$ and $\times$ that ensure we can do category theory: we say that $(\mathbf{Set}, \times, \mathbf{1})$ is a **monoidal category**.[384] Now, **enriched category theory** is done by replacing **Set** with another category that has a monoidal structure.

[384] The specific properties are not too relevant for us right now, but know that $\times$ and $\mathbf{1}$ are called the **tensor** and **unit** of the monoidal category.

Examples F.35. 1. The category $\mathbf{1}$ is a monoidal category with the tensor and unit being trivial (there is only one object, so there is no choice). A category enriched in $\mathbf{1}$ is simply a collection $\mathbf{C}_0$ because there is no choice when defining $\mathrm{Hom}_{\mathbf{C}}(A,B) \in \mathbf{1}_0$, $\circ_{A,B,C} \in \mathbf{1}_1$ and $\mathrm{id}_A \in \mathbf{1}_1$.

2. Recall that categories can be seen as generalizations of monoids where elements have a source and target, and you can only multiply elements when they are composable. If we started from rings instead, we would have to say how morphisms can be added. For instance in **Ab**, given two parallel morphisms $f, f' : A \to B$, we can add them pointwise $(f + f')(a) = f(a) + f'(a)$.[385] This operation makes $\mathrm{Hom}_{\mathbf{Ab}}(A,B)$ an abelian group. Moreover, you can check that, just as multiplication commutes with addition in a ring, $g \circ (f + f') = (g \circ f) + (g \circ f')$ and $(f + f') \circ h = (f \circ h) + (f' \circ h)$.[386] This is equivalent to saying
$$\circ_{A,B,C} : \mathrm{Hom}_{\mathbf{Ab}}(B,C) \times \mathrm{Hom}_{\mathbf{Ab}}(A,B) \to \mathrm{Hom}_{\mathbf{Ab}}(A,C)$$
is a bilinear map, or equivalently,
$$\circ_{A,B,C} \in \mathrm{Hom}_{\mathbf{Ab}}\left(\mathrm{Hom}_{\mathbf{Ab}}(B,C) \otimes \mathrm{Hom}_{\mathbf{Ab}}(A,B), \mathrm{Hom}_{\mathbf{Ab}}(A,C)\right),$$
where $\otimes$ denotes the tensor product of abelian groups. Noting that $(\mathbf{Ab}, \otimes, \mathbb{Z})$ is a monoidal category, we simply say that **Ab** is enriched in **Ab**. You can check that $\mathbf{Vect}_k$ is also **Ab**–enriched.[387] Moreover, just like a monoid is the same thing as a single object category, a ring is the same thing as a single object **Ab**–enriched category.

[385] The group operation in B is denoted by $+$ because it is commutative.

[386] However, in general,
$$(g + g') \circ (f + f') \neq (g \circ f) + (g' \circ f').$$

[387] You might encounter abelian categories in the wild, these are a special kind of **Ab**–enriched categories.

3. The category **Cat** of small categories is monoidal with the tensor being $\times$ and the unit being $\mathbf{1}$. A category enriched in **Cat** is a strict 2–category. For instance, the 2–category of categories is a collection $\mathbf{Cat}_0$ of objects, a category $\mathbf{Cat}(\mathbf{C}, \mathbf{D}) = [\mathbf{C}, \mathbf{D}]$ for every $\mathbf{C}, \mathbf{D} \in \mathbf{Cat}_0$, a functor $\mathrm{id}_{\mathbf{C}} : \mathbf{1} \rightsquigarrow [\mathbf{C}, \mathbf{C}]$ that picks the identity functor and, as you will show in Exercise F.34, a morphism
$$\circ_{\mathbf{C},\mathbf{D},\mathbf{E}} \in \mathrm{Hom}_{\mathbf{Cat}}([\mathbf{D}, \mathbf{E}] \times [\mathbf{C}, \mathbf{D}], [\mathbf{C}, \mathbf{E}]).$$

The diagrams corresponding to (147), (148), and (149) (now they live in **Cat**) commute by results we have shown in this chapter.

4. Generalizing the previous item, a strict n–category is a category enriched in the category of strict $(n-1)$–categories.

5.

6. The posetal category $([0,\infty], \geq)$ is monoidal with the tensor being $+$ (addition) and unit being 0.[388] A category enriched in $[0,\infty]$ is
   - a collection of objects X,
   - for every $x, y \in X$ an element $X(x,y) \in [0,\infty]$, and
   - for every $x, y, z \in X$, an element of $\mathrm{Hom}_{[0,\infty]}(X(y,z) + X(x,y), X(x,z))$.

   We can see the second point as a function $X \times X \to [0,\infty]$, and the third point says that $X(x,z) \leq X(x,y) + X(y,z)$.[389] This looks like a triangle inequality, and in fact all of X looks like a metric space, but where the distance can be infinite, the distance is not symmetric, and two distinct elements can be at distance 0.[390] A $[0,\infty]$–enriched category is also called a Lawvere metric space. If you are enjoying this introduction to enriched category theory, you can try to define *enriched functors*. You should find that for $[0,\infty]$, an enriched functor is a nonexpansive map between Lawvere metric spaces.[391]

[388] We define addition with ∞ in the intuitive way, $x + \infty = \infty + x = \infty$ for all $x \in [0,\infty]$.

[389] Recall there is an element in $\mathrm{Hom}_{[0,\infty]}(r,s)$ if and only if $r \geq s$.

[390] The fact that $X(x,x) = 0$ is witnessed by the identity morphism in $\mathrm{Hom}_{[0,\infty]}(0, X(x,x))$.

[391] This is one reason to define **Met** as we did.

F.3 Equivalences

Up to now, we supposedly have been doing everything up to isomorphism. However, in a 2–category and in particular in **Cat**, this can be too restrictive. Fortunately, the new "dimension" of natural transformations allows us to define a relaxed version of equality between categories called equivalence.

Recall that an isomorphism of categories is an isomorphism in the category **Cat**, namely, a functor $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ with an inverse $G : \mathbf{D} \rightsquigarrow \mathbf{C}$ such that $F \circ G = \mathrm{id}_{\mathbf{D}}$ and $G \circ F = \mathrm{id}_{\mathbf{C}}$. As is typical in mathematics, one cannot distinguish between isomorphic categories as they only differ in notations and terminology.[392]

In many situations, we will describe an isomorphism between **C** and **D** by identifying the objects and morphisms in **C** with the objects and morphisms in **D**. That is, the functors are implicit in the discussion. For instance, in Example E.44 we argued that $\mathbf{1}/\mathbf{Set}$ and $\mathbf{Set}_*$ are the same category. We really meant that they are isomorphic.[393] Only in rare cases (see Example F.36.5 below) will we explicitly define the functor and its inverse.

[392] For example, the monoid isomorphism $\mathbb{N} \cong \{\mathtt{a}\}^*$ offers two ways to talk about the same mathematical object. In particular, it identifies 1 with a, 2 with aa, 3 with aaa, etc.

[393] The details of the construction of the isomorphisms are left to you.

Examples F.36. Here are other examples of isomorphic categories that we have already seen and a couple of new ones.

1. It was already shown in Example F.13 (the details were implicit) that for a group G, the category $[\mathbf{B}G, \mathbf{Set}]$ is isomorphic to the category of G–sets with G–equivariant maps as morphisms.

Another example for readers who know a bit of advanced algebra. Let k be a field and G a finite group, the categories of $k[G]$–modules ($k[G]$ is the group ring of k over G) and of k–linear representations of G are isomorphic.

2. In Example F.14, three other isomorphisms were implicitly given:

$$[\mathbf{1}, \mathbf{C}] \cong \mathbf{C} \qquad [\mathbf{1}+\mathbf{1}, \mathbf{C}] \cong \mathbf{C} \times \mathbf{C} \qquad [\mathbf{2}, \mathbf{C}] \cong \mathbf{C}^{\rightarrow}.$$

3. The category **Rel** of sets with relations is isomorphic to $\mathbf{Rel}^{\mathrm{op}}$.[394] The functor $\mathbf{Rel} \rightsquigarrow \mathbf{Rel}^{\mathrm{op}}$ is the identity on objects and sends a relation $R \subseteq X \times Y$ to the opposite relation $\reflectbox{$R$} \subseteq Y \times X$ (which is a morphism $X \to Y$ in $\mathbf{Rel}^{\mathrm{op}}$) defined by $(y, x) \in \reflectbox{$R$} \Leftrightarrow (x, y) \in R$. The inverse is defined similarly.

[394] An arbitrary category **C** is not always isomorphic to its opposite. While the opposite functors $(-)^{\mathrm{op}}_{\mathbf{C}} : \mathbf{C} \rightsquigarrow \mathbf{C}^{\mathrm{op}}$ and $(-)^{\mathrm{op}}_{\mathbf{C}^{\mathrm{op}}} : \mathbf{C}^{\mathrm{op}} \rightsquigarrow \mathbf{C}$ are inverses of each other, they are contravariant functors, i.e. they are not morphisms in **Cat**.

4. Let **C** and **D** be categories the functor $\mathsf{swap} : \mathbf{C} \times \mathbf{D} \rightsquigarrow \mathbf{D} \times \mathbf{C}$ sends (A, B) to (B, A) and (f, g) to (g, f). It is easy to check that swap is a functor with inverse $\mathsf{swap}^{-1} : \mathbf{D} \times \mathbf{C} \rightsquigarrow \mathbf{C} \times \mathbf{D}$ defined in the obvious way.

5. Given three categories **C**, **D** and **E**, there is an isomorphism[395]

$$[\mathbf{C} \times \mathbf{D}, \mathbf{E}] \cong [\mathbf{C}, [\mathbf{D}, \mathbf{E}]].$$

[395] You might recognize a similarity with exponentials which rely on an isomorphism $\mathrm{Hom}_{\mathbf{C}}(B \times X, A) \cong \mathrm{Hom}_{\mathbf{C}}(B, A^X)$. The example here is more than an instance of exponentials of categories because the isomorphism is not only as sets but as categories.

The construction of the isomorphism follows the intuition of currying and uncurrying of functions, so the definitions are straightforward. Still, you will see that verifying the straightforward defintions are well-typed is cumbersome (but simple) because there are several levels of functors and natural transformations.

Let $F : \mathbf{C} \times \mathbf{D} \rightsquigarrow \mathbf{E}$, the currying of F is $\Lambda F : \mathbf{C} \rightsquigarrow [\mathbf{D}, \mathbf{E}]$ defined as follows. For $X \in \mathbf{C}_0$, the functor $\Lambda F(X)$ sends $Y \in \mathbf{D}_0$ to $F(X, Y)$ and $g \in \mathbf{D}_1$ to $F(\mathrm{id}_X, g)$. We showed in Exercise B.44 that $\Lambda F(X) = F(X, -)$ is a functor. For $f \in \mathrm{Hom}_{\mathbf{C}}(X, X')$, we define the natural transformation $\Lambda F(f) : F(X, -) \Rightarrow F(X', -)$ by

$$\Lambda F(f)_Y = F(f, \mathrm{id}_Y) : F(X, Y) \to F(X', Y).$$

$$\begin{tikzcd} F(X,Y) \arrow[r, "F(\mathrm{id}_X, g)"] \arrow[d, "F(f, \mathrm{id}_Y)"'] & F(X,Y') \arrow[d, "F(f, \mathrm{id}_{Y'})"] \\ F(X',Y) \arrow[r, "F(\mathrm{id}_{X'}, g)"'] & F(X',Y') \end{tikzcd} \tag{150}$$

The naturality square (150) is commutative because, by functoriality of F, the top and bottom path are equal to $F(f, g)$. We also have to show ΛF is a functor, namely $\Lambda F(\mathrm{id}_X) = \mathbb{1}_{F(X,-)}$ and $\Lambda F(f \circ f') = \Lambda F(f) \cdot \Lambda F(f')$. We can verify this componentwise using functoriality of F.

$$\Lambda F(\mathrm{id}_X)_Y = F(\mathrm{id}_X, \mathrm{id}_Y) = \mathrm{id}_{F(X,Y)}$$
$$\Lambda F(f \circ f')_Y = F(f \circ f', \mathrm{id}_Y) = F(f, \mathrm{id}_Y) \circ F(f', \mathrm{id}_Y) = \Lambda F(f)_Y \circ \Lambda F(f')_Y.$$

It remains to define $\Lambda -$ on morphisms. Given a natural transformation $\phi : F \Rightarrow F'$, we define $\Lambda\phi : \Lambda F \Rightarrow \Lambda F'$ at component $X \in \mathbf{C}_0$ by the natural transformation:

$$\Lambda\phi(X) = \phi_{X,-} : F(X, -) \Rightarrow F'(X, -).$$

We showed in Exercise F.7 that $\phi_{X,-}$ is natural. Finally, we can check that $\Lambda -$ is a functor with the following derivations.[396]

[396] The second equation on the second line can be verified componentwise, i.e. for every $Y \in \mathbf{D}_0$

$$(\phi \cdot \eta)_{X,Y} = \phi_{X,Y} \circ \eta_{X,Y} = (\phi_{X,-} \cdot \eta_{X,-})_Y.$$

$$\Lambda \mathbb{1}_F(X) = (\mathbb{1}_F)_{X,-} = \mathbb{1}_{F(X,-)}$$
$$\Lambda(\phi \cdot \eta)(X) = (\phi \cdot \eta)_{X,-} = \phi_{X,-} \cdot \eta_{X,-} = \Lambda\phi \cdot \Lambda\eta$$

Conversely, let $F : \mathbf{C} \rightsquigarrow [\mathbf{D}, \mathbf{E}]$, the uncurrying of F is $\Lambda^{-1}F : \mathbf{C} \times \mathbf{D} \rightsquigarrow \mathbf{E}$ defined as follows. We use Exercise B.45 to define $\Lambda^{-1}F$ componentwise. Fixing $X \in \mathbf{C}_0$, we know that $F(X)$ is a functor, so we set $\Lambda^{-1}F(X, -) = F(X)$. Fixing $Y \in \mathbf{D}_0$, we define $\Lambda^{-1}F(-, Y)$ on objects by sending $X \in \mathbf{C}_0$ to $F(X)(Y)$ and $f \in \mathbf{C}_1$ to $F(f)_Y$.[397] To show $\Lambda^{-1}F(-, Y)$ is a functor, we use the functoriality of F as follows.

[397] As a sanity check, if $f : X \to X'$, $F(f) : F(X) \Rightarrow F(X')$, thus the component at Y is $F(f)_Y : F(X)(Y) \to F(X')(Y)$ as desired.

$$\Lambda^{-1}F(\mathrm{id}_X, Y) = F(\mathrm{id}_X)_Y = \mathbb{1}_{F(X)_Y} = \mathrm{id}_{F(X)(Y)}$$
$$\Lambda^{-1}F(f \circ f', Y) = F(f \circ f')_Y = (F(f) \cdot F(f'))_Y = F(f)_Y \circ F(f')_Y.$$

Now, for every $f : X \to X'$ and $g : Y \to Y'$, the naturality of $F(f)$ implies the square in (151) commutes. This means we can let $\Lambda^{-1}F(f, g)$ be the diagonal, i.e.

$$\Lambda^{-1}F(f, g) := \Lambda^{-1}F(X', g) \circ \Lambda^{-1}F(f, Y) = \Lambda^{-1}F(f, Y') \circ \Lambda^{-1}F(X, g),$$

$$\begin{tikzcd} F(X)(Y) \arrow[r, "F(X)(g)"] \arrow[d, "F(f)_Y"'] & F(X)(Y') \arrow[d, "F(f)_{Y'}"] \\ F(X')(Y) \arrow[r, "F(X')(g)"'] & F(X')(Y') \end{tikzcd} \tag{151}$$

and conclude by Exercise B.45 that $\Lambda^{-1}F : \mathbf{C} \times \mathbf{D} \rightsquigarrow \mathbf{E}$ is a functor.

Given a natural transformation $\phi : F \Rightarrow F'$, we define $\Lambda^{-1}\phi : \Lambda^{-1}F \Rightarrow \Lambda^{-1}F'$ by $(\Lambda^{-1}\phi)_{X,Y} := (\phi_X)_Y$. By Exercise F.7, it is enough to show naturality in one component at a time. Fix $X \in \mathbf{C}_0$, by hypothesis (ϕ_X is a morphism in $[\mathbf{D}, \mathbf{E}]$), $\phi_X : F(X) \Rightarrow F'(X)$ is natural in Y. Fix $Y \in \mathbf{D}_0$, we need to show the following square commutes.

$$\begin{tikzcd} F(X)(Y) \arrow[r, "\Lambda^{-1}F(f,Y)"] \arrow[d, "(\phi_X)_Y"'] & F(X')(Y) \arrow[d, "(\phi_{X'})_Y"] \\ F'(X)(Y) \arrow[r, "\Lambda^{-1}F'(f,Y)"'] & F'(X')(Y) \end{tikzcd} \tag{152}$$

Recalling that $\Lambda^{-1}F(f, Y) = F(f)_Y$ and $\Lambda^{-1}F'(f, Y) = F'(f)_Y$, we recognize this square as $\mathsf{NAT}(\phi, X, X', f)$ evaluated at Y. Finally, we can check that $\Lambda^{-1}-$ is a functor with the following derivations.

$$(\Lambda^{-1}\mathbb{1}_F)_{X,Y} = ((\mathbb{1}_F)_X)_Y = \mathrm{id}_{F(X)(Y)} = (\mathbb{1}_{\Lambda^{-1}F})_{X,Y}$$
$$(\Lambda^{-1}\phi \cdot \eta)_{X,Y} = ((\phi \cdot \eta)_X)_Y = (\phi_X)_Y \circ (\eta_X)_Y = (\Lambda^{-1}\phi)_{X,Y} \cdot (\Lambda^{-1}\eta)_{X,Y}$$

The last step (I promise) of this proof is to show that $\Lambda-$ and $\Lambda^{-1}-$ are inverses of each other. The mindless computations below suffice.

$$\Lambda\Lambda^{-1}F(X)(Y) = \Lambda^{-1}F(X, Y) = F(X)(Y)$$
$$\Lambda\Lambda^{-1}F(f)_Y = \Lambda^{-1}F(f, Y) = F(f)_Y$$

$$\Lambda^{-1}\Lambda F(X, Y) = \Lambda F(X)(Y) = F(X, Y)$$
$$\Lambda^{-1}\Lambda F(f, g) = \Lambda F(X')(g) \circ \Lambda F(f)_Y = F(\mathrm{id}_{X'}, g) \circ F(f, \mathrm{id}_Y) = F(f, g)$$

Of course, the list above is not exhaustive, but it is time to introduce equivalences. Instead of requiring the round trips between $\mathbf{C}$ and $\mathbf{D}$ to be the identities, we merely require they are naturaly isomorphic to the identities.

Definition F.37 (Equivalence)**.** A functor $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ is an **equivalence** of categories if there exists a functor $G : \mathbf{D} \rightsquigarrow \mathbf{C}$ such that $F \circ G \cong \mathrm{id}_{\mathbf{D}}$ and $G \circ F \cong \mathrm{id}_{\mathbf{C}}$.[398] This is clearly symmetric, so we say two categories $\mathbf{C}$ and $\mathbf{D}$ are **equivalent**, denoted $\mathbf{C} \simeq \mathbf{D}$, if there is an equivalence between them. Moreover, we say that G is a **quasi-inverse** of F and vice-versa.

[398] Recall that $\cong$ between functors stands for natural isomorphisms.

This is certainly weaker than an isomorphism of categories, but it is still quite strong. In order to gain more intuition on how equivalences equate two categories, let us observe what properties this forces on the functor F. For all $f \in \mathrm{Hom}_{\mathbf{C}}(A, B)$, the following square commutes where ϕ_A and ϕ_B are isomorphisms.[399]

[399] Naturality of ϕ only gives us $GF(f) \circ \phi_A = \phi_B \circ f$, but by composing with ϕ_A^{-1} or ϕ_B^{-1}, we obtain the commutativity of all of (153). In particular, we have $GF(f) = \phi_B \circ f \circ \phi_A^{-1}$.

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, shift left, "\phi_A"] & B \arrow[d, shift right, "\phi_B"'] \\ GF(A) \arrow[r, "GF(f)"'] \arrow[u, shift left, "\phi_A^{-1}"] & GF(B) \arrow[u, shift right, "\phi_B^{-1}"'] \end{tikzcd} \tag{153}$$

This implies that the map $f \mapsto GF(f) : \mathrm{Hom}_{\mathbf{C}}(A, B) \to \mathrm{Hom}_{\mathbf{C}}(GF(A), GF(B))$ is a bijection. Indeed, pre-composition by ϕ_A^{-1} and post-composition by ϕ_B are both bijections,[400] so

$$f \mapsto \phi_B \circ f \circ \phi(A)^{-1} = GF(f)$$

is a bijection. Since A and B are arbitrary, we conclude $G \circ F$ is a fully faithful functor and a symmetric argument shows $F \circ G$ is also fully faithful. Then, it is easy to conclude that F and G must be fully faithful as well.[401]

[400] Recall the definitions of monomorphisms and epimorphisms and the fact that isomorphisms are monic and epic.

[401] Recall Exercise B.32

What is more, the existence of an isomorphism $\eta_A : A \to FG(A)$ for any object A implies F (symmetrically G) has the following property.

Definition F.38 (Essentially surjective)**.** A functor $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ is **essentially surjective** if for any $X \in \mathbf{D}_0$, there exists $Y \in \mathbf{C}_0$ such that $X \cong F(Y)$.[402]

[402] Intuitively, this property means that while the image of F may not be all of $\mathbf{D}$, everything outside the image is at least isomorphic to somethig in the image.

We will show that these two properties (full faithfulness and essential surjectivity) are necessary and sufficient for F to be an equivalence.

Theorem F.39. *A functor $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ is an equivalence of categories if and only if F is fully faithful and essentially surjective.*

Proof. ($\Rightarrow$) Shown above.

($\Leftarrow$) We construct a functor $G : \mathbf{D} \rightsquigarrow \mathbf{C}$ such that $G \circ F \cong \mathrm{id}_{\mathbf{C}}$ and $F \circ G \cong \mathrm{id}_{\mathbf{D}}$.[403] Since F is essentially surjective, for any $A \in \mathbf{D}_0$, there exists an object $G(A) \in \mathbf{C}_0$ and an isomorphism $\phi_A : F(G(A)) \cong A$. Hence, $A \mapsto G(A)$ is a good candidate to describe the action of G on objects.

[403] The quasi-inverse of F. We can say *the* thanks to Exercise F.40.

Next, similarly to the converse direction, note that for any $A, B \in \mathbf{D}_0$, the map

$$f \mapsto \phi_B \circ f \circ \phi_A^{-1}$$

is a bijection from $\mathrm{Hom}_{\mathbf{D}}(A, B)$ to $\mathrm{Hom}_{\mathbf{D}}(FG(A), FG(B))$. Moreover, since the functor F is fully faithful, it induces a bijection

$$F_{GA,GB} : \mathrm{Hom}_{\mathbf{C}}(G(A), G(B)) \to \mathrm{Hom}_{\mathbf{D}}(FG(A), FG(B))$$

which in turns yields a bijection

$$G_{A,B} : \mathrm{Hom}_{\mathbf{D}}(A,B) \to \mathrm{Hom}_{\mathbf{C}}(G(A),G(B)) = f \mapsto F^{-1}_{GA,GB}(\phi_B \circ f \circ \phi_A^{-1}).$$

This is the action of G on morphisms. Observe that the construction of G ensures that $F \circ G \cong \mathrm{id}_{\mathbf{D}}$ through the natural transformation ϕ. It remains to show that G is indeed a functor and find a natural isomorphism $\eta : G \circ F \cong \mathrm{id}_{\mathbf{C}}$.

For any composable morphisms $(f,g) \in \mathbf{D}_2$, it is easy to verify that

$$F(G(f) \circ G(g)) = FG(f) \circ FG(g) = FG(f \circ g),$$

so functoriality of G because F is faithful. To find η, recall that the definition of G yields commutativity of (154) for any $f \in \mathrm{Hom}_{\mathbf{C}}(A,B)$.

$$\begin{tikzcd} F(A) \arrow[r, "F(f)"] \arrow[d, leftrightarrow, "\phi_{FA}"'] & F(B) \arrow[d, leftrightarrow, "\phi_{FB}"] \\ FGF(A) \arrow[r, "FGF(f)"'] & FGF(B) \end{tikzcd} \tag{154}$$

Then, because F is fully faithful, (**??**) also commutes in $\mathbf{C}$ where $\eta_X = F^{-1}_{X,GFX}(\phi_{FX})$, and we conclude that η is a natural isomorphism $\mathrm{id}_{\mathbf{C}} \cong G \circ F$.[404]

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, leftrightarrow, "\eta_A"'] & B \arrow[d, leftrightarrow, "\eta_B"] \\ GF(A) \arrow[r, "GF(f)"'] & GF(B) \end{tikzcd} \tag{155}$$

[404] You can manually derive that η_X is an isomorphism or use the fact that fully faithful functors reflect isomorphisms (Exercise C.52).

□

The insight to extract from this argument is that two categories are equivalent if they describe the same objects and morphisms with the only relaxation that isomorphic objects can appear any number of times in either category. In contrast, categories can only be isomorphic if they have exactly the same objects and morphisms.

When constructing the quasi-inverse of F in Theorem F.39, we had to pick one $G(A)$ for every A such that $A \cong FG(A)$ and one isomorphism $\phi_A : A \cong FG(A)$. These choices rely on the axiom of choice. There is some literature on doing category theory constructively and it relies on anafunctors. Those were defined precisely to avoid the axiom of choice in the proof above.

OL Exercise F.40 (NOW!)**.** Let $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ and $G, G' : \mathbf{D} \rightsquigarrow \mathbf{C}$ be two quasi-inverses to F. Show that $G \cong G'$.

OL Exercise F.41. Let $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ be an equivalence. Show that if $F \cong F'$, then F' is an equivalence.

We will detail a couple of *easy* examples of equivalences and briefly mention a few *harder* ones. Of course, all the isomorphisms of categories we saw earlier are examples of equivalences where the natural isomorphisms are identities.

Examples F.42 (Easy)**.** 1. Consider the full subcategory of **FinSet** consisting only of the sets $\emptyset, \{1\}, \{1,2\}, \dots, \{1,\dots,n\}, \dots$, we denoted it by **FinOrd**. The inclusion functor is fully faithful by definition and we claim it is essentially surjective. Indeed, any set $X \in \mathbf{FinSet}_0$ has a finite cardinality n, so $X \cong \{1,\dots,n\}$ and the latter belongs to **FinOrd**.

2. In a very similar fashion, an early result in linear algebra says that any finite dimensional vector space over a field k is isomorphic to k^n for some $n \in \mathbb{N}$. Thus, the category whose objects are k^n for all $n \in \mathbb{N}$ and morphisms are $m \times n$ matrices with entries in k,[405] which we denote $\mathbf{Mat}(k)$, is equivalent to the category of finite dimensional vector spaces.

[405] After making a choice of basis for all k^n, an $m \times n$ matrix with entries in k corresponds to a linear map $k^n \to k^m$.

3. A **partial** function $f : X \rightharpoonup Y$ is a function that may not be defined on all of X.[406] There is category **Par** of sets and partial functions where identity morphism and composition are defined straightforwardly.[407] We can view a partial function $f : X \rightharpoonup Y$ as a total function $f' : X \to Y + \mathbf{1}$ which sneds x to $f(x)$ when the latter is defined and to $* \in \mathbf{1}$ otherwise. Further extending f' to $[f', \mathrm{id}_{\mathbf{1}}] : X + \mathbf{1} \to Y + \mathbf{1}$, we can see any partial function as a function between pointed sets where the distinguished element corresponds to being undefined.

   This yields a fully faithful functor $F : \mathbf{Par} \rightsquigarrow \mathbf{Set}_*$ sending X to $(X + \mathbf{1}, *)$ and $f : X \rightharpoonup Y$ to $[f', \mathrm{id}_{\mathbf{1}}]$.[408] This functor is essentially surjective because for every pointed set (X, x), we find an isomorphism $(X \setminus \{x\} + \mathbf{1}, *) \to (X, x)$ that sends $y \in X \setminus \{x\}$ to y and $*$ to x. We infer the quasi-inverse to F sends a pointed set (X, x) to $X \setminus \{x\}$ and a function $f : (X, x) \to (Y, y)$ to the partial function $X \setminus \{x\} \to Y \setminus \{y\}$ that acts like f but is undefined whenever $f(a) = y$.

[406] In this context, a *normal* function defined on all of X is called **total**.
[407] You can view **Par** as the subcategory of **Rel** where you only take the relations $R \subseteq X \times Y$ satisfing for any $x \in X$ (cf. Remark B.24),

$$|\{y \in Y \mid (x, y) \in R\}| \leq 1.$$

[408] We have already seen in Corollary D.74 that $[f', \mathrm{id}_{\mathbf{1}}] = [g', \mathrm{id}_{\mathbf{1}}]$ if and only if $f' = g'$. It should be clear from the definition that $f' = g'$ if and only if $f = g$.

The first two examples and many other simple examples of equivalences are examples of skeletons. They are morally a subcategory where all the isomorphic copies are removed.

Definition F.43 (Skeleton)**.** A category is called **skeletal** if there it contains no two isomorphic objects. A **skeleton** of a category is an equivalent skeletal category.

Examples F.44. We have said that $\mathbf{FinOrd} \simeq \mathbf{FinSet}$ and $\mathbf{Mat}(k) \simeq \mathbf{FDVect}_k$ and we leave to you the easy task to check that these are examples of skeletons.[409]

Any posetal category is skeletal because whenever $x \leq y$ and $y \leq x$, we have $x = y$ which means no two distinct object can be isomorphic.

[409] Namely, you should show that no two sets in **FinOrd** are isomorphic and no two spaces in $\mathbf{Mat}(k)$ are isomorphic.

A category always has a skeleton if you assume the axiom of choice and the next result justifies us calling it *the* skeleton of a category.

OL Exercise F.45. Show that all skeletons of a category are isomorphic.

Here are other more interesting examples of equivalent categories.

Example F.46 (Medium)**.** Let $\mathbf{C}$ be a category, the functor $\mathsf{id} : \mathbf{C} \rightsquigarrow \mathbf{C}^{\to}$ sends X to id_X and $f : X \to Y$ to the commutative square in (156). This functor is an equivalence if and only if all morphisms in $\mathbf{C}$ are isomorphisms.[410] It is clearly fully faithful, so it is left to show id is essentially surjective if and only if $\mathbf{C}$ is a groupoid.

(⇒) For any $f : X \to Y \in \mathbf{C}_1$, by hypothesis, there exists $A \in \mathbf{C}_0$ such that $\mathrm{id}_A \cong f$ in $\mathbf{C}^{\to}$. Let $(s : A \to X, t : A \to Y)$ be the isomorphism, its inverse must be (s^{-1}, t^{-1}). Looking at the chain of commutative squares in (157), we can infer that $s \circ t^{-1}$ is the inverse of f.[411]

[410] Such a category is called a **groupoid**.

$$\begin{tikzcd} X \arrow[r, "\mathrm{id}_X"] \arrow[d, "f"'] & X \arrow[d, "f"] \\ Y \arrow[r, "\mathrm{id}_Y"'] & Y \end{tikzcd} \tag{156}$$

[411] The composition $f \circ s \circ t^{-1}$ is the top path of the combined two leftmost squares, the bottom path is $t \circ t^{-1} \circ \mathrm{id}_Y = \mathrm{id}_Y$. The composition $s \circ t^{-1} \circ f$ is the bottom path of the combined two rightmost squares, the top path is $\mathrm{id}_X \circ s \circ s^{-1} = \mathrm{id}_X$.

$$\begin{tikzcd}
Y \arrow[r, "t^{-1}"] \arrow[d, "\mathrm{id}_Y"'] & A \arrow[r, "s"] \arrow[d, "\mathrm{id}_A"'] & X \arrow[r, "s^{-1}"] \arrow[d, "f"'] & A \arrow[r, "s"] \arrow[d, "\mathrm{id}_A"'] & X \arrow[d, "\mathrm{id}_X"] \\
Y \arrow[r, "t^{-1}"'] & A \arrow[r, "t"'] & Y \arrow[r, "t^{-1}"'] & A \arrow[r, "s"'] & X
\end{tikzcd} \tag{157}$$

($\Leftarrow$) Let $f : X \to Y$ be an object of $\mathbf{C}^{\to}$, the inverse of f satisfies $f \circ f^{-1} = \mathrm{id}_Y$ and $f^{-1} \circ f = \mathrm{id}_X$, so the squares in (158) are isomorphisms in $\mathbf{C}^{\to}$ (they are inverses of each other). Thus, we find that f is isomorphic to id_X which is in the image of id.

$$\begin{tikzcd}
X \arrow[r, "\mathrm{id}_X"] \arrow[d, "\mathrm{id}_X"'] & X \arrow[d, "f"] \\
X \arrow[r, "f"'] & Y
\end{tikzcd} \qquad
\begin{tikzcd}
X \arrow[r, "f"] \arrow[d, "\mathrm{id}_X"'] & Y \arrow[d, "f^{-1}"] \\
X \arrow[r, "\mathrm{id}_X"'] & X
\end{tikzcd} \tag{158}$$

OL **Exercise F.47.** The category **Setoid** is the full subcategory of 2**Rel** containing only objects (X, R) where R is an equivalence relation. Is **Set** equivalent to **Setoid**?

Examples F.48 (Hard)**.** Examples of significant equivalences are all over the place in higher mathematics. However, they require a bit of work to describe them, thus let us only say a few words on a couple of them.

1. The equivalence between the category of affine schemes and the opposite of the category of commutative rings is a seminal result in scheme theory, a huge part of modern algebraic geometry.

2. The equivalence between Boolean lattices and Stone spaces is again seminal in the theory of Stone-type dualities. These can lead to deep connections between topology and logic. One application in particular is the study of the behavior of computer programs through formal semantics.

OL **Exercise F.49.** Show that equivalence of categories is an equivalence relation.

OL **Exercise F.50.** Show that $\mathbf{C} \simeq \mathbf{C}'$ and $\mathbf{D} \simeq \mathbf{D}'$ implies $[\mathbf{C}, \mathbf{D}] \simeq [\mathbf{C}', \mathbf{D}']$.

G Yoneda Lemma

We first defined an element of an object $X \in \mathbf{C}_0$ to be a morphism $\mathbf{1} \to X$. Our inspiration came from **Set** where $\mathrm{Hom}_{\mathbf{Set}}(\mathbf{1}, X) \cong X$. This is not a perfect categorification of the notion of element, because it works in some categories (e.g. **Poset**, **Top**, **Met**), but not in others (e.g. **Grp**, **Cat**[412], categories with no terminal object). In Exercise E.45, we found a workaround for **Grp**, namely, elements of G correspond to morphisms $\mathbb{Z} \to G$.

[412] In **Cat**, a morphism $\mathbf{1} \to \mathbf{C}$ corresponds to an object of $\mathbf{C}_0$, but depending on the context, it may be more relevant to define an element of $\mathbf{C}$ to be a morphism of $\mathbf{C}_1$.

Armed with our new abstract tools from last chapter, in particular natural isomorphisms, we can rigorously explain why $\mathbf{1}$ seems to *represent* the choice of an element in **Set**, why $\mathbb{Z}$ plays that role in **Grp**, and go further to find other things that are *representable*.

This journey quickly leads to the Yoneda lemma which formalizes our conviction[413] that studying mathematical objects through their interactions with other objects is "enough".

[413] Hopefully, you have been convinced by earlier chapters.

G.1 Representable Functors

Throughout this chapter, let $\mathbf{C}$ be a locally small category. Recall that for an object $A \in \mathbf{C}_0$, there are two Hom functors from $\mathbf{C}$ to **Set**. The covariant one, $\mathrm{Hom}_{\mathbf{C}}(A, -)$, sends an object $B \in \mathbf{C}_0$ to $\mathrm{Hom}_{\mathbf{C}}(A, B)$ and a morphism $f : B \to B'$ to $f \circ (-)$. The contravariant one, $\mathrm{Hom}_{\mathbf{C}}(-, A)$, sends an object $B \in \mathbf{C}_0$ to $\mathrm{Hom}_{\mathbf{C}}(B, A)$ and a morphism $f : B \to B'$ to $(-) \circ f$. In order to lighten the notation, we denote these functors $H^A : \mathbf{C} \rightsquigarrow \mathbf{Set}$ and $H_A : \mathbf{C}^{\mathrm{op}} \rightsquigarrow \mathbf{Set}$ respectively.[414]

[414] It is somewhat standard to use sub- and superscript as an indication for the *variance* of a notation. Note however that, while H^A is covariant and H_A contravariant, we are not talking about this. Instead we are interested in their *variance* in the parameter A, and we will, given a morphism $f : A \to A'$, construct a natural transformation $H^{A'} \Rightarrow H^A$ which means H^A is contravariant in A, and similarly, H_A is covariant in A.

Although these functors are sometimes interesting on their own, their full power is unleashed when they are related to other functors through natural transformations. Before doing that, let us investigate how nice Hom functors are. For instance, many Hom functors can be described in simpler terms.

Examples G.1. We are just revisiting things we already know.

1. Let $\mathbf{1} = \{*\}$ be the terminal object in **Set**, then what is the action of $H^{\mathbf{1}}$? For any object B,
$$H^{\mathbf{1}}(B) = \mathrm{Hom}_{\mathbf{Set}}(\mathbf{1}, B)$$
is easy to describe because for any element $b \in B$, there is a unique function $f : \mathbf{1} \to B = * \mapsto b$. Hence, there is an isomorphism from $H^{\mathbf{1}}(B)$ to B for any $B \in \mathbf{Set}_0$, it sends f to $f(*)$ and its inverse sends $b \in B$ to the map $* \mapsto b$.

Moreover, these isomorphisms are natural in B because (159) clearly commutes for any $f : B \to B'$, yielding a natural isomorphism $H^{\mathbf{1}} \cong \mathrm{id}_{\mathbf{Set}}$.

$$\begin{tikzcd} H^{\mathbf{1}}(B) \arrow[r, "f\circ(-)"] \arrow[d, leftrightarrow] & H^{\mathbf{1}}(B') \arrow[d, leftrightarrow] \\ B \arrow[r, "f"'] & B' \end{tikzcd} \tag{159}$$

2. Consider again the terminal object but in the category **Grp**, namely, the group **1** only containing an identity element. Then, for any group G, the set $H^{\mathbf{1}}(G)$ is a singleton because any homomorphism $f : \mathbf{1} \to G$ must send the identity to the identity and no other choice can be made. Therefore, unlike in **Set**, $H^{\mathbf{1}}$ is very uninteresting and acts like the constant functor $\Delta(\mathbf{1}) : \mathbf{Grp} \rightsquigarrow \mathbf{Set}$, i.e. $H^{\mathbf{1}} \cong \Delta(\mathbf{1})$.

3. A better choice of object to mimic the behavior of $\mathrm{id}_{\mathbf{Grp}}$ is the additive group $\mathbb{Z}$. Indeed, for any $g \in G$, there is a unique homomorphism $f : \mathbb{Z} \to G$ sending 1 to g.[415] A very similar argument as above yields a natural isomorphism between $H^{\mathbb{Z}}$ and the forgetful functor $U : \mathbf{Grp} \rightsquigarrow \mathbf{Set}$. (The identity functor on **Grp** does not have the same type as $H^{\mathbb{Z}}$, but $\mathrm{id}_{\mathbf{Set}}$ can be viewed as a forgetful functor from **Set** to itself.)

4. The terminal object in **Cat** is the category **1** with a single object $\bullet$ and no morphism other than the identity. For any category $\mathbf{C} \in \mathbf{Cat}_0$, a functor $\mathbf{1} \rightsquigarrow \mathbf{C}$ is just a choice of object. Therefore, the same argument will show that $H^{\mathbf{1}} \cong (-)_0$, where $(-)_0$ sends a category to its set[416] of objects and a functor to its action restricted on objects.

   In order to obtain a similar way to extract morphisms, consider the category **2** with two objects and a single morphism between them. One obtains a natural isomorphism $H^{\mathbf{2}} \cong (-)_1$.[417]

[415] Note that f is completely determined by $f(1)$ because the homomorphism properties imply that $f(n) = f(1) + \overset{n}{\cdots} + f(1)$, $f(-n) = f(n)^{-1}$, and $f(0)$ must be the identity.

[416] Recall that **Cat** only contains small categories.

[417] You can prove this as we did for $H^{\mathbf{1}} \cong (-)_0$ or use Example F.14.3.

Just like we benefitted from recognizing a category was isomorphic to a functor category (e.g. Theorem F.16 and Corollary F.17), we can benefit from finding a natural isomorphism between a functor and a Hom functor. For instance, we already know that the Hom functors are continuous,[418] and with Example G.1.4 we can infer

$$\left(\prod_{i\in I} \mathbf{C}_i\right)_0 = \prod_{i\in I} (\mathbf{C}_i)_0 \text{ and } \left(\prod_{i\in I} \mathbf{C}_i\right)_1 = \prod_{i\in I} (\mathbf{C}_i)_1.$$

In words, the objects of a product of categories are tuples of objects of each category and similarly for morphisms.[419] This suggest carefully studying representable functors.

[418] Theorem D.88 and Corollary D.89. Also recall that a functor naturally isomorphic to a continuous functor is also continuous, see Exercise F.12.

[419] We already knew that for the case of binary products, see Exercise D.11.

Definition G.2 (Representable functor)**.** A covariant functor $F : \mathbf{C} \rightsquigarrow \mathbf{Set}$ is **representable** if there is an object $X \in \mathbf{C}_0$ such that F is naturally isomorphic to $\mathrm{Hom}_{\mathbf{C}}(X, -)$. If F is contravariant, then it is representable if it is naturally isomorphic to $\mathrm{Hom}_{\mathbf{C}}(-, X)$. We call X the **representing** object of F.[420]

[420] The use of the definite article *the* is justified by Proposition G.5.

Examples G.3. Let us give examples of the contravariant kind.

1. Recall from Example C.2.2 the contravariant powerset functor $2^{-} : \mathbf{Set} \rightsquigarrow \mathbf{Set}$. It sends a set X to its powerset $2^X = \mathcal{P}(X)$ and a function $f : X \to Y$ to the inverse image $2^f = f^{-1} : \mathcal{P}(Y) \to \mathcal{P}(X)$. We can identify subsets of a given set with functions from this set into $\Omega = \{\perp, \top\}$.[421] This yields a bijection $2^X \cong H_{\Omega}(X)$

[421] See our discussion of subobject classifers in **Set**.

that is natural in X. Indeed, for all $f : X \to Y$, you can check (160) commutes,[422] so $2^- \cong H_\Omega$.

$$\begin{tikzcd} H_\Omega(X) \arrow[r, shift left, "p\mapsto p^{-1}(\top)"'] & 2^X \arrow[l, shift left, "S\mapsto\chi_S"'] \\ H_\Omega(Y) \arrow[u, "-\circ f"] \arrow[r, shift left, "p\mapsto p^{-1}(\top)"'] & 2^X \arrow[l, shift left, "S\mapsto\chi_S"'] \arrow[u, "2^f=f^{-1}"'] \end{tikzcd} \tag{160}$$

[422] Starting with $p : Y \to \Omega$ in the bottom left. The top path yields

$$(p \circ f)^{-1}(\top) = \{x \in X \mid p(f(x)) = \top\}.$$

The bottom path yields

$$f^{-1}(p^{-1}(\top)) = \{x \in X \mid p(f(x)) = \top\}.$$

2. Our first example of natural isomorphism (Example F.8.1) was the currying of a morphism $\lambda : \mathrm{Hom}_{\mathbf{C}}(- \times X, A) \cong \mathrm{Hom}_{\mathbf{C}}(-, A^X)$, where A^X is an exponential object. It turns out exponential objects can be defined via this natural isomorphism. Namely, there is an isomorphism $\ell : \mathrm{Hom}_{\mathbf{C}}(- \times X, A) \cong \mathrm{Hom}_{\mathbf{C}}(-, E)$ if and only if E is the exponential object and $\ell_E^{-1}(\mathrm{id}_E)$ is the evaluation morphism.[423]

[423] The expression $\ell_E^{-1}(\mathrm{id}_E)$ might look like it comes out of nowhere, but it is not so mysterious. Given the natural isomorphism ℓ, if we are looking for a moprhism of type $E \times X \to A$, then we may as well look for a morphism of type $E \to E$ and use the bijection ℓ_E^{-1}. What morphism of type $E \to E$ do we have? Only one is guaranteed to exist, the identity id_E. This chapter contains several instances of this kind of forced choice.

($\Leftarrow$) This was already shown in Example F.8.1 modulo the fact that $\lambda^{-1}\mathrm{id}_{A^X} = \mathsf{ev}$. For the latter, it suffices to note that $\lambda\mathsf{ev}$ must be id_{A^X} to make (161) commute.

$$\begin{tikzcd} A & A^X \times X \arrow[l, "\mathsf{ev}"'] \\ & A^X \times X \arrow[ul, "\mathsf{ev}"] \arrow[u, dashed, "\lambda\mathsf{ev}\times\mathrm{id}_X"'] \end{tikzcd} \tag{161}$$

($\Rightarrow$) Given ℓ, we show E is the exponential. For any $g : B \times X \to A$, we claim that $\ell_B(g)$ makes (162) commute. The naturality of ℓ^{-1} yields the following commutative square.

$$\begin{tikzcd} A & E \times X \arrow[l, "\ell_E^{-1}(\mathrm{id}_E)"'] \\ & B \times X \arrow[ul, "g"] \arrow[u, dashed, "\ell_B(g)\times\mathrm{id}_X"'] \end{tikzcd} \tag{162}$$

$$\begin{tikzcd} \mathrm{Hom}_{\mathbf{C}}(B \times X, A) & \mathrm{Hom}_{\mathbf{C}}(B, E) \arrow[l, "\ell_B^{-1}"'] \\ \mathrm{Hom}_{\mathbf{C}}(E \times X, A) \arrow[u, "-\circ(\ell_B(g)\times\mathrm{id}_X)"] & \mathrm{Hom}_{\mathbf{C}}(E, E) \arrow[l, "\ell_E^{-1}"] \arrow[u, "-\circ\ell_B(g)"'] \end{tikzcd} \tag{163}$$

Starting in the bottom right with id_E, the bottom path sends it to $\ell_E^{-1}(\mathrm{id}_E) \circ (\ell_B(g) \times \mathrm{id}_X)$ and the top path sends to $\ell_B^{-1}(\ell_B(g)) = g$. Commutativity lets us conclude $\ell_E^{-1}(\mathrm{id}_E) \circ (\ell_B(g) \times \mathrm{id}_X) = g$, i.e. (162) commutes.

In the first items of Examples G.1 and G.3, we made an arbitrary choice of set. That is, we could have taken any singleton instead of $\mathbf{1}$ in the first case and any set with two elements instead of Ω in the second. More generally, one can show that if $A \cong B$, then $H_A \cong H_B$ and $H^A \cong H^B$.

OL **Exercise G.4** (NOW!). Let $A, B \in \mathbf{C}_0$ be isomorphic objects. Show that $H^A \cong H^B$. Dually, show that $H_A \cong H_B$.

In particular, for any object E isomorphic to the exponential A^X, we have

$$H_E \cong H_{A^X} \cong \mathrm{Hom}_{\mathbf{C}}(- \times X, A),$$

which means E is also the exponential. In Exercise E.14, we also showed that if E satisfies the same universal property as A^X, then they must be isomorphic. In order to prove this using the natural isomorphism instead of the universal property, we would need a converse to Exercise G.4.

Perhaps surprisingly, it is true and it will follow from the Yoneda lemma, but we prove it on its own first as a warm-up for the proof of the lemma.

Proposition G.5. *Let $A, B \in \mathbf{C}_0$ be such that $H^A \cong H^B$, then $A \cong B$.*

Proof. The natural isomorphism gives two natural transformations $\phi : H^A \Rightarrow H^B$ and $\eta : H^B \Rightarrow H^A$ such that for any object $X \in \mathbf{C}_0$,

$$\eta_X \circ \phi_X : H^A(X) \to H^A(X) \quad \text{and} \quad \phi_X \circ \eta_X : H^B(X) \to H^B(X)$$

are identities. In order to show $A \cong B$, we will find two morphisms $f : B \to A$ and $g : A \to B$ such that $f \circ g = \mathrm{id}_A$ and $g \circ f = \mathrm{id}_B$. With the given data, there is no freedom to construct f and g. Since $\mathbf{C}$, A and B are arbitrary, there are only two morphisms that are required to exist, id_A and id_B. Next, we note that $\mathrm{id}_A \in H^A(A)$ and $\mathrm{id}_B \in H^B(B)$, hence, we can set $f := \phi_A(\mathrm{id}_A)$ and $g := \eta_B(\mathrm{id}_B)$.[424]

[424] To emphasize the point about *no freedom*, try to convince yourself that any morphisms of type $B \to A$ and $A \to B$ that we can construct from id_A, id_B, ϕ and η (the only data we have) must be equal to f and g as we defined them.

Now, $\phi_A(\mathrm{id}_A)$ is a morphism from B to A, so (164) commutes by naturality of η.

$$\begin{tikzcd} H_B(A) \arrow[r, "\eta_A"] & H_A(A) \\ H_B(B) \arrow[u, "\phi_A(\mathrm{id}_A)\circ(-)"] \arrow[r, "\eta_B"'] & H_A(B) \arrow[u, "\phi_A(\mathrm{id}_A)\circ(-)"'] \end{tikzcd} \tag{164}$$

We conclude, by starting with id_B in the bottom left, that

$$g \circ f = \phi_A(\mathrm{id}_A) \circ \eta_B(\mathrm{id}_B) = \eta_A(\phi_A(\mathrm{id}_A)) = \mathrm{id}_A.$$

A dual argument shows that

$$f \circ g = \eta_B(\mathrm{id}_B) \circ \phi_A(\mathrm{id}_A) = \phi_B(\eta_B(\mathrm{id}_B)) = \mathrm{id}_B,$$

and we have shown $A \cong B$. □

Corollary G.6 (Dual)**.** *Let $A, B \in \mathbf{C}_0$ be such that $H_A \cong H_B$, then $A \cong B$.*

Steve Awodey calls **Yoneda principle** the equivalences[425]

[425] They are the combination of Exercise G.4, Proposition G.5 and Corollary G.6.

$$H^A \cong H^B \Leftrightarrow A \cong B \Leftrightarrow H_A \cong H_B.$$

This is the formalization of the philosophical point we mentionned a few times already: an object is determined up to isomorphism by all its relations with all other objects. The Hom functor H^A (or H_A) makes for an efficient description of all the relations between A and all other objects.

Let us give two more concrete examples of representable functors.

Example G.7 (G acting on itself)**.** Any group G acts on itself by multiplication on the left. The corresponding functor, abusively denoted by $\mathbf{B}G \rightsquigarrow \mathbf{Set}$, sends $*$ to the set G and $g \in G$ to the bijection $h \mapsto gh$.[426] Note that this is the Hom functor

[426] Its inverse is $h \mapsto g^{-1}h$.

$\mathrm{Hom}_{\mathbf{B}G}(*,*) = H^*$. Indeed, it sends $*$ to $\mathrm{Hom}_{\mathbf{B}G}(*,-) = G$, and for any $g \in G$ $\mathrm{Hom}_{\mathbf{B}G}(*,g) : G \to G$ is the left multiplication by g because $g \circ h = gh$.

Fix another group action $F \in [\mathbf{B}G, \mathbf{Set}]$, we showed in Example F.13 that a natural transformation $f : H^* \Rightarrow F$ is a G–equivariant map, i.e. its only component f_* makes (165) commute for every $g \in G = H^*(*)$.

$$\begin{tikzcd} G \arrow[r, "f_*"] \arrow[d, "g-"'] & F(*) \arrow[d, "g\star-"] \\ G \arrow[r, "f_*"'] & F(*) \end{tikzcd} \tag{165}$$

Starting with 1_G on the top left, we find that $f_*(g) = g \star f_*(1_G)$. Thus, the equivariant map f_* is completely determined by where it sends 1_G. Since there is no constraint on that choice, we get a bijection between natural transformations $H^* \Rightarrow F$ and elements of $F(*)$.

The assignment $F \mapsto F(*)$ is functorial as we have seen when defining Ev, and you can also see it as the forgetful functor $U : [\mathbf{B}G, \mathbf{Set}] \rightsquigarrow \mathbf{Set}$ that forgets about the action of G. Thus, we can ask whether the bijection above is natural in F, i.e. does (166) commute for every $\phi : F \Rightarrow F'$? It does commute as both paths send f to $\phi_*(f_*(1_G))$, hence we find that U is representable with $U \cong \mathrm{Hom}_{[\mathbf{B}G,\mathbf{Set}]}(H^*, -)$.

$$\begin{tikzcd} \mathrm{Hom}_{[\mathbf{B}G,\mathbf{Set}]}(H^*,F) \arrow[r, "f\mapsto f_*(1_G)"] \arrow[d, "\phi\cdot-"'] & F(*) \arrow[d, "\phi_*"] \\ \mathrm{Hom}_{[\mathbf{B}G,\mathbf{Set}]}(H^*,F') \arrow[r, "f\mapsto f_*(1_G)"'] & F'(*) \end{tikzcd} \tag{166}$$

Example G.8 (Elements of a ring)**.** In **Ring** just like in **Grp**, the terminal object is the ring containing only one element that is the zero and identity at the same time. Thus, there can be no morphism $\mathbf{1} \to R$ unless $R = \mathbf{1}$.[427]

[427] A ring homomorphism must send 0 to 0 and 1 to 1, so if $0 = 1$ in the source then 0 must equal 1 in the target as well.

Let us try what we did for **Grp**: replace $\mathbf{1}$ with $\mathbb{Z}$. Unfortunately, a ring homomoprhism $f : \mathbb{Z} \to R$ is too constrained. We must have $f(0) = 0_R$ and $f(1) = 1_R$, and any other value is forced by the homomorphism properties:

$$f(n) = f(1) + \overset{n}{\cdots} + f(1) = 1_R + \overset{n}{\cdots} + 1_R \text{ and } f(-n) = -f(n).$$

This means $\mathbb{Z}$ is the initial ring, and we can prove $H^{\mathbb{Z}}$ is naturally isomorphic to the constant functor $\Delta(\mathbf{1}) : \mathbf{Ring} \rightsquigarrow \mathbf{Set}$ (see Exercise G.18).

We need to add one element, say x, to $\mathbb{Z}$ so that f can map x anywhere, but no other choice can be made.[428] For the "map x anywhere" part, we must make sure that x is free of any constraint other than the properties of a ring. That is, it has an additive inverse $-x$, it satifies $x + x = (1x + 1x) = (1+1)x = 2x$ and other similar equations, it has powers like $x^2 = x \cdot x$ and $x^3 = x \cdot x \cdot x$, there are combinations like $5 + 2x + 4x^5$, and so on. The "no other choice" part is a consequence of the homomorphism properties. If the image of x is known, then the images of all the multiples and powers of x and combinations of them and other elements of $\mathbb{Z}$ are known too.

[428] This is essentially what we have done to go from $\mathbf{1}$ to $\mathbb{Z}$ in **Grp**. The integers can be seen as the group $\mathbf{1} = \{0\}$ where we add 1 (it is not the identity), its inverse -1 and letting the group operation do its thing. For instance $2 = 1 + 1$, $3 = 1 + 1 + 1$, etc.

In short, we are talking about the ring $\mathbb{Z}[x]$ of polynomials with one variable and coefficients in $\mathbb{Z}$. A ring homomorphism $\mathbb{Z}[x] \to R$ is completely determined by where it sends x, and we leave you to show $H^{\mathbb{Z}[x]}$ is naturally isomorphic to the forgetful functor $\mathbf{Ring} \rightsquigarrow \mathbf{Set}$.[429]

[429] With the Yoneda principle, we now have the promised categorical definition of polynomials from Example D.58.3. Exercise G.9 generalizes this to multivariate polynomials with non-integer coefficients.

With a slight modification, we can show the units functor $(-)^\times : \mathbf{Ring} \rightsquigarrow \mathbf{Set}$ (the functor from Example F.5 is composed with the forgetful functor from **Grp** to **Set**) is representable. The ring $\mathbb{Z}[x, x^{-1}]$ is $\mathbb{Z}[x]$ where we add a multiplicative inverse to x. It satisfies all the expected equations (e.g. $x \cdot x^{-1} = 1$, $x^2 \cdot x^{-3} = x^{-1}$, etc.) and no other. A ring homomoprhism $f : \mathbb{Z}[x, x^{-1}] \to R$ must send x^{-1} to the inverse of $f(x)$. Therefore, $f(x)$ is now restricted to $R^\times$. We leave you to show $H^{\mathbb{Z}[x,x^{-1}]} \cong (-)^\times$.

OL **Exercise G.9.** Let $U : \mathbf{Ring} \rightsquigarrow \mathbf{Set}$ be the forgetful functor and, for any $n \in \mathbb{N}$, $(-)^n : \mathbf{Ring} \rightsquigarrow \mathbf{Ring}$ the n–wise product functor.

1. Show that $H^{\mathbb{Z}[x_1,\dots,x_n]}$ is naturally isomorphic to the composition $U \circ (-)^n$.
2. For any ring R, show that $H^{R[x]} \cong H^R \times U$.[430]
3. Make up a categorical definition of $R[x_1, \dots, x_n]$ using this characterization. Does item 1 make you more confident in your definition?

[430] For this to typecheck, the R.H.S. must be the product inside $[\mathbf{Ring}, \mathbf{Set}]$, i.e. $(H^R \times U)(S) = \mathrm{Hom}(R, S) \times US$.

G.2 Yoneda Lemma

Taking a closer look at our solution to Exercise G.4, we find the assignments $A \mapsto H^A$ and $A \mapsto H_A$ are functorial.

Definition G.10 (Yoneda embeddings)**.** The contravariant **Yoneda embedding**[431] $H^{(-)} : \mathbf{C}^{\mathrm{op}} \rightsquigarrow [\mathbf{C}, \mathbf{Set}]$ sends $A \in \mathbf{C}_0$ to the Hom functor H^A and a morphism $f : A' \to A$ to the natural transformation $H^f : H^A \Rightarrow H^{A'}$ defined by $H^f_B = \mathrm{Hom}_{\mathbf{C}}(f, B) = (-) \circ f$ for every $B \in \mathbf{C}_0$. The naturality of H^f follows from associativity[432]: for any $g : B \to B'$, (167) commutes.

[431] Yoneda embeddings and the Yoneda lemma are named in honor of Nobuo Yoneda.

[432] Starting with h in the top left. The top path sends it to $g \circ (h \circ f)$ and the bottom path sends it to $(g \circ h) \circ f$. Since composition is associative, both paths are the same function.

$$\begin{tikzcd} H^A(B) \arrow[r, "(-)\circ f"] \arrow[d, "g\circ(-)"'] & H^{A'}(B) \arrow[d, "g\circ(-)"] \\ H^A(B') \arrow[r, "(-)\circ f"'] & H^{A'}(B') \end{tikzcd} \tag{167}$$

The covariant embedding $H_{(-)} : \mathbf{C} \rightsquigarrow [\mathbf{C}^{\mathrm{op}}, \mathbf{Set}]$ sends $B \in \mathbf{C}_0$ to the Hom functor H_B and a morphism $f : B \to B'$ to the natural transformation $H_f : H_B \Rightarrow H_{B'}$ defined by $(H_f)_A = \mathrm{Hom}_{\mathbf{C}}(A, f) = f \circ (-)$ for any $A \in \mathbf{C}_0$.[433] In order to harmonize the notation, we write H^A_f instead of $(H_f)_A$. Now the subscript of H always goes in the target of the Hom, and the superscript alawys goes in the source.

[433] Naturality follows from associativity of composition again.

Another way to obtain these embeddings (incidentally proving they are functors) is to curry the Hom bifunctor. Indeed, you can verify that

$$H^- = \Lambda\mathrm{Hom}(-,-) \text{ and } H_- = \Lambda(\mathrm{Hom}(-,-) \circ \mathsf{swap}).$$

The embeddings are called like that (c.f. Exercise C.20) because both functors are injective on objects[434] and fully faithful as will follow from the Yoneda lemma.

[434] If $A \neq B$, then $H^A(A)$ contains id_A but $H^B(A)$ does not, so $H^A \neq H^B$.

We now understand how an object $A \in \mathbf{C}_0$ can be understood by studying the representable H^A. In some sense, H^A tells us how A views the category it is in. Since the representable H^A is an object of the category $[\mathbf{C}, \mathbf{Set}]$, it is daring to try and understand it via the representable H^{H^A}. In other words, how does H^A see other functors in $[\mathbf{C}, \mathbf{Set}]$.

We have already got a problem. Even if $\mathbf{C}$ is locally small, there is no guarantee that $[\mathbf{C}, \mathbf{Set}]$ is locally small. Thus, $H^{H^A} = \mathrm{Hom}_{[\mathbf{C},\mathbf{Set}]}(H^A, -)$ might no be a well-defined functor.[435] To avoid confusing or cluttered notation, we write instead

[435] We do not know what category it lands in.

$\mathrm{Nat}(H^A, -)$ because, for a functor $F : \mathbf{C} \rightsquigarrow \mathbf{Set}$, $\mathrm{Nat}(H^A, F)$ is the collection of natural transformations from H^A to F.

We already saw that for every morphism $f : B \to A$ in $\mathbf{C}$, there is an element $H^f \in \mathrm{Nat}(H^A, H^B)$. Does every natural transformation of this type arise like that? Given a natural transformation $\alpha : H^A \Rightarrow H^B$ constructed from an unknown morphism $B \to A$, we can figure out what is that morphism by looking at $\alpha_A(\mathrm{id}_A)$.[436] Indeed, if $\alpha = H^f$ for some given $f : B \to A$, then

$$\alpha_A(\mathrm{id}_A) = H^f_A(\mathrm{id}_A) = \mathrm{id}_A \circ f = f.$$

Even if we do not know such an f, $\alpha_A(\mathrm{id}_A)$ is still a morphism $B \to A$. It turns out we can exploit naturality to show α must be the natural transformation $H^{\alpha_A(\mathrm{id}_A)}$.

[436] Once again, this choice is forced on us by the data we have. We are only given α and we need to find an element of $\mathrm{Hom}(B, A)$. It turns out α_A has type $\mathrm{Hom}(A, A) \to \mathrm{Hom}(B, A)$, so it remains to find an element of $\mathrm{Hom}(A, A)$. Since we know nothing else about $\mathbf{C}$, we can only pick id_A, because $\mathrm{Hom}(A, A)$ might contain no other morphism.

What can we say when the target of α is not representable? i.e. $\alpha : H^A \Rightarrow F$ for some functor $F : \mathbf{C} \rightsquigarrow \mathbf{Set}$. Our trick from above tells us every such α yields an element $\alpha_A(\mathrm{id}_A) \in F(A)$. Again relying on naturality, we can show every element $a \in F(A)$ gives a transformation $\alpha : H^A \Rightarrow F$ satisfying $\alpha_A(\mathrm{id}_A) = a$.

In short, the surprising relation described by the Yoneda lemma is an isomorphism between $\mathrm{Nat}(H^A, F)$ and $F(A)$ that is natural in F and A. We first show the isomorphism and then the naturality.

Lemma G.11 (Yoneda lemma I). *For any $A \in \mathbf{C}_0$ and $F : \mathbf{C} \rightsquigarrow \mathbf{Set}$,*

$$\mathrm{Nat}(H^A, F) \cong F(A).$$

Proof. Let $\phi_{A,F} : \mathrm{Nat}(H^A, F) \to F(A)$ be defined by $\alpha \mapsto \alpha_A(\mathrm{id}_A)$.[437] In the opposite direction, let $\eta_{A,F} : F(A) \to \mathrm{Nat}(H^A, F)$ send an element $a \in F(A)$ to the natural transformation with components $(\eta_{A,F}(a))_B : \mathrm{Hom}_{\mathbf{C}}(A, B) \to F(B) = f \mapsto F(f)(a)$ for each $B \in \mathbf{C}_0$.[438] Checking (168) commutes for any $g : B \to B'$ shows that $\eta_{A,F}(a)$ is a natural transformation. Starting with f in the top left, the top path sends it to $F(g)(F(f)(a))$ and the bottom path sends it to $F(g \circ f)(a)$. These two are equal by functoriality, i.e. $F(g) \circ F(f) = F(g \circ f)$.

[437] As we said earlier, this is the only way to obtain an element of $F(A)$ from the given data.

[438] Again this definition is the only one that type-checks. With a functor F, an element of $F(A)$, and a morphism in $\mathrm{Hom}_{\mathbf{C}}(A, B)$, we can apply $F(f) : F(A) \to F(B)$ to get an element of $F(B)$.

$$\begin{tikzcd} H^A(B) \arrow[r, "F(-)(a)"] \arrow[d, "g\circ(-)"'] & F(B) \arrow[d, "F(g)"] \\ H^A(B') \arrow[r, "F(-)(a)"'] & F(B') \end{tikzcd} \tag{168}$$

We now check that $\phi_{A,F}$ and $\eta_{A,F}$ are inverses. First, $(\eta \circ \phi)_{A,F}$ sends $\alpha \in \mathrm{Nat}(H^A, F)$ to $\eta_{A,F}(\alpha_A(\mathrm{id}_A))$, and at any $B \in \mathbf{C}_0$, we have

$$\begin{aligned} (\eta_{A,F}(\alpha_A(\mathrm{id}_A)))_B(f) &= F(f)(\alpha_A(\mathrm{id}_A)) && \text{def of } \eta \\ &= \alpha_B(H^A(f)(\mathrm{id}_A)) && \mathsf{NAT}(\alpha, A, B, f) \\ &= \alpha_B(f \circ \mathrm{id}_A) && \text{def of } H^A \\ &= \alpha_B(f), \end{aligned}$$

thus $\alpha = (\eta \circ \phi)_{A,F}(\alpha)$.

Conversely, $(\phi \circ \eta)_{A,F}$ sends $a \in F(A)$ to $\eta_{A,F}(a)_A(\mathrm{id}_A) = F(\mathrm{id}_A)(a) = a$, and we can conclude that $\eta_{A,F}$ and $\phi_{A,F}$ are natural isomorphisms. $\square$

Corollary G.12 (Dual). *For any $A \in \mathbf{C}_0$ and $F : \mathbf{C}^{\mathrm{op}} \rightsquigarrow \mathbf{Set}$, $\mathrm{Nat}(H_A, F) \cong F(A)$.*

We already mentionned a consequence of this result.

Corollary G.13. *The Yoneda embeddings $H^{(-)}$ and $H_{(-)}$ are fully faithful.*[439]

[439] Recall from Exercises C.51 and C.52 that when a functor F is fully faithful, $A \cong B$ if and only if $FA \cong FB$. Thus, Exercise G.4, Proposition G.5 and Corollary G.6 are all corollaries of this.

Proof. Applying the lemma with $F = H^B$, we find an isomorphism

$$\mathrm{Nat}(H^A, H^B) \cong H^B(A) = \mathrm{Hom}_{\mathbf{C}}(B, A)$$

In the right to left direction, this isomorphism sends $f : B \to A$ to $H^f : H^A \Rightarrow H^B$.[440] This is the action of the functor $H^{(-)}$ on the homset $\mathrm{Hom}_{\mathbf{C}}(B, A)$. Therefore, for all $A, B \in \mathbf{C}_0$, $f \mapsto H^f$ is a bijection, which means $H^{(-)}$ is fully faithful.

[440] By unrolling the definition of $\eta_{A,H^B}(f)$, we find its component at $A' \in \mathbf{C}_0$ sends $h \in \mathrm{Hom}_{\mathbf{C}}(A, A')$ to $h \circ f \in \mathrm{Hom}_{\mathbf{C}}(B, A')$. So $\eta_{A,H^B}(f) = H^f$.

The dual argument shows $H_{(-)}$ is fully faithful. □

Another consequence is that $\mathrm{Nat}(H^A, F)$ is a set (because it is isomorphic to $F(A)$ which is a set), and this allows us to formally state the second part of the Yoneda lemma.[441]

[441] That $\phi_{A,F}$ and $\eta_{A,F}$ are natural in A and F.

The assignment $(A, F) \mapsto \mathrm{Nat}(H^A, F)$ is a functor $\mathbf{C} \times [\mathbf{C}, \mathbf{Set}] \rightsquigarrow \mathbf{Set}$ with the action on morphisms given by[442]

$$(g, \mu) : (A, F) \to (A', F') \mapsto \mu \cdot (-) \cdot H^g : \mathrm{Nat}(H^A, F) \to \mathrm{Nat}(H^{A'}, F').$$

[442] If $g : A \to A'$, $\mu : F \Rightarrow F'$, and $\eta \in \mathrm{Nat}(H^A, F)$, we have the composite

$$H^{A'} \overset{H^g}{\Longrightarrow} H^A \overset{\eta}{\Rightarrow} F \overset{\mu}{\Rightarrow} F' \in \mathrm{Nat}(H^{A'}, F').$$

We can check this preserves identities and composition. The identity morphism on (A, F) is $(\mathrm{id}_A, \mathbb{1}_F)$, and it is sent to $\mathbb{1}_F \cdot (-) \cdot H^{\mathrm{id}_A}$, that is pre- and post-composition by the identities.[443] Given two morphisms $(g, \mu) : (A, F) \to (A', F')$ and $(g', \mu') : (A', F') \to (A'', F'')$, associativity of vertical composition implies

[443] It follows from functoriality of $H^{(-)}$ that $H^{\mathrm{id}_A} = \mathbb{1}_{H^A}$.

$$(\mu' \cdot (-) \cdot H^{g'}) \circ (\mu \cdot (-) \cdot H^g) = (\mu' \cdot \mu) \cdot (-) \cdot (H^g \cdot H^{g'}) = (\mu' \cdot \mu) \cdot (-) \cdot H^{g' \circ g}.$$

The type of $\mathrm{Nat}(H^-, -)$ can be confusing. Just for a moment, think of $\mathrm{Nat}(-, -)$ as a Hom bifunctor.[444] Then, instead of seeing H^- as a functor $\mathbf{C}^{\mathrm{op}} \rightsquigarrow [\mathbf{C}, \mathbf{Set}]$, see it instead as $\mathbf{C} \rightsquigarrow [\mathbf{C}, \mathbf{Set}]^{\mathrm{op}}$. Then, $\mathrm{Nat}(H^-, -)$ is the composite

[444] Strictly speaking $[\mathbf{C}, \mathbf{Set}]$ might not be locally small, so the functor $\mathrm{Nat}(-, -)$ is not well-defined.

$$\mathbf{C} \times [\mathbf{C}, \mathbf{Set}] \xrightarrow{H^- \times \mathrm{id}} [\mathbf{C}, \mathbf{Set}]^{\mathrm{op}} \times [\mathbf{C}, \mathbf{Set}] \xrightarrow{\mathrm{Nat}(-,-)} \mathbf{Set} \; .$$

The assignment $(A, F) \mapsto F(A)$ is another functor of the same type. We denoted it by Ev,[445] its action on morphisms is defined by

[445] See Example F.36.5.

$$(g, \mu) : (A, F) \to (A', F') \mapsto F'(g) \circ \mu_A = \mu_{A'} \circ F(g) : F(A) \to F'(A').$$

Lemma G.14 (Yoneda lemma II). *There is a natural isomorphism* $\mathrm{Nat}(H^-, -) \cong \mathsf{Ev}$.

Proof. The components of this isomorphism are the ones described in the first part. It remains to show that ϕ is natural in (A, F).[446] For any $(g, \mu) : (A, F) \to (A', F')$, we need to show the following square commutes.

[446] By Exercise F.7, it is enough to show it is natural in A and natural in F separately. We do both at the same time because it is not much harder.

$$\begin{tikzcd}
\mathrm{Nat}(H^A, F) \arrow[r, "\phi_{A,F}"] \arrow[d, "\mu \cdot (-) \cdot H^g"'] & F(A) \arrow[d, "F'(g) \circ \mu_A"] \\
\mathrm{Nat}(H^{A'}, F') \arrow[r, "\phi_{A',F'}"'] & F'(A')
\end{tikzcd} \tag{169}$$

Starting with a natural transformation $\alpha \in \mathrm{Nat}(H^A, F)$, the bottom path sends it to $(\mu \cdot \alpha \cdot H^g)_{A'}(\mathrm{id}_{A'})$ and the top path sends it to $(F'(g) \circ \mu_A)(\alpha_A(\mathrm{id}_A))$. The following derivation shows they are equal.

$$\begin{aligned}
(\mu \cdot \alpha \cdot H^g)_{A'}(\mathrm{id}_{A'}) &= (\mu_{A'} \circ \alpha_{A'})(H^g_{A'}(\mathrm{id}_{A'})) && \text{def of } \cdot \\
&= (\mu_{A'} \circ \alpha_{A'})(g) && \text{def of } H^g_{A'} \\
&= (\mu_{A'} \circ \alpha_{A'})(H^A_g(\mathrm{id}_A)) && \text{def of } H^A_g \\
&= (\mu_{A'} \circ \alpha_{A'} \circ H^A_g)(\mathrm{id}_A) && \\
&= (\mu_{A'} \circ F(g) \circ \alpha_A)(\mathrm{id}_A) && \mathsf{NAT}(\alpha, A, A', g) \\
&= (F'(g) \circ \mu_A)(\alpha_A(\mathrm{id}_A)) && \mathsf{NAT}(\mu, A, A', g) \qquad \square
\end{aligned}$$

Corollary G.15 (Dual)**.** *There is a natural isomorphism* $\mathrm{Nat}(H_-, -) \cong \mathsf{Ev}$.[447]

[447] We can typecheck this as before. We see H_- as a functor $\mathbf{C}^{\mathrm{op}} \rightsquigarrow [\mathbf{C}^{\mathrm{op}}, \mathbf{Set}]^{\mathrm{op}}$ (c.f. Exercise C.9). Then $\mathrm{Nat}(H_-, -) = \mathrm{Nat}(-,-) \circ H_- \times \mathrm{id}$.

While the Yoneda lemma is called a lemma, it is extremely important and powerful. We already said how it gives category theorists reasons to study an object through its relations to other objects (via the Yoneda principle). In a shallow exploration of category theory, this might seem like the only point[448] of the Yoneda lemma.

[448] I find it already quite grandiose.

Another result with a similar status in mathematics — it looks motivated only by philosophical and meta considerations — is Cayley's theorem. It states that any group is isomorphic to the subgroup of a permutation group.[449] Remarkably, the Yoneda lemma can be understood as a generalization of Cayley's theorem. This is our first application of Yoneda.

[449] It is important to group theorists because they are interested in studying symmetries of geometric shapes or other things, and these can easily be seen as subgroups of permutation groups. Thus, the abstract notion of group is made more concrete by Cayley's theorem.

Example G.16 (Cayley's theorem with the Yoneda lemma)**.** Recall the first part of the Yoneda lemma which states that for a category $\mathbf{C}$, a functor $F : \mathbf{C} \rightsquigarrow \mathbf{Set}$ and an object $A \in \mathbf{C}_0$, we have

$$\mathrm{Nat}(\mathrm{Hom}(A, -), F) \cong F(A).$$

Moreover, we know the explicit maps, namely, a natural transformation ϕ in the L.H.S. is mapped to $\phi_A(\mathrm{id}_A)$ and an element $a \in F(A)$ is mapped to the natural transformation whose component at $B \in \mathbf{C}_0$ is $\phi_B = f \mapsto F(f)(a)$.

Let us apply this to $\mathbf{C}$ being the delooping of a group G. Recall that any functor $F : \mathbf{B}G \rightsquigarrow \mathbf{Set}$ sends $*$ to a set S and any $g \in G$ to a permutation of S, it corresponds to an action of G on S.

To use the Yoneda lemma, our only choice of object for A is $*$ and we will choose for F the Hom functor $F = \mathrm{Hom}_{\mathbf{B}G}(*, -)$. The Yoneda lemma yields

$$\mathrm{Nat}(\mathrm{Hom}_{\mathbf{B}G}(*, -), \mathrm{Hom}_{\mathbf{B}G}(*, -)) \cong \mathrm{Hom}_{\mathbf{B}G}(*, *).$$

We already know that the R.H.S. is G,[450] but we have to do a bit of work to understand the L.H.S. First, observe that a natural transformation $\phi : \mathrm{Hom}_{\mathbf{B}G}(*, -) \Rightarrow \mathrm{Hom}_{\mathbf{B}G}(*, -)$ is just one morphism $\phi_* : \mathrm{Hom}_{\mathbf{B}G}(*, *) \to \mathrm{Hom}_{\mathbf{B}G}(*, *)$. Namely, it is a map from G to G. Second, recalling that $\mathrm{Hom}_{\mathbf{B}G}(*, g) = g \circ (-)$ and that $*$ is

[450] By definition of $\mathbf{B}G$.

the only object in $\mathbf{C}_0$, we get that ϕ_* must only make (170) commute.

$$\begin{array}{ccc} G & \xrightarrow{\phi_*} & G \\ {\scriptstyle g\circ(-)}\downarrow & & \downarrow{\scriptstyle g\circ(-)} \\ G & \xrightarrow[\phi_*]{} & G \end{array} \tag{170}$$

This is equivalent to $\phi_*(g \cdot h) = g \cdot \phi_*(h)$, and we get that each ϕ_* is a G–equivariant map from G to itself.[451] Denote the set of such maps by $\mathrm{Hom}_G(G, G)$. We obtain that, as sets,

$$\mathrm{Hom}_G(G, G) \cong G.$$

Now, we can check that $\mathrm{Hom}_G(G, G)$ is a subgroup of Σ_G (the group of permutations of the set G) and that the bijection is in fact an group isomorphism. Cayley's theorem follows.

We have to show that id_G is in $\mathrm{Hom}_G(G, G)$, that maps in $\mathrm{Hom}_G(G, G)$ are bijective, and that they are stable under composition and taking inverses. First, we have $\mathrm{id}_G(g \cdot h) = g \cdot h = g \cdot \mathrm{id}_G(h)$, so $\mathrm{id}_G \in \mathrm{Hom}_G(G, G)$. Second, let f be a G–equivariant map. For any $g \in G$, we have $f(g) = f(g \cdot 1) = g \cdot f(1)$, that is f acts on G by right multiplication by $f(1)$. Thus, it is bijective with its inverse being right multiplication by $f(1)^{-1}$. Third, if f and f' are both G–equivariant map, then

$$(f \circ f')(g \cdot h) = f(f'(g \cdot h)) = f(g \cdot f'(h)) = g \cdot (f \circ f')(h),$$

hence $f \circ f'$ is G–equivariant. Finally, we saw f^{-1} is right multiplication by $f(1)^{-1}$, and it is G–equivariant as $f^{-1}(g \cdot h) = g \cdot h \cdot f(1)^{-1} = g \cdot f^{-1}(h)$. We conclude that $\mathrm{Hom}_G(G, G)$ is a subgroup of Σ_G.

The final check is that the Yoneda bijection $G \to \mathrm{Hom}_G(G, G)$ sending g to $(-) \cdot g$ is a group homomorphism.[452] It is clear that it sends the identity to the identity and for any $g, h \in G$

$$(-) \cdot gh = ((-) \cdot g) \cdot h = ((-) \cdot h) \circ ((-) \cdot g),$$

so this is a group homomorphism.

I would like to believe this book is not a "shallow exploration of category theory", so we will also see more concrete uses of Yoneda.

Example G.17 (Exponentials in $\mathbf{DGph}$)**.** We saw in Chapter E that $\mathbf{DGph}$ is a topos, so it has exponentials, but we did not write a nice description for them.[453] We will do this here relying on Yoneda and the isomorphism $\mathbf{DGph} \cong [V \rightrightarrows E, \mathbf{Set}]$ outlined in Example F.22.

G.3 Universality as Representability

Representability is one of the two ways to describe universal constructions that we hinted at at the end of Chapter E. In this section, we will explore how any universal property is equivalent to representability of some functor. Since (co)limits and

[451] We see G as a G–set with the action of left multiplication as in Example G.7.

[452] isomorphism follows because it is a bijection.

[453] Theoretically, we know how to compute them because we have seen how to take power objects in Example E.24 and (co)limits in Example F.22, but we will take a more direct approach here.

universal morphisms are initial or terminal objects in some category, there is a first trivial way to express universality as representability.

OL **Exercise G.18** (NOW!). Let $X \in \mathbf{C}_0$ and $\Delta(\mathbf{1}) : \mathbf{C} \rightsquigarrow \mathbf{Set}$ be the constant functor at the singleton $\mathbf{1} = \{\star\}$. Show that $\mathrm{Hom}_{\mathbf{C}}(X, -) \cong \Delta(\mathbf{1})$ if and only if X is initial. Dually, $\mathrm{Hom}_{\mathbf{C}}(-, X) \cong \Delta(\mathbf{1})$ if and only if X is terminal.[454]

[454] In the dual statement, the source of $\Delta(\mathbf{1})$ is $\mathbf{C}^{\mathrm{op}}$.

It turns out this result is not very useful.

Proposition G.19. *Let $X, Y \in \mathbf{C}_0$. The product of X and Y exists if and only if there exists $P \in \mathbf{C}_0$ such that $\mathrm{Hom}_{\mathbf{C}\times\mathbf{C}}(\Delta_{\mathbf{C}}(-), (X, Y)) \cong \mathrm{Hom}_{\mathbf{C}}(-, P)$. The product is P.*

Proof. ($\Rightarrow$) Let $P = X \times Y$, for any $A \in \mathbf{C}_0$, there is an isomorphism

$$\mathrm{Hom}_{\mathbf{C}\times\mathbf{C}}((A, A), (X, Y)) \cong \mathrm{Hom}_{\mathbf{C}}(A, X \times Y)$$

which sends the pair $(f : A \to X, g : A \to Y)$ to $\langle f, g\rangle : A \to X \times Y$.[455] In the other direction, $p : A \to X \times Y$ is sent to the pair $(\pi_X \circ p, \pi_Y \circ p)$. Let us show it is natural in A. For any $m : A' \to A$, (171) commutes because the top path sends the pair (f, g) to the morphism $\langle f, g\rangle$ then to $\langle f, g\rangle \circ m = \langle f \circ m, g \circ m\rangle$ and the bottom path sends (f, g) to $(f, g) \circ (m, m) = (f \circ m, g \circ m)$ which is then sent to $\langle f \circ m, g \circ m\rangle$.

[455] Recall that $\langle f, g\rangle$ is the unique morphism satisfying $\pi_X \circ \langle f, g\rangle = f$ and $\pi_Y \circ \langle f, g\rangle = g$. Be careful not to confuse it with a pair of morphisms.

$$\begin{tikzcd}
\mathrm{Hom}_{\mathbf{C}\times\mathbf{C}}((A, A), (X, Y)) \arrow[r, "\sim"] \arrow[d, "-\circ(m,m)"'] & \mathrm{Hom}_{\mathbf{C}}(A, X \times Y) \arrow[d, "-\circ m"] \\
\mathrm{Hom}_{\mathbf{C}\times\mathbf{C}}((A', A'), (X, Y)) \arrow[r, "\sim"'] & \mathrm{Hom}_{\mathbf{C}}(A', X \times Y)
\end{tikzcd} \tag{171}$$

($\Leftarrow$) First, we define π_X and π_Y to be the pair of morphisms corresponding to id_P under the isomorphism $\mathrm{Hom}_{\mathbf{C}\times\mathbf{C}}((P, P), (X, Y)) \cong \mathrm{Hom}_{\mathbf{C}}(P, P)$.[456] Given two morphisms $f : A \to X$ and $g : A \to Y$, the isomorphism

[456] Once more, we are making a forced choice. To define the projections, we need two morphims $P \to X$ and $P \to Y$. By the natural isomorphism of the hypothesis, it is enough to find a morphism $P \to P$. We can only take id_P as we know nothing else about $\mathbf{C}$.

$$\mathrm{Hom}_{\mathbf{C}\times\mathbf{C}}((A, A), (X, Y)) \cong \mathrm{Hom}_{\mathbf{C}}(A, P)$$

yields a unique morphism $! : A \to P$. To see that $\pi_X \circ ! = f$ and $\pi_Y \circ ! = g$ we start with id_P in the top right of (172) which commutes by hypothesis.

$$\begin{tikzcd}
\mathrm{Hom}_{\mathbf{C}\times\mathbf{C}}((P, P), (X, Y)) \arrow[d, "-\circ(!,!)"'] & \mathrm{Hom}_{\mathbf{C}}(P, P) \arrow[l, "\sim"'] \arrow[d, "-\circ !"] \\
\mathrm{Hom}_{\mathbf{C}\times\mathbf{C}}((A, A), (X, Y)) & \mathrm{Hom}_{\mathbf{C}}(A, P) \arrow[l, "\sim"]
\end{tikzcd} \tag{172}$$

□

Corollary G.20 (Dual). *Let $X, Y \in \mathbf{C}_0$. The coproduct of X and Y exists if and only if there exists $S \in \mathbf{C}_0$ such that $\mathrm{Hom}_{\mathbf{C}\times\mathbf{C}}((X, Y), \Delta_{\mathbf{C}}(-)) \cong \mathrm{Hom}_{\mathbf{C}}(S, -)$. The coproduct is S.*[457]

[457] We implicitly use the fact that $(\mathbf{C} \times \mathbf{C})^{\mathrm{op}} \cong \mathbf{C}^{\mathrm{op}} \times \mathbf{C}^{\mathrm{op}}$.

In order to generalize these two results to arbitrary (co)limits, we define the generalized version of $\Delta_{\mathbf{C}}$.

Definition G.21 (Generalized diagonal functor)**.** Let $\mathbf{J}$ be a small category, the **generalized diagonal functor** $\Delta^{\mathbf{J}}_{\mathbf{C}} : \mathbf{C} \rightsquigarrow [\mathbf{J}, \mathbf{C}]$ sends an object $X \in \mathbf{C}_0$ to the constant functor at X and a morphism $f : X \to Y \in \mathbf{C}_1$ to the natural transformation whose components are all $f : X \to Y$.

We have $\Delta^{\mathbf{J}}_{\mathbf{C}}(f) : X \Rightarrow Y$ because for any $a \in \mathbf{J}_1$, the square below commutes.

$$\begin{tikzcd}
X \arrow[r, "X(a)=\mathrm{id}_X"] \arrow[d, "f"'] & X \arrow[d, "f"] \\
Y \arrow[r, "Y(a)=\mathrm{id}_Y"'] & Y
\end{tikzcd}$$

Remark G.22. This is a generalization of the diagonal functor $\Delta_{\mathbf{C}} : \mathbf{C} \rightsquigarrow \mathbf{C} \times \mathbf{C}$ because, with the isomorphism $[\mathbf{1}+\mathbf{1}, \mathbf{C}] \cong \mathbf{C} \times \mathbf{C}$ described in Example F.14.2, we can identify $\Delta_{\mathbf{C}}$ with $\Delta^{\mathbf{1}+\mathbf{1}}_{\mathbf{C}}$.

Proposition G.23. *Let $F : \mathbf{J} \rightsquigarrow \mathbf{C}$ be a diagram. The limit of F exists if and only if there is an object $L \in \mathbf{C}_0$ such that* $\mathrm{Nat}(\Delta^{\mathbf{J}}_{\mathbf{C}}(-), F) \cong \mathrm{Hom}_{\mathbf{C}}(-, L)$.[458] *The tip of the limit cone is L.*

[458] Recall that

$$\mathrm{Nat}(\Delta^{\mathbf{J}}_{\mathbf{C}}(-), F) = \mathrm{Nat}(-, F) \circ \Delta^{\mathbf{J}}_{\mathbf{C}}.$$

For this to be a functor $\mathbf{C}^{\mathrm{op}} \rightsquigarrow \mathbf{Set}$, it is important that $\mathbf{J}$ is small and $\mathbf{C}$ is locally small as it guarantees the functor category $[\mathbf{J}, \mathbf{C}]$ to be locally small too, hence $\mathrm{Nat}(\Delta^{\mathbf{J}}_{\mathbf{C}}(X), F)$ is a set for any $X \in \mathbf{C}_0$.

Proof. First, we note that for any $X \in \mathbf{C}_0$, a natural transformation $\psi : \Delta^{\mathbf{J}}_{\mathbf{C}}(X) \Rightarrow F$ is a cone over F with tip X. Indeed, for any $a : A \to B \in \mathbf{J}_1$, the naturality square in (173) is commutative.

$$\begin{tikzcd}
X \arrow[r, "X(a)=\mathrm{id}_X"] \arrow[d, "\psi_A"'] & X \arrow[d, "\psi_B"] \\
FA \arrow[r, "F(a)"'] & FB
\end{tikzcd} \tag{173}$$

This is equivalent to $\{\psi_A : X \to FA\}_{A \in \mathbf{J}_0}$ being a cone over F. Furthermore, a morphism of cones $\phi \to \psi$ is a morphism f between the tips such that $\forall A \in \mathbf{J}_0, \phi_A = \psi_A \circ f$. By looking at (174), we see this condition is equivalent to $\phi = \psi \cdot \Delta^{\mathbf{J}}_{\mathbf{C}}(f)$.

$$\begin{tikzcd}
& Y \arrow[rr, "\mathrm{id}_Y"] \arrow[dl, "f"'] \arrow[dd, "\phi_A" near end] & & Y \arrow[dl, "f"'] \arrow[dd, "\phi_B"] \\
X \arrow[rr, "\mathrm{id}_X" near start] \arrow[dr, "\psi_A"'] & & X \arrow[dr, "\psi_B"'] & \\
& FA \arrow[rr, "F(a)"'] & & FB
\end{tikzcd} \tag{174}$$

($\Rightarrow$) Let $\{\psi_A : L \to FA\}_{A \in \mathbf{J}_0}$ be the terminal cone over F (i.e. the limit) and see it as a natural transformation $\psi : \Delta^{\mathbf{J}}_{\mathbf{C}}(L) \Rightarrow F$. We need to define a natural isomorphism $\mathrm{Nat}(\Delta^{\mathbf{J}}_{\mathbf{C}}(-), F) \cong \mathrm{Hom}_{\mathbf{C}}(-, L)$. Similarly to the proofs of the previous section, we will see that we only need to see where id_L is sent to and the rest of the natural transformation will *construct itself*. Our only choice for the cone corresponding to id_L is ψ (it is the only cone we know exists).

$$\begin{tikzcd}
\mathrm{Nat}(\Delta^{\mathbf{J}}_{\mathbf{C}}(L), F) \arrow[r, leftrightarrow] \arrow[d, "- \cdot \Delta^{\mathbf{J}}_{\mathbf{C}}(f)"'] & \mathrm{Hom}_{\mathbf{C}}(L, L) \arrow[d, "- \circ f"] \\
\mathrm{Nat}(\Delta^{\mathbf{J}}_{\mathbf{C}}(X), F) \arrow[r, leftrightarrow] & \mathrm{Hom}_{\mathbf{C}}(X, L)
\end{tikzcd} \tag{175}$$

Indeed, for any $f : X \to L$ the naturality square in (175) means the cone corresponding to $f : X \to L$ is $\{\psi_A \circ f : X \to FA\}_{A \in \mathbf{J}_0}$ by starting with id_L in the top right. Now, since ψ is the terminal cone, for any cone $\{\phi_A : X \to FA\}_{A \in \mathbf{J}_0}$, there is a unique morphism of cones $f : X \to L$ which satisfies $\forall A \in \mathbf{J}_0, \psi_A \circ f = \phi_A$. We conclude that $f \mapsto \psi \cdot \Delta^{\mathbf{J}}_{\mathbf{C}}(f)$ is a natural isomorphism.

($\Leftarrow$) Let $\psi : \Delta^{\mathbf{J}}_{\mathbf{C}}(L) \Rightarrow F$ be the cone corresponding to $\mathrm{id}_L \in \mathrm{Hom}_{\mathbf{C}}(L, L)$ under the natural isomorphism, we will show it is terminal. By the commutativity of (175) and bijectivity of the horizontal arrows, for any cone $\phi : \Delta^{\mathbf{J}}_{\mathbf{C}}(X) \Rightarrow F$, there is a unique morphism $f : X \to L$ such that $\phi = \psi \cdot \Delta^{\mathbf{J}}_{\mathbf{C}}(f)$. By the first paragraph of the proof, this is the unique morphism of cones showing ψ is terminal. □

Corollary G.24 (Dual)**.** *Let $F : \mathbf{J} \rightsquigarrow \mathbf{C}$ be a diagram. The colimit of F exists if and only if there is an object $L \in \mathbf{C}_0$ such that* $\mathrm{Nat}(F, \Delta^{\mathbf{J}}_{\mathbf{C}}(-)) \cong \mathrm{Hom}_{\mathbf{C}}(L, -)$. *The tip of the colimit cone is L.*

Proposition G.25. *Let $U : \mathbf{Mon} \rightsquigarrow \mathbf{Set}$ be the forgetful functor, A be a set and A^* be the free monoid on A, we have* $\mathrm{Hom}_{\mathbf{Set}}(A, U-) \cong \mathrm{Hom}_{\mathbf{Mon}}(A^*, -)$.

Proof. We have already shown before Definition E.2 that sending $h : A \to M$ to $h^* : A^* \to M$ is a bijection of the desired type.[459] Now, we need to show it is natural in M. For any monoid homomorphism $f : M \to N$, (176) commutes (we omitted applications of U) because starting with $h : A \to M$, we have $(f \circ h)^* = f \circ h^*$.[460]

$$\begin{tikzcd} \mathrm{Hom}_{\mathbf{Set}}(A, M) \arrow[r, "\sim"] \arrow[d, "f\circ-"'] & \mathrm{Hom}_{\mathbf{Mon}}(A^*, M) \arrow[d, "f\circ-"] \\ \mathrm{Hom}_{\mathbf{Set}}(A, N) \arrow[r, "\sim"'] & \mathrm{Hom}_{\mathbf{Mon}}(A^*, N) \end{tikzcd} \tag{176}$$

□

[459] In the other direction, $h : A^* \to M$ is sent to $U(h) \circ i$ where $i : A \hookrightarrow A^*$ is the inclusion.

[460] To check this, let $w = a_1 \cdots a_n \in A^*$, we have

$$\begin{aligned} (f \circ h)^*(w) &= fh(a_1) \cdots fh(a_n) \\ &= f(h(a_1) \cdots h(a_n)) \\ &= f(h^*(w)). \end{aligned}$$

In the next Proposition, we will generalize this result to see how any universal morphism corresponds to some kind of representability and we will even give a converse direction. The generalizations of the proof is straightforward, so I suggest you try to get familiar with a specific case in the next exercise.

OL **Exercise G.26.** Let $\mathbf{C}$ be a category and $X \in \mathbf{C}_0$ be such that $- \times X$ is a functor. An object $A \in \mathbf{C}_0$ has an exponential $A^X \in \mathbf{C}_0$ if and only if $\mathrm{Hom}_{\mathbf{C}}(- \times X, A) \cong \mathrm{Hom}_{\mathbf{C}}(-, A^X)$.

Proposition G.27. *Let $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ be a functor and $X \in \mathbf{D}_0$. There is a universal morphism from X to F if and only if there exists $A \in \mathbf{C}_0$ such that $\mathrm{Hom}_{\mathbf{D}}(X, F-) \cong \mathrm{Hom}_{\mathbf{C}}(A, -)$.*

Proof. ($\Rightarrow$) Let $a : X \to FA$ be a universal morphism, by definition, for any $b : X \to FB$, there is a unique morphism $\phi_B(b) : A \to B$ such that $F(\phi_B(b)) \circ a = b$. In the other direction, ϕ_B^{-1} sending $f : A \to B$ to $Ff \circ a$ is the inverse of ϕ_B.[461] Let us now check that ϕ_B is natural. For any $m : B \to B'$, (177) commutes because when starting with $f : A \to B$ in the top right, the top path sends it to $Ff \circ a$ then to $Fm \circ Ff \circ a$ and the bottom path sends it to $m \circ f$ then to $F(m \circ f) \circ a$.

[461] We check they are inverses:

$$\begin{aligned} \phi_B^{-1}(\phi_B(b)) &= F(\phi_B(b)) \circ a = b \\ \phi_B(\phi_B^{-1}(f)) &= \phi_B(Ff \circ a) = f. \end{aligned}$$

$$\begin{tikzcd} \mathrm{Hom}_{\mathbf{C}}(X, FB) \arrow[d, "Fm\circ-"'] & \mathrm{Hom}_{\mathbf{D}}(A, B) \arrow[l, "\sim"'] \arrow[d, "m\circ-"] \\ \mathrm{Hom}_{\mathbf{C}}(X, FB') & \mathrm{Hom}_{\mathbf{D}}(A, B') \arrow[l, "\sim"] \end{tikzcd} \tag{177}$$

($\Leftarrow$) Let $a : X \to FA$ be the image of $\mathrm{id}_A : A \to A$ under the isomorphism $\mathrm{Hom}_{\mathbf{C}}(X, FA) \cong \mathrm{Hom}_{\mathbf{D}}(A, A)$, we claim that a is a universal morphism from X to F. Given $b : X \to FB$, let $\phi_B(b)$ be its image under the isomorphism $\mathrm{Hom}_{\mathbf{C}}(X, FB) \cong \mathrm{Hom}_{\mathbf{D}}(A, B)$, it satisfies $F(\phi_B(b)) \circ a = b$ because (178) commutes (start with id_A in the top right corner). The morphism $\phi_B(b)$ is unique with this property because any other $f : A \to B$ is the image of some $b' \neq b$ under ϕ_B yielding $Ff \circ a = b' \neq b$.

$$\begin{tikzcd} \mathrm{Hom}_{\mathbf{C}}(X, FA) \arrow[d, "F(\phi_B(b))\circ-"'] & \mathrm{Hom}_{\mathbf{D}}(A, A) \arrow[l, "\sim"'] \arrow[d, "\phi_B(b)\circ-"] \\ \mathrm{Hom}_{\mathbf{C}}(X, FB) & \mathrm{Hom}_{\mathbf{D}}(A, B) \arrow[l, "\sim"] \end{tikzcd} \tag{178}$$

□

Corollary G.28 (Dual)**.** *Let $F : \mathbf{C} \rightsquigarrow \mathbf{D}$ be a functor and $X \in \mathbf{D}_0$. There is a universal morphism from F to X if and only if there exists $A \in \mathbf{C}_0$ such that $\mathrm{Hom}_{\mathbf{D}}(F-, X) \cong \mathrm{Hom}_{\mathbf{C}}(-, A)$.*

Comparing Propositions G.23 and G.27 and their duals, we infer that (co)limits satisfy universal properties.

Theorem G.29. *Let $F \in [\mathbf{J}, \mathbf{C}]_0$ be a diagram.*

- *The limit of F exists if and only if there is a universal morphism from $\Delta^{\mathbf{J}}_{\mathbf{C}}$ to F.*
- *The colimit of F exists if and only if there is a universal morphism from F to $\Delta^{\mathbf{J}}_{\mathbf{C}}$.*

In the next chapter, we will lift these correspondence to a more global version. Namely, we will see how to assemble the universal morphisms for all diagrams of shape $\mathbf{J}$ (if they all exist) into something called a right adjoint to $\Delta^{\mathbf{J}}_{\mathbf{C}}$.

H Adjunctions

Remark H.1. Adjunctions are very much everywhere in mathematics (once you learn to recognize them), and this inevitably means there are many angles to approach a first understanding. We will only get to see my favorite here, it can be roughly summarized in "adjunctions are global universal constructions", but of course I suggest you visit other resources to round out your intuitions.[462]

[462] I think feeling comfortable with adjunctions is a good signal that you are done with your journey in so-called basic category theory, and you are ready for the harder stuff (or you can apply basic category theory to other stuff).

In Chapter E on universal properties, we gave categorical descriptions of important constructions in mathematics. We defined the free monoid *on a set*, the abelianization *of a group*, and the exponential *of a set* by another one. The given set (resp. group) on which the constructions are applied is part of the definitions we gave, but we know that they can be applied to any set (resp. group). Therefore, one might ask if it is possible to define (categorically) the construction as a whole. For instance, the action of taking free monoids sends a set to a monoid, so it could be the action on objects of a functor from **Set** to **Mon**.

We start by explainig how this functor arises simply from the existence of free monoids on every set.[463] More abstractly, we show that having an object FX with a universal property based on X for every X means that F is a functor. Moreover, we will see that F is closely related to the functor used in the universal property. This relation is what we call an adjunction. The rest of the chapter will be dedicated to learning more about adjunctions through examples and properties.

[463] We spend a lot of time on this example, so you might want to revisit your understanding of free monoids before moving on.

H.1 Equivalent Definitions

There are four very commonly used defintions of an adjunction.[464] We will start from the one that is most directly linked to the concrete setting of free monoids, and then develop the details (in the abstract setting) to get the other definitions. Finally, we will prove the equivalence between the definitions.

[464] Morally only three because one is dual to another.

Let us have two categories $\mathbf{C}$ and $\mathbf{D}$ and a functor $R : \mathbf{D} \rightsquigarrow \mathbf{C}$.[465] Suppose that for any $X \in \mathbf{C}_0$, we have a universal morphism from X to R, namely, we have an object $LX \in \mathbf{D}_0$ and a morphism $\eta_X : X \to RLX$ satisfying a universal property as in Definition E.29 and summarized below.[466]

[465] In our concrete running example, $\mathbf{C} = \mathbf{Set}$, $\mathbf{D} = \mathbf{Mon}$ and R is the forgetful functor.

[466] For free monoids, LX is the free monoid on X, i.e. X^*, and η_X is the inclusion of X inside X^* (R only forgets the monoid structure).

$$\begin{tikzcd}[row sep=small]
\text{in } \mathbf{C} & & & \text{in } \mathbf{D} \\
X \arrow[r, "\eta_X"] \arrow[ddr, "h"'] & RLX \arrow[dd, dashed, "R!"] & & LX \arrow[dd, dashed, "!"] \\
& {} & {} \arrow[l, "R"'] & \\
& RA & & A
\end{tikzcd} \tag{179}$$

We first show that the action $X \mapsto LX$ is functorial (yielding a functor $L : \mathbf{C} \rightsquigarrow \mathbf{D}$). For any $f : X \to Y$, the universality of η_X yields a unique morphism $Lf : LX \to LY$ satisfying $RLf \circ \eta_X = \eta_Y \circ f$ as summarized in (180).[467]

$$\begin{tikzcd}[row sep=small]
\text{in } \mathbf{C} & & & \text{in } \mathbf{D} \\
X \arrow[r, "\eta_X"] \arrow[dd, "f"'] \arrow[ddr, "\eta_Y \circ f"] & RLX \arrow[dd, dashed, "RLf"] & & LX \arrow[dd, dashed, "Lf"] \\
& {} & {} \arrow[l, "R"'] & \\
Y \arrow[r, "\eta_Y"'] & RLY & & LY
\end{tikzcd} \tag{180}$$

The functoriality follows from the following equations showing that $L(\mathrm{id}_X) = \mathrm{id}_{LX}$ and $L(g \circ f) = Lg \circ Lf$ because these morphisms make the relevant diagrams commute:[468]

$$R(\mathrm{id}_{LX}) \circ \eta_X = \mathrm{id}_{RLX} \circ \eta_X = \eta_X = \eta_X \circ \mathrm{id}_X$$
$$R(Lg \circ Lf) \circ \eta_X = RLg \circ RLf \circ \eta_X = RLg \circ \eta_Y \circ f = \eta_Z \circ (g \circ f).$$

Note that the definition of L on morphisms readily gives us that η is a natural transformation $\mathrm{id}_{\mathbf{C}} \Rightarrow RL$. The functor L constructed like that is called the left adjoint to R.[469]

[467] For free monoids, $Lf : X^* \to Y^*$ is the homomorphism defined inductively by $Lf(\varepsilon) = \varepsilon$ and $Lf(w \cdot x) = Lf(w) \cdot f(x)$. Concretely, it applies f to every letter of the word.

[468] The equations respectively show that id_{LX} makes (179) commute when h is replaced by id_X and $Lg \circ Lf$ does it when h is replaced by $g \circ f$.

[469] For free monoids, L is the free monoid functor $\mathbf{Mon} \rightsquigarrow \mathbf{Set}$ sending X to X^* and it is the left adjoint to the forgetful functor $\mathbf{Mon} \rightsquigarrow \mathbf{Set}$.

Definition H.2 (Left adjoint)**.** Let $R : \mathbf{D} \rightsquigarrow \mathbf{C}$ be a functor. A functor $L : \mathbf{C} \rightsquigarrow \mathbf{D}$ is called the **left adjoint** to R if there exists a natural transformation $\eta : \mathrm{id}_{\mathbf{C}} \Rightarrow RL$ such that for every X, $\eta_X : X \to RLX$ is a universal morphism from X to R, equivalently, η_X is initial in $\Delta(X) \downarrow R$.

Following the construction of L with another family of universal morphisms to R would yield another left adjoint. Thus, to justify the use of the definite article *the*, we can prove that the two left adjoints would be naturally isomorphic.

Proposition H.3. *Let $R : \mathbf{D} \rightsquigarrow \mathbf{C}$ be a functor, and $L, L' : \mathbf{C} \rightsquigarrow \mathbf{D}$ be two left adjoints to R. Then, $L \cong L'$.*

Proof. Let $\eta : \mathrm{id}_{\mathbf{C}} \Rightarrow RL$ and $\eta' : \mathrm{id}_{\mathbf{C}} \Rightarrow RL'$ be the natural transformations witnessing L and L' respectively as left adjoints to R. For any X, since both $\eta_X : X \to RLX$ and $\eta'_X : X \to RL'X$ are initial in $\Delta(X) \downarrow R$, they must be isomorphic inside this comma category. This means there is an (unique) isomorphism $\phi_X : LX \to LX'$ making (181) commute. It is an isomorphism in $\Delta(X) \downarrow R$, but we find it is also an isomorphism in $\mathbf{D}$ by applying the forgetful functor $U_R : \Delta(X) \downarrow R \rightsquigarrow \mathbf{D}$ from Exercise E.34 (recall Exercise C.51.4).

$$\begin{tikzcd}
X \arrow[r, "\eta_X"] \arrow[dr, "\eta'_X"'] & RLX \arrow[d, "R\phi_X"] \\
& RL'X
\end{tikzcd} \tag{181}$$

It remains to show these components assemble into a natural transformation, i.e. that for any $f : X \to Y$, $L'f \circ \phi_X = \phi_Y \circ Lf$. We start by drawing the following two commutative diagrams.

$$
\begin{tikzcd}
X \arrow[r, "\eta_X"] \arrow[dd, "f"'] \arrow[dr, phantom, "\text{(a)}"] & RLX \arrow[d, "RLf"] \\
& RLY \arrow[d, "R\phi_Y"] \\
Y \arrow[r, "\eta'_Y"'] \arrow[ur, "\eta_Y"] \arrow[ur, phantom, "\text{(b)}"', shift right=3] & RL'Y
\end{tikzcd}
\qquad
\begin{tikzcd}
X \arrow[r, "\eta_X"] \arrow[dd, "f"'] \arrow[dr, "\eta'_X"'] \arrow[dr, phantom, "\text{(c)}", shift left=3] & RLX \arrow[d, "R\phi_X"] \\
& RL'X \arrow[d, "RL'f"] \\
Y \arrow[r, "\eta'_Y"'] \arrow[ur, phantom, "\text{(d)}"] & RL'Y
\end{tikzcd}
\tag{182}
$$

Showing (182) commutes:
(a) $\mathsf{NAT}(\eta, X, Y, f)$.
(b) Definition of ϕ (181).
(c) Definition of ϕ (181).
(d) $\mathsf{NAT}(\eta', X, Y, f)$.

We find that both $\phi_Y \circ Lf$ and $L'f \circ \phi_X$ make (179) commute when h is replaced by $\eta'_Y \circ f$. Thus, by uniqueness, they must be equal. We conclude that ϕ is a natural isomorphism $L \Rightarrow L'$. □

The dual concept is called a right adjoint.

Definition H.4 (Right adjoint)**.** Let $L : \mathbf{C} \rightsquigarrow \mathbf{D}$ be a functor. A functor $R : \mathbf{D} \rightsquigarrow \mathbf{C}$ is called the right adjoint to L is there exists a natural transformation $\varepsilon : LR \Rightarrow \mathrm{id}_{\mathbf{D}}$ such that for every X, $\varepsilon_X : LRX \to X$ is a universal morphism from L to X, equivalently, ε_X is terminal in $L \downarrow \Delta(X)$.

Corollary H.5 (Dual)**.** *If $R, R' : \mathbf{D} \rightsquigarrow \mathbf{C}$ are two right adjoints to $L : \mathbf{C} \rightsquigarrow \mathbf{D}$, then $R \cong R'$.*

Example H.6 (Cartesian closedness)**.** Let $\mathbf{C}$ be a category with all finite products (in particular, binary ones and a terminal object). Given two objects $A, X \in \mathbf{C}_0$, recall that their exponential exists if and only if there is a universal morphism $\mathsf{ev} : A^X \times X \to A$ from $- \times X$ to A.

Fixing X, if this exponential exists for every $A \in \mathbf{C}_0$, then a dual argument to the one preceding Definition H.2 shows that the assignment $A \mapsto A^X$ yields a functor $\mathbf{C} \rightsquigarrow \mathbf{C}$ that is right adjoint to $- \times X : \mathbf{C} \rightsquigarrow \mathbf{C}$ from Exercise E.11, and moreover the evaluation morphisms are components of a natural transformation $(-)^X \times X \Rightarrow \mathrm{id}_{\mathbf{C}}$. By Definition E.15, $\mathbf{C}$ is cartesian closed precisely when all functors $- \times X$ have a right adjoint.

Example H.7 (Free monoids)**.** We saw that the free monoid functor $(-)^* : \mathbf{Set} \rightsquigarrow \mathbf{Mon}$ is left adjoint to the forgetful functor $U : \mathbf{Mon} \rightsquigarrow \mathbf{Set}$. We can also show that U is right adjoint to $(-)^*$. For any monoid $M \in \mathbf{Mon}_0$, we need to define a monoid homomorphism $UM^* \to M$. Since an element $w \in UM^*$ is a word whose letters are elements of M, we can multiply all the letters together with the monoid operation (the order does not matter thanks to associativity) to get one element of M. We call this function $c : UM^* \to M$, and the fact that it is a homomorphism also follows from associativity.

Now, for any set A and homomorphism $h : A^* \to M$, we know that the action of h is completely determined by where it sends the single-letter words.[470] More precisely, we know that if $w = \mathtt{a}_1 \cdots \mathtt{a}_n$ is a word in A^*, then $h(w) = h(\mathtt{a}_1) \cdots h(\mathtt{a}_n)$, where $\cdots$ denotes here the multiplciation in M. If we instead see $h(\mathtt{a}_1) \cdots h(\mathtt{a}_n)$ as a word in UM^*, i.e. $\cdots$ denotes concatenation of letters, it can be obtained by applying

[470] You can see this as a consequence of either the classical Definition E.1 or the categorical Definition E.2 of free monoids.

the restriction of h to A to every letter in w, i.e. $h(\mathsf{a}_1)\cdots h(\mathsf{a}_n) = h|_A{}^*(\mathsf{a}_1\cdots\mathsf{a}_n) = h|_A{}^*(w)$. This lets us see that $h|_A : A \to UM$ is the unique function satisfying $c(h|_A{}^*) = h$, and we conclude that c satisfies the appropriate universal property summarized in (183).

$$\begin{tikzcd} \text{in } \mathbf{Mon} & & & \text{in } \mathbf{Set} \\ M & UM^* \arrow[l, "c"'] & & UM \\ & A^* \arrow[ul, "h"] \arrow[u, dashed, "h|_A{}^*"'] & & A \arrow[u, dashed, "g"'] \arrow[l, phantom, ""] \end{tikzcd} \quad \xleftarrow{(-)^*} \qquad (183)$$

As for exponentials, we find that U is right adjoint to $(-)^*$.

In our running example, we now have a pair of functors ($(-)^*$ and U) adjoint to each other, one left adjoint and the other right adjoint. It turns out we can develop Example H.7 abstractly and show that when L is left adjoint to R, then R is right adjoint to L, and vice-versa by duality.

Proposition H.8. *Let $L : \mathbf{C} \rightsquigarrow \mathbf{D}$ and $R : \mathbf{D} \rightsquigarrow \mathbf{C}$ be two functors. If L is left adjoint to R, then R is right adjoint to L.*

Proof. Let $\eta : \mathrm{id}_{\mathbf{C}} \Rightarrow RL$ be the natural transformation witnessing L as left adjoint to R. We first define the components of a natural transformation $\varepsilon : LR \Rightarrow \mathrm{id}_{\mathbf{D}}$. For $X \in \mathbf{D}_0$, we need a morphism $LRX \to X$ in $\mathbf{D}$, and we know from the universal property of η_{RX} that it is enough to find a morphism $RX \to RX$. Of course we take the identity, and we let ε_X be the unique morphism given by the universality of η_{RX} such that $R(\varepsilon_X) \circ \eta_{RX} = \mathrm{id}_{RX}$ (see (184)).

$$\begin{tikzcd} \text{in } \mathbf{C} & & \text{in } \mathbf{D} \\ RX \arrow[r, "\eta_{RX}"] \arrow[dr, "\mathrm{id}_{RX}"'] & RLRX \arrow[d, dashed, "R\varepsilon_X"] & LRX \arrow[d, dashed, "\varepsilon_X"] \\ & RX & X \end{tikzcd} \quad \xleftarrow{R} \qquad (184)$$

Next, we show that $\varepsilon_X : LRX \to X$ is a universal morphism from L to X. For any $f : LA \to X$, if $g : A \to RX \in \mathbf{C}_1$ is such that $f = \varepsilon_X \circ Lg$, then applying R and pre-composing with η_A, we obtain

$$\begin{aligned} Rf \circ \eta_A &= R\varepsilon_X \circ RLg \circ \eta_A \\ &= R\varepsilon_X \circ \eta_{RX} \circ g && \mathsf{NAT}(\eta, A, RX, g) \\ &= \mathrm{id}_{RX} \circ g && \text{definition of } \varepsilon_X \\ &= g. \end{aligned}$$

We conclude that $g := Rf \circ \eta_A$ is the unique morphism satisfying that $f = \varepsilon_X \circ Lg$, hence ε_X is universal.

Finally, we show that $\varepsilon : LR \Rightarrow \mathrm{id}_{\mathbf{D}}$ is natural. For any $f : X \to Y \in \mathbf{D}_1$, by universality, there is a unique morphism $g : RX \to RY$ such that $f \circ \varepsilon_X = \varepsilon_Y \circ Lg$ (see (185)) and by our derivation above, $g = Rf \circ R\varepsilon_X \circ \eta_{RX} \overset{(184)}{=} Rf$. Thus, we find that $f \circ \varepsilon_X = \varepsilon_Y \circ LRf$, namely ε is natural. ☐

$$\begin{tikzcd} \text{in } \mathbf{D} & & \text{in } \mathbf{C} \\ Y & LRY \arrow[l, "\varepsilon_Y"'] & RY \\ X \arrow[u, "f"] & LRX \arrow[l, "\varepsilon_X"] \arrow[u, dashed, "Lg"'] \arrow[ul, "\varepsilon_X \circ f"] & RX \arrow[u, dashed, "g"'] \end{tikzcd} \quad \xleftarrow{L} \qquad (185)$$

As a sanity check, notice that using the definition of ε_M in the case of free monoids, we get back the homomorphism c from Example H.7. Indeed, instantiating (184), we find $\varepsilon_M : UM^* \to M$ is the unique homomorphism that acts like identity on single-letter words M (recall η_{UM} sends $x \in UM$ to the word $x \in UM^*$). It is easy to check c also acts like identity on single-letter words, so ε_M and c coincide by uniqueness.

Corollary H.9 (Dual). *If R is right adjoint to L, then L is left adjoint to R.*

This makes Definitions H.2 and H.4 a bit unsatisfactory because they seem to focus on one side of relation between two functors. To resolve this, we bring up two important properties that arise from having a left and right adjoint, and we will see these also characterize adjoints.

First, we note that $\eta : \mathrm{id}_{\mathbf{C}} \Rightarrow RL$ and $\varepsilon : LR \Rightarrow \mathrm{id}_{\mathbf{D}}$ seem to have the right type to give rise to an equivalence between $\mathbf{C}$ and $\mathbf{D}$. However, in general, nothing guarantees the components of η and ε are isomorphisms.[471] There is still some kind of invertibility property: η and ε satisfy the the **triangle identities** shown in (186) and (187) (they are commutative diagrams in $[\mathbf{C}, \mathbf{D}]$ and $[\mathbf{D}, \mathbf{C}]$ respectively).

[471] It is clearly not the case in the free monoids example.

$$\begin{tikzcd} L \arrow[r, "L\eta"] \arrow[dr, "\mathbb{1}_L"'] & LRL \arrow[d, "\varepsilon L"] \\ & L \end{tikzcd} \quad (186) \qquad \begin{tikzcd} RLR \arrow[d, "R\varepsilon"'] & R \arrow[l, "\eta R"'] \arrow[dl, "\mathbb{1}_R"] \\ R & \end{tikzcd} \quad (187)$$

The second one holds by definition of ε_X (for any $X \in \mathbf{D}_0$, $R\varepsilon_X \circ \eta_{RX} = \mathrm{id}_{RX}$). For the first one, by universality of ε_X, there is a unique morphism $g : X \to RLX$ such that $\varepsilon_{LX} \circ Lg = \mathrm{id}_{LX}$ (see (188)), and by our derivation in the previous proof, $g = R(\mathrm{id}_{LX}) \circ \eta_X = \eta_X$. We find that $\varepsilon_{LX} \circ L\eta_X = \mathrm{id}_{LX}$ as desired.

$$\begin{tikzcd} \text{in } \mathbf{D} & & & \text{in } \mathbf{C} \\ LX & LRLX \arrow[l, "\varepsilon_{LX}"'] & & RLX \\ & LX \arrow[ul, "\mathrm{id}_{LX}"] \arrow[u, dashed, "Lg"'] & & X \arrow[u, dashed, "g"'] \arrow[ll, phantom, "\xleftarrow{\ L\ }"] \end{tikzcd} \quad (188)$$

It is simple, but not very illuminating to see how these triangle identities hold in the free monoids example. Conversely, the next characterization of adjoints is in the spotlight of our running example. It abstractly states the slogan that it is the same thing to give a homomorphism out of the free monoid A^* or a function out of the set A.

Formally, we find a natural isomorphism[472]

$$\Phi : \mathrm{Hom}_{\mathbf{C}}(-, R-) \cong \mathrm{Hom}_{\mathbf{D}}(L-, -) : \Phi^{-1}.$$

[472] For free monoids, this gives

$$\mathrm{Hom}_{\mathbf{Set}}(A, M) \cong \mathrm{Hom}_{\mathbf{Mon}}(A^*, M),$$

which is inded what the slogan means.

For $g : X \to RY$, we define $\Phi_{X,Y}(g) = \varepsilon_Y \circ Lg$ and for $f : LX \to Y$, we define $\Phi^{-1}_{X,Y}(f) = Rf \circ \eta_X$.[473] The derivations below show these are inverses (and it only relies on the triangle identities and naturality):

[473] You can certainly infer these definitions just by looking at the types. Also note because it will be useful that $\Phi_{X,Y}(\mathrm{id}_{RX}) = \varepsilon_X$ and $\Phi^{-1}_{X,Y}(\mathrm{id}_{LX}) = \eta_X$.

$$\Phi^{-1}_{X,Y}(\Phi_{X,Y}(g)) = R\varepsilon_Y \circ RLg \circ \eta_X = R\varepsilon_Y \circ \eta_{RY} \circ g = g \quad (189)$$

$$\Phi_{X,Y}(\Phi^{-1}_{X,Y}(f)) = \varepsilon_Y \circ LRf \circ L\eta_X = f \circ \varepsilon_{LX} \circ L\eta_X = f. \quad (190)$$

To show that Φ is natural, we need to show that (191) commutes for any $x : X' \to X$ and $y : Y \to Y'$. Starting with $g : X \to RY$ in the top left, the bottom path sends it to $Ry \circ g \circ x$ then to $\varepsilon_{Y'} \circ LRy \circ Lg \circ Lx$ and the top path sends g to $\varepsilon_Y \circ Lg$ then to $y \circ \varepsilon_Y \circ Lg \circ Lx$. The end results are equal by $\mathsf{NAT}(\varepsilon, Y, Y', y)$.

$$\begin{tikzcd} \mathrm{Hom}_{\mathbf{C}}(X, RY) \arrow[r, leftrightarrow, "\Phi_{X,Y}"] \arrow[d, "Ry \circ - \circ x"'] & \mathrm{Hom}_{\mathbf{D}}(LX, Y) \arrow[d, "y \circ - \circ Lx"] \\ \mathrm{Hom}_{\mathbf{C}}(X', RY') \arrow[r, leftrightarrow, "\Phi_{X',Y'}"'] & \mathrm{Hom}_{\mathbf{D}}(LX', Y') \end{tikzcd} \quad (191)$$

We can now give an unbiased definition (not focused on one side) of adjunction.

Definition H.10 (Adjunction)**.** An **adjunction** between a functor $L : \mathbf{C} \rightsquigarrow \mathbf{D}$ and $R : \mathbf{D} \rightsquigarrow \mathbf{C}$ is the following data:

- A natural transformation $\eta : \mathrm{id}_{\mathbf{C}} \Rightarrow RL$ called the **unit** such that η_X is initial in $\Delta(X) \downarrow R$ for each $X \in \mathbf{C}_0$.

- A natural transformation $\varepsilon : LR \Rightarrow \mathrm{id}_{\mathbf{D}}$ called the **counit** such that ε_X is terminal in $L \downarrow \Delta(X)$ for each $X \in \mathbf{D}_0$.

- The unit η and counit ε satisfy the triangle identities.

- A natural isomorphism $\Phi : \mathrm{Hom}_{\mathbf{C}}(-, R-) \cong \mathrm{Hom}_{\mathbf{D}}(L-, -) : \Phi^{-1}$ such that $\Phi_{RX,X}(\mathrm{id}_{RX}) = \varepsilon_X$ and $\Phi^{-1}_{X,LX}(\mathrm{id}_{LX}) = \eta_X$.[474]

[474] It follows by naturality that $\Phi_{X,Y}(g) = \varepsilon_Y \circ Lg$ and $\Phi^{-1}_{X,Y}(f) = Rf \circ \eta_X$, as we had above.

We denote $\mathbf{C} : L \dashv R : \mathbf{D}$ when there is an adjunction between $L : \mathbf{C} \rightsquigarrow \mathbf{D}$ and $R : \mathbf{D} \rightsquigarrow \mathbf{C}$ and we call L the left adjoint and R the right adjoint, and we say L and R are adjoints.[475]

[475] When they are clear from the context or irrelevant, we omit the categories from the notation and write $L \dashv R$.

Example H.11 (Boring)**.** The identity functor on any category is self-adjoint: $\mathrm{id}_{\mathbf{C}} \dashv \mathrm{id}_{\mathbf{C}}$. Both the unit and counit are $\mathbb{1}_{\mathrm{id}_{\mathbf{C}}}$.[476]

[476] You can prove this easily but it also follows from Proposition H.19 and the fact that $\mathrm{id}_{\mathbf{C}}$ is its own inverse.

While we resolved the bias in our definitions of adjoints, it cost us brevity. The culminating point of this section is the proof that all this data is not necessary to define an adjunction, giving only one of the fours points is enough. In other words, Definition H.10 gives in fact four equivalent definitions of an adjunction.[477]

[477] There are still more equivalent definitions, but we have to limit ourselves to a finite list and we mentioned the parts of an adjunction that are most commonly used. One notable omission is that of adjunctions as Kan extensions.

Theorem H.12. *Two functors $L : \mathbf{C} \rightsquigarrow \mathbf{D}$ and $R : \mathbf{D} \rightsquigarrow \mathbf{C}$ are adjoints if at least one of the following holds.*

i. There is a natural transformation $\eta : \mathrm{id}_{\mathbf{C}} \Rightarrow RL$ such that η_X is initial in $\Delta(X) \downarrow R$ for each $X \in \mathbf{C}_0$.

ii. There is a natural transformation $\varepsilon : LR \Rightarrow \mathrm{id}_{\mathbf{D}}$ such that ε_X is terminal in $L \downarrow \Delta(X)$ for each $X \in \mathbf{D}_0$.

iii. There are two natural transformations $\eta : \mathrm{id}_{\mathbf{C}} \Rightarrow RL$ and $\varepsilon : LR \Rightarrow \mathrm{id}_{\mathbf{D}}$ that satisfy the triangle identities.[478]

[478] They satisfy

$$\varepsilon L \cdot L\eta = \mathbb{1}_L \qquad R\varepsilon \cdot \eta R = \mathbb{1}_R.$$

iv. There is a natural isomorphism $\Phi : \mathrm{Hom}_{\mathbf{C}}(-, R-) \cong \mathrm{Hom}_{\mathbf{D}}(L-, -) : \Phi^{-1}$.

Proof. We have already shown that (i.) gives rise to all the data of an adjunction at the start of the chapter.

For (ii.), we can use duality. Indeed, taking the dual of Definition H.10, we see that $\mathbf{C} : L \dashv R : \mathbf{D}$ if and only if $\mathbf{D}^{\mathrm{op}} : R^{\mathrm{op}} \dashv L^{\mathrm{op}} : \mathbf{C}^{\mathrm{op}}$ and η and ε swap their roles as unit and counit. Hence, from ε, we can derive an adjunction $R^{\mathrm{op}} \dashv L^{\mathrm{op}}$ as we did at the start of the chapter and duality yields $L \dashv R$.

For (iii.), it is enough to show the components of the unit η_X are initial in $\Delta(X) \downarrow R$ and use (i.).[479] Recall from (189) and (190) that for any $g : X \to RY \in \mathbf{C}_1$, there is a unique (because the components of Φ and Φ^{-1} are bijections) morphism $\Phi_{X,Y}(g) = \varepsilon_Y \circ Lg$ such that $R(\Phi_{X,Y}(g)) \circ \eta_X = \Phi^{-1}_{X,Y}(\Phi_{X,Y}(g)) = g$. Thus, η_X is a universal morphism as required.

[479] You can check that the triangle identities ensure that the adjunction constructed from (i.) will have ε as a counit.

For (iv.), we will construct a unit satisfying (i.). Fix $X \in \mathbf{C}_0$, we have a natural isomorphism $\Phi_{X,-} : \mathrm{Hom}_{\mathbf{C}}(X, R-) \cong \mathrm{Hom}_{\mathbf{D}}(LX, -)$. By Proposition G.27, there is a universal morphism $\eta_X : X \to RLX$ from X to R.[480] This yields a natural transformation $\eta : \mathrm{id}_{\mathbf{C}} \Rightarrow RL$ because for any $f : X \to Y$, the commutativity of (192) implies (by starting with id_{LX} and id_{LY} in the top left and top right corners

[480] From the proof of Proposition G.27, we recover $\eta_X = \Phi^{-1}_{X,LX}(\mathrm{id}_{LX})$.

respectively) $RLf \circ \eta_X = \Phi^{-1}_{X,LY}(Lf) = \eta_Y \circ f$.

$$\begin{tikzcd}
\mathrm{Hom}_{\mathbf{D}}(LX, LX) \arrow[r, "Lf\circ-"] \arrow[d, leftrightarrow, "\Phi_{X,LX}"'] & \mathrm{Hom}_{\mathbf{D}}(LX, LY) \arrow[d, leftrightarrow, "\Phi_{X,LY}"'] & \mathrm{Hom}_{\mathbf{D}}(LY, LY) \arrow[l, "-\circ Lf"'] \arrow[d, leftrightarrow, "\Phi_{Y,LY}"] \\
\mathrm{Hom}_{\mathbf{C}}(X, RLX) \arrow[r, "RLf\circ-"'] & \mathrm{Hom}_{\mathbf{C}}(X, RLY) & \mathrm{Hom}_{\mathbf{C}}(Y, RLY) \arrow[l, "-\circ f"]
\end{tikzcd} \tag{192}$$

You can check the natural isomorphism constructed with (i.) coincides with Φ. □

Each item of Theorem H.12 can be seen as a definition of adjunctions.[481] We would like to spend a bit more time on point (iv.) which is, in our opinion, the hardest definition to internalize and yet the easiest one to use in concrete contexts. The definition of an adjunction according to (iv.) can be stated as follows.

[481] That is how most textbooks present it.

Two functors $L : \mathbf{C} \rightsquigarrow \mathbf{D}$ and $R : \mathbf{D} \rightsquigarrow \mathbf{C}$ are adjoint if there is a natural isomorphism[482]

$$\mathrm{Hom}_{\mathbf{C}}(-, R-) \cong \mathrm{Hom}_{\mathbf{D}}(L-, -).$$

[482] We use Remark C.10 to define

$$\mathrm{Hom}_{\mathbf{C}}(-, R-) := \mathrm{Hom}_{\mathbf{C}}(-,-) \circ (\mathrm{id}_{\mathbf{C}^{\mathrm{op}}} \times R)$$
$$\mathrm{Hom}_{\mathbf{D}}(L-, -) := \mathrm{Hom}_{\mathbf{D}}(-,-) \circ (L^{\mathrm{op}} \times \mathrm{id}_{\mathbf{D}})$$

Less concisely, for any $X \in \mathbf{C}_0$ and $Y \in \mathbf{D}_0$, there is an isomorphism $\Phi_{X,Y}$: $\mathrm{Hom}_{\mathbf{C}}(X, RY) \cong \mathrm{Hom}_{\mathbf{D}}(LX, Y)$ such that for any $f : X \to X' \in \mathbf{C}_1$ and $g : Y \to Y' \in \mathbf{D}_1$, (193) commutes. We split the naturality in two squares because we will often use one square on its own[483] as we did on both sides of (192).

[483] This is possible by Exercise F.7.

$$\begin{tikzcd}
\mathrm{Hom}_{\mathbf{C}}(X', RY) \arrow[r, "-\circ f"] \arrow[d, leftrightarrow, "\Phi_{X',Y}"'] & \mathrm{Hom}_{\mathbf{C}}(X, RY) \arrow[r, "Rg\circ-"] \arrow[d, leftrightarrow, "\Phi_{X,Y}"'] & \mathrm{Hom}_{\mathbf{C}}(X, RY') \arrow[d, leftrightarrow, "\Phi_{X,Y'}"] \\
\mathrm{Hom}_{\mathbf{D}}(LX', Y) \arrow[r, "-\circ Lf"'] & \mathrm{Hom}_{\mathbf{D}}(LX, Y) \arrow[r, "g\circ-"'] & \mathrm{Hom}_{\mathbf{D}}(LX, Y')
\end{tikzcd} \tag{193}$$

In a very informal sense, the bijections $\Phi_{X,Y}$ let us embed $\mathbf{C}$ in $\mathbf{D}$ and vice-versa in a compatible way, that is, morphisms between $X \in \mathbf{C}_0$ and $Y \in \mathbf{D}_0$ can be seen either by viewing X in $\mathbf{D}$ via L or viewing Y in $\mathbf{C}$ via R.[484]

[484] For the adjunction $\mathbf{Set} : (-)^* \dashv U : \mathbf{Mon}$, any set can be viewed as the monoid of words over it, and any monoid can be viewed as a set by forgetting the operation.

To make proofs go smoother, we will often use the superscript notation $(-)^{\mathrm{t}}$ to denote an application of a component of Φ or Φ^{-1}. That is, for any $X \in \mathbf{C}_0$ and $Y \in \mathbf{D}_0$, we have

$$(-)^{\mathrm{t}} : \mathrm{Hom}_{\mathbf{C}}(X, RY) \cong \mathrm{Hom}_{\mathbf{D}}(LX, Y) : (-)^{\mathrm{t}}.$$

We call f^{t} the **transpose** of f.[485]

[485] Unfortunately, the term *transpose* is probably inspired by matrix transposition, but I do not know of a technical way to realize one as an instance of the other. Some authors also write f^* or $f^\sharp$ for the transpose of f.

H.2 Results and Examples

There are a couple of very important results in this section (Theorem H.25 and Theorem H.30), but we will start slow.

We already proved in Proposition H.3 that two left adjoints to the same functor must be isomorphic.[486] That proof used the first definition of left adjoints we saw with a natural family of universal morphisms. Let us prove the same thing, but relying on our two new definitions instead.[487]

[486] With our new notation: if $L \dashv R$ and $L' \dashv R$, then $L \cong L'$, and dually if $L \dashv R$ and $L \dashv R'$, then $R \cong R'$.

[487] We omit the second item in Definition H.10 because it is dual to the proof we already gave.

Proof of Proposition H.3 via triangle identities. Let η and ε be the unit and counit of the adjunction $\mathbf{C} : L \dashv R : \mathbf{D}$, η' and ε' be those of $\mathbf{C} : L' \dashv R : \mathbf{D}$. Guided by the types, it is easy to compose the natural transformations we have to obtain two new natural transformations of type $L \Rightarrow L'$ and $L' \Rightarrow L$:

$$\phi = L \overset{L\eta'}{\Longrightarrow} LRL' \overset{\varepsilon L'}{\Longrightarrow} L' \quad \text{and} \quad \phi^{-1} = L' \overset{L'\eta}{\Longrightarrow} L'RL \overset{\varepsilon' L}{\Longrightarrow} L.$$

It remains to show ϕ^{-1} is the inverse of ϕ. We show $\phi^{-1} \circ \phi = \mathbb{1}_L$ by paving the following diagram (it lives in $[\mathbf{C}, \mathbf{D}]$).

$$\begin{tikzcd}
L \arrow[r, "L\eta'"] \arrow[rr, bend left, "\phi"] \arrow[dd, equal] & LRL' \arrow[r, "\varepsilon L'"] \arrow[dr, "LRL'\eta"'] & L' \arrow[r, "L'\eta"] \arrow[rr, bend left, "\phi^{-1}"] \arrow[d, phantom, "\text{(b)}"] & L'RL \arrow[r, "\varepsilon' L"] & L \arrow[dd, equal] \\
{} \arrow[r, phantom, "\text{(a)}"] & {} & LRL'RL \arrow[ur, "\varepsilon L'RL"'] \arrow[dr, "LR\varepsilon' L"] \arrow[d, phantom, "\text{(d)}"] & {} \arrow[r, phantom, "\text{(c)}"] & {} \\
L \arrow[r, "L\eta"] \arrow[drr, equal] & LRL \arrow[ur, "L\eta' RL"] \arrow[rr, equal] \arrow[dr, phantom, "\text{(e)}", near start] & {} & LRL \arrow[r, "\varepsilon L"] & L \arrow[dll, equal] \\
{} & {} & L & {} & {}
\end{tikzcd} \tag{194}$$

Showing (194) commutes:

(a) Apply $L(-)'$ to $\mathsf{HOR}(\eta', \eta)$.

(b) By $\mathsf{HOR}(\varepsilon L', \eta)$ or $\mathsf{HOR}(\varepsilon, L'\eta)$.

(c) Apply $(-)L$ to $\mathsf{HOR}(\varepsilon, \varepsilon')$.

(d) Apply $L(-)L$ to the triangle identity (187) instantiated for η' and ε'.

(e) Apply the triangle identity (186) for η and ε.

We leave you to show $\phi \circ \phi^{-1}$ by paving a similar diagram (where L, η and ε swap roles with L', η' and ε'). ☐

Proof of Proposition H.3 via transposes. For any $X \in \mathbf{C}_0$, we define $\phi_X : LX \to L'X$ to be the image of $\mathrm{id}_{L'X} \in \mathrm{Hom}_{\mathbf{D}}(L'X, L'X)$ under the composition of the natural isomorphisms

$$\mathrm{Hom}_{\mathbf{D}}(L'X, L'X) \cong \mathrm{Hom}_{\mathbf{C}}(X, RL'X) \cong \mathrm{Hom}_{\mathbf{D}}(LX, L'X).$$

Then, for any $f : X \to Y$, the naturality squares in (195) imply $L'f \circ \phi_X = \phi_Y \circ Lf$.[488]

$$\begin{tikzcd}
\mathrm{Hom}_{\mathbf{D}}(L'X, L'X) \arrow[r, "L'f\circ -"] \arrow[d, leftrightarrow] & \mathrm{Hom}_{\mathbf{D}}(L'X, L'Y) \arrow[d, leftrightarrow] & \mathrm{Hom}_{\mathbf{D}}(L'Y, L'Y) \arrow[l, "-\circ L'f"'] \arrow[d, leftrightarrow] \\
\mathrm{Hom}_{\mathbf{C}}(X, RL'X) \arrow[r, "RL'f\circ -"] \arrow[d, leftrightarrow] & \mathrm{Hom}_{\mathbf{C}}(X, RLY) \arrow[d, leftrightarrow] & \mathrm{Hom}_{\mathbf{C}}(Y, RLY) \arrow[l, "-\circ f"'] \arrow[d, leftrightarrow] \\
\mathrm{Hom}_{\mathbf{D}}(LX, L'X) \arrow[r, "L'f\circ -"'] & \mathrm{Hom}_{\mathbf{D}}(LX, L'Y) & \mathrm{Hom}_{\mathbf{D}}(LY, L'Y) \arrow[l, "-\circ Lf"]
\end{tikzcd} \tag{195}$$

[488] Start with $\mathrm{id}_{L'X}$ and $\mathrm{id}_{L'Y}$ at the top left and top right respectively and compare the results at the bottom middle.

We conclude that $\phi : L \Rightarrow L'$ is natural. With a symmetric argument, we construct $\phi^{-1} : L' \Rightarrow L$[489] and we check that they are inverses with (196) and (197).

[489] i.e.: ϕ_X^{-1} is the image of id_{LX} under

$\mathrm{Hom}_{\mathbf{D}}(LX, LX) \cong \mathrm{Hom}_{\mathbf{C}}(X, RLX) \cong \mathrm{Hom}_{\mathbf{D}}(L'X, LX)$.

$$\begin{tikzcd}
\mathrm{Hom}_{\mathbf{D}}(LX, LX) \arrow[r, "\phi_X\circ -"] \arrow[d, leftrightarrow] & \mathrm{Hom}_{\mathbf{D}}(LX, L'X) \arrow[d, leftrightarrow] \\
\mathrm{Hom}_{\mathbf{D}}(L'X, LX) \arrow[r, "\phi_X\circ -"'] & \mathrm{Hom}_{\mathbf{D}}(L'X, L'X)
\end{tikzcd} \tag{196}$$

$$\begin{array}{ccc} \mathrm{Hom}_{\mathbf{D}}(L'X, L'X) & \xrightarrow{\phi_X^{-1}\circ -} & \mathrm{Hom}_{\mathbf{D}}(L'X, LX) \\ \updownarrow & & \updownarrow \\ \mathrm{Hom}_{\mathbf{D}}(LX, L'X) & \xrightarrow[\phi_X^{-1}\circ -]{} & \mathrm{Hom}_{\mathbf{D}}(LX, LX) \end{array} \tag{197}$$

Starting with id_{LX} in the top left of (196) and reaching the top right, we find that the image of $\phi_X \circ \phi_X^{-1}$ under the isomorphism is ϕ_X which is the image of $\mathrm{id}_{L'X}$, thus $\phi_X \circ \phi_X^{-1} = \mathrm{id}_{L'X}$. We proceed with a symmetric argument for (197). □

Of the three different proofs of Proposition H.3, the second one using the triangle identities seems to be the quickest. You can judge for yourself which proof you prefer. In the rest of this chapter, we will see many examples of adjunctions and results about adjoint functors and try to have a balance between the different definitions we use.[490]

[490] We try to care about which definition is easiest to use.

We start with a converse to Proposition H.3. When L has a right adjoint R and R' is isomorphic to R, then R' is also right adjoint to L.

OL **Exercise H.13.** Show that if $\mathbf{C} : L \dashv R : \mathbf{D}$ is an adjunction and $R \cong R'$, then $L \dashv R'$. State the dual statement and prove it.

Our main point in the introduction to this chapter was that grouping universal morphisms together as we did into an adjunction yields a notion of *global* universal construction. In particular, we can characterize when a category has *all* (co)limits of shape $\mathbf{J}$.

Theorem H.14. *A category* $\mathbf{C}$ *has all limits of shape* $\mathbf{J}$ *if (and only if)*[491] *the functor* $\Delta^{\mathbf{J}}_{\mathbf{C}}$ *has a right adjoint.*

[491]

Proof. ($\Rightarrow$) For each diagram $F : \mathbf{J} \rightsquigarrow \mathbf{C}$, we pick (with the axiom of choice) a limit $\lim_{\mathbf{J}} F$ given by completeness and a universal morphism $\Delta^{\mathbf{J}}_{\mathbf{C}} \to F$ given by Theorem G.29. By our argument at the start of the chapter, we get an adjunction $\Delta^{\mathbf{J}}_{\mathbf{C}} \dashv \lim_{\mathbf{J}}$.

($\Leftarrow$) Suppose $\mathbf{C} : \Delta^{\mathbf{J}}_{\mathbf{C}} \dashv L : [\mathbf{J}, \mathbf{C}]$ with unit η and let $F : \mathbf{J} \rightsquigarrow \mathbf{C}$ be a diagram. By definiton, $\eta_F : \Delta^{\mathbf{J}}_{\mathbf{C}} L(F) \to F$ is a universal morphism from $\Delta^{\mathbf{J}}_{\mathbf{C}}$ to F. Thus, by Theorem G.29, $L(F)$ is the limit of F. □

Corollary H.15 (Dual)**.** *A category* $\mathbf{C}$ *has all colimits of shape* $\mathbf{J}$ *if and only if the functor* $\Delta^{\mathbf{J}}_{\mathbf{C}}$ *has a left adjoint.*

We saw how families of universal morphisms give rise to an adjunction, so we could make our examples from Chapter E into adjunctions. Here, we carry out a similar but new example.

Example H.16. Recall from Exercise D.28 the maybe functor $- + \mathbf{1}$. Denote $\mathbf{1} = \{*\}$ for the terminal object of $\mathbf{Set}$. We consider a very similar functor $- + \mathbf{1} : \mathbf{Set} \rightsquigarrow \mathbf{Set}_*$ sending a set X to $(X + \mathbf{1}, *)$ and $f : X \to Y$ to $f + \mathrm{id}_{\mathbf{1}} : X + \mathbf{1} \to Y + \mathbf{1}$. In the other direction, we have the forgetful functor $U : \mathbf{Set}_* \rightsquigarrow \mathbf{Set}$ that forgets about the distinguished element of a pointed set. We claim that $- + \mathbf{1} \dashv U$.

First, for every set X, we need to define $\eta_X : X \to U((X + \mathbf{1}, *)) = X + \mathbf{1}$. The only obvious choice is to let η_X be the inclusion of X in $X + \mathbf{1}$ and one can check it makes η into a natural transformation $\mathrm{id}_{\mathbf{Set}} \Rightarrow U(- + \mathbf{1})$.

Check η and ε are natural:

$$\begin{array}{ccc} X & \xrightarrow{\eta_X} & X + \mathbf{1} \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle f + \mathrm{id}_{\mathbf{1}}} \\ Y & \xrightarrow[\eta_Y]{} & Y + \mathbf{1} \end{array} \qquad \begin{array}{ccc} (X, x) & \xrightarrow{\varepsilon_{(X,x)}} & (X + \mathbf{1}, *) \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle f + \mathrm{id}_{\mathbf{1}}} \\ (Y, y) & \xrightarrow[\varepsilon_{(Y,y)}]{} & (Y + \mathbf{1}, *) \end{array}$$

Second, for every pointed set (X, x), we need to define $\varepsilon_{(X,x)} : (X + \mathbf{1}, *) \to (X, x)$. Again, there is one clear choice, i.e.: acting like the identity on X and sending $*$ to x, we will denote $\varepsilon_{(X,x)} = [\mathrm{id}_X, * \mapsto x]$.

Finally, after checking the triangle identities which we instantiate below,[492] we conclude that $- + \mathbf{1} \dashv U$.

$$\begin{tikzcd} (X+\mathbf{1},*) \arrow[r, "\eta_X+\mathrm{id}_{\mathbf{1}}"] \arrow[dr, "\mathrm{id}_{X+\mathbf{1}}"'] & ((X+\mathbf{1})+\mathbf{1},\star) \arrow[d, "{[\mathrm{id}_{X+\mathbf{1}},\star\mapsto *]}"] \\ & (X+\mathbf{1},*) \end{tikzcd} \tag{198}$$

$$\begin{tikzcd} X \arrow[r, "\eta_X"] \arrow[dr, "\mathrm{id}_X"'] & X+\mathbf{1} \arrow[d, "{[\mathrm{id}_X,*\mapsto x]}"] \\ & X \end{tikzcd} \tag{199}$$

A good exercise in categorical thinking is to generalize this example to an arbitrary category $\mathbf{C}$ with binary coproducts and a terminal object.[493]

[492] When dealing with a set $(X+\mathbf{1})+\mathbf{1}$, we will denote $*$ for the element of the inner $\mathbf{1}$ and $\star$ for the outer one.
In (199), $X = U(X,x)$.

[493] See ... for a solution.

Example H.17 (Top). Let $U : \mathbf{Top} \rightsquigarrow \mathbf{Set}$ be the forgetful functor sending a topological space to its underlying set. We will find a left and a right adjoint to U.

Left adjoint: Fix a topological space (X,τ) and a set Y. We need to find a topological space (LY,λ) so that continuous functions $(LY,\lambda) \to (X,\tau)$ are in correspondence with functions $Y \to X$. It turns out there is a trivial topology that we can put on Y that makes any function $f : Y \to X$ continuous, it is called the **discrete topology** and contains all the subsets of Y.[494] We can check that any function $f : Y \to X$ is continuous relative to the discrete topology because for any open set $U \in \tau$, $f^{-1}(U)$ is a subset of Y and hence it is open in $(Y,\mathcal{P}(Y))$. After checking that sending Y to $(Y,\mathcal{P}(Y))$ and $f : Y \to Y'$ to $f : (Y,\mathcal{P}(Y)) \to (Y',\mathcal{P}(Y'))$ is a functor, we denote it disc, we find can conclude that disc $\dashv U$.

[494] It is clear that the set of all subsets of Y is a topology because any union or intersection of subsets is still a subset.

Right adjoint: Fix a topological space (X,τ) and a set Y. We need to find a topological space (LY,λ) so that continuous functions $(X,\tau) \to (LY,\lambda)$ are in correspondence with functions $X \to Y$. Again, there is a trivial topology that we can put on Y that makes any function $f : X \to Y$ continuous, it is called the **codiscrete topology** and contains only the empty set and the full space Y.[495] We can check that any function $f : X \to Y$ is continuous relative to the codiscrete topology because the $f^{-1}(\emptyset) = \emptyset$ and $f^{-1}(Y) = X$ must be open by the definition of a topology. After checking that sending Y to $(Y,\{\emptyset,Y\})$ and $f : Y \to Y'$ to $f : (Y,\{\emptyset,Y\}) \to (Y',\{\emptyset,Y'\})$ is a functor, we denote it codisc, we can conclude that $U \dashv$ codisc.

[495] Since $\emptyset \cap Y = \emptyset$ and $\emptyset \cup Y$, we conclude that $\{\emptyset, Y\}$ is closed under any union and intersection, hence it is a topology.

We found our first chain of adjunctions disc $\dashv U \dashv$ codisc. Another interesting one is $\mathrm{colim}_{\mathbf{J}} \dashv \Delta^{\mathbf{J}}_{\mathbf{C}} \dashv \lim_{\mathbf{J}}$ in a category $\mathbf{C}$ with all limits and colimits of shape $\mathbf{J}$. A less interesting one is $\cdots \dashv \mathrm{id}_{\mathbf{C}} \dashv \mathrm{id}_{\mathbf{C}} \dashv \mathrm{id}_{\mathbf{C}} \dashv \cdots$. Here is a chain of five adjunctions.

OL **Exercise H.18.** Let $\mathbf{C}$ be a category and $\mathsf{id}, \mathsf{s}, \mathsf{t}$ be the functors described in Exercise E.37. Show they are related by the adjunctions $\mathsf{t} \dashv \mathsf{id} \dashv \mathsf{s}$. Suppose furthermore that $\mathbf{C}$ has an initial object $\emptyset$ and a terminal object $\mathbf{1}$. Show that the constant functor at $\mathrm{id}_\emptyset$ is left adjoint to t and the constant functor at $\mathrm{id}_{\mathbf{1}}$ is right adjoint to s.

As a final example, we show that any equivalence gives rise to two adjunctions. In this sense[496], one can see a left (resp. right) adjoint to a functor F as an approximation to a left (resp. right) inverse that is even coarser than a quasi-inverse.[497]

[496] And in another sense related to Kan extensions.

[497] Furthermore, it follows from Proposition H.3 (resp. Corollary H.5) that the left (resp. right) adjoint of F is the left (resp. right) inverse or quasi-inverse when the latter exists.

Proposition H.19. *Let $L : \mathbf{C} \rightsquigarrow \mathbf{D}$ and $R : \mathbf{D} \rightsquigarrow \mathbf{C}$ be quasi-inverses, then $L \dashv R$ and $R \dashv L$.*

Proof. It is enough to show $L \dashv R$ as the definition of quasi-inverses is symmetric. $\square$

Proposition H.20. *Let $\mathbf{C} : L \dashv R : \mathbf{D}$ be adjoint functors and $X, Y \in \mathbf{D}_0$. If $X \times Y$ exists, then $R(X \times Y)$ with the projections $R(\pi_X)$ and $R(\pi_Y)$ is the product $R(X) \times R(Y)$.*[498]

[498] In other words, right adjoints preserve binary products.

Proof. Let $p_X : A \to RX$ and $p_Y : A \to RY$ be such that (200) commutes.

$$\begin{tikzcd}
& A \arrow[dl, "p_X"'] \arrow[dr, "p_Y"] & \\
RX & R(X \times Y) \arrow[l, "R\pi_X"] \arrow[r, "R\pi_Y"'] & RY
\end{tikzcd} \tag{200}$$

We need to show there is a unique mediating morphism $A \to R(X \times Y)$. First, we will get rid of the applications of R at the bottom, in order to use the universal property of the product $X \times Y$. To do this, we apply L to (200) and use the counit $\varepsilon : LR \Rightarrow \mathrm{id}_{\mathbf{D}}$ to obtain (201).

$$\begin{tikzcd}
& LA \arrow[dl, "Lp_X"'] \arrow[dr, "Lp_Y"] & \\
LRX \arrow[d, "\varepsilon_X"'] & LR(X \times Y) \arrow[l, "LR\pi_X"] \arrow[r, "LR\pi_Y"'] \arrow[d, "\varepsilon_{X\times Y}"'] & LRY \arrow[d, "\varepsilon_Y"'] \\
X & X \times Y \arrow[l, "\pi_X"] \arrow[r, "\pi_Y"'] & Y
\end{tikzcd} \tag{201}$$

The universal property of $X \times Y$ tells us there is a unique $! : LA \to X \times Y$ such that $\pi_X \circ ! = \varepsilon_X \circ Lp_X$ and $\pi_Y \circ ! = \varepsilon_Y \circ Lp_Y$. We claim that $!^{\mathrm{t}}$ is the mediating morphism of (200), i.e.: $R\pi_X \circ !^{\mathrm{t}} = p_X$ and $R\pi_Y \circ !^{\mathrm{t}} = p_Y$. Using the adjunction $L \dashv R$, we obtain the following commutative square.

$$\begin{tikzcd}
& LA \arrow[dl, "Lp_X"'] \arrow[dr, "Lp_Y"] \arrow[dd, dashed, "!"] & \\
LRX \arrow[d, "\varepsilon_X"'] & & LRY \arrow[d, "\varepsilon_Y"'] \\
X & X \times Y \arrow[l, "\pi_X"] \arrow[r, "\pi_Y"'] & Y
\end{tikzcd}$$

$$\begin{tikzcd}
\mathrm{Hom}_{\mathbf{D}}(LA, X \times Y) \arrow[r, leftrightarrow] \arrow[d, "\pi_X \circ -"'] & \mathrm{Hom}_{\mathbf{C}}(A, R(X \times Y)) \arrow[d, "R\pi_X \circ -"] \\
\mathrm{Hom}_{\mathbf{D}}(LA, X) \arrow[r, leftrightarrow] & \mathrm{Hom}_{\mathbf{C}}(A, RX)
\end{tikzcd} \tag{202}$$

Now, starting with ! on the top left corner, we obtain the following derivation.

$$\begin{aligned}
p_X &= {p_X}^{\mathrm{t}\mathrm{t}} \\
&= (\varepsilon_X \circ Lp_X)^{\mathrm{t}} \\
&= (\pi_X \circ !)^{\mathrm{t}} && \text{definition of } ! \\
&= R\pi_X \circ !^{\mathrm{t}} && \text{commutativity of (202)}
\end{aligned}$$

Replacing X with Y in the previous argument shows $!^{\mathrm{t}}$ makes (203) commute. For the uniqueness, note that if $m : A \to R(X \times Y)$ can replace $!^{\mathrm{t}}$, then (204) commutes

$$\begin{tikzcd}
& A \arrow[dl, "p_X"'] \arrow[dr, "p_Y"] \arrow[d, "!^{\mathrm{t}}"'] & \\
RX & R(X \times Y) \arrow[l, "R\pi_X"] \arrow[r, "R\pi_Y"'] & RY
\end{tikzcd} \tag{203}$$

which implies by uniqueness of ! that $m^{\mathrm{t}} = \varepsilon_{X\times Y} \circ Lm = !$. Transposing yields $!^{\mathrm{t}} = m$.

$$\begin{tikzcd} & LA \arrow[dl, "Lp_X"'] \arrow[d, "Lm"] \arrow[dr, "Lp_Y"] & \\ LRX \arrow[d, "\varepsilon_X"'] & LR(X\times Y) \arrow[l, "LR\pi_X"] \arrow[r, "LR\pi_Y"'] \arrow[d, "\varepsilon_{X\times Y}"'] & LRY \arrow[d, "\varepsilon_Y"'] \\ X & X\times Y \arrow[l, "\pi_X"] \arrow[r, "\pi_Y"'] & Y \end{tikzcd} \tag{204}$$

□

Corollary H.21 (Dual)**.** *Let $\mathbf{C} : L \dashv R : \mathbf{D}$ be adjoint functors and $A, B \in \mathbf{C}_0$. If $A + B$ exists, then $L(A+B)$ with the coprojections $L\kappa_A$ and $L\kappa_B$ is the coproduct $LA \times LB$.*[499]

[499] In other words, left adjoints preserve binary coproducts.

Proposition H.22. *Let $\mathbf{C} : L \dashv R : \mathbf{D}$ be adjoint functors. If $g : X \to Y \in \mathbf{D}_1$ is monic, then $R(g)$ is monic.*[500]

[500] In other words, right adjoints preserve monomorphisms.

Proof. Let $h_1, h_2 : Z \to R(X)$ be such that $R(g) \circ h_1 = R(g) \circ h_2$, we need to show that $h_1 = h_2$. Since $L \dashv R$, we have the following commutative square.

$$\begin{tikzcd} \mathrm{Hom}_{\mathbf{C}}(Z, RX) \arrow[r, leftrightarrow] \arrow[d, "Rg\circ -"'] & \mathrm{Hom}_{\mathbf{D}}(LZ, X) \arrow[d, "g\circ -"] \\ \mathrm{Hom}_{\mathbf{C}}(Z, RY) \arrow[r, leftrightarrow] & \mathrm{Hom}_{\mathbf{D}}(LZ, Y) \end{tikzcd} \tag{205}$$

Starting with h_1 and h_2 in the top left corner, we find that[501]

$$g \circ {h_1}^{\mathrm{t}} = (Rg \circ h_1)^{\mathrm{t}} = (Rg \circ h_2)^{\mathrm{t}} = g \circ {h_2}^{\mathrm{t}},$$

[501] The first and last equality follow from commutativity of (205) and the middle equality is a hypothesis.

which, by monicity of g implies ${h_1}^{\mathrm{t}} = {h_2}^{\mathrm{t}}$. This in turn means that $h_1 = h_2$ because $(-)^{\mathrm{t}}$ is a bijection. □

Corollary H.23 (Dual)**.** *Let $\mathbf{C} : L \dashv R : \mathbf{D}$ be adjoint functors. If $f : A \to B \in \mathbf{C}_1$ is epic, then $L(f)$ is epic.*[502]

[502] In other words, left adjoints preserve epimorphisms.

Remark H.24. We want to put the emphasis on a crucial step in the proof above which was to derive $g \circ {h_1}^{\mathrm{t}} = (Rg \circ h_1)^{\mathrm{t}}$ from (205). By varying the arguments slightly (i.e.: going around the square in another direction or considering the naturality square involving pre-composition), we cook up four similar equations that can be helpful.[503]

[503] For instance, (207) was a crucial step in the proof of Proposition H.20: we used (202) to derive $(\pi_X \circ !)^{\mathrm{t}} = R\pi_X \circ !^{\mathrm{t}}$.

$$\forall g : X \to Y, f : Z \to RX, \qquad g \circ f^{\mathrm{t}} = (Rg \circ f)^{\mathrm{t}} \tag{206}$$

$$\forall g : X \to Y, f : LZ \to X, \qquad (g \circ f)^{\mathrm{t}} = Rg \circ f^{\mathrm{t}} \tag{207}$$

$$\forall g : LX \to Y, f : Z \to X, \qquad g^{\mathrm{t}} \circ f = (g \circ Lf)^{\mathrm{t}} \tag{208}$$

$$\forall g : X \to RY, f : Z \to X, \qquad (g \circ f)^{\mathrm{t}} = g^{\mathrm{t}} \circ Lf \tag{209}$$

Theorem H.25. *Right adjoints are continuous.*

Proof. Let $\mathbf{C} : L \dashv R : \mathbf{D}$ be an adjunction and $F : \mathbf{J} \rightsquigarrow \mathbf{D}$ be a diagram in $\mathbf{D}$ whose limit cone is $\{\ell_X : \mathrm{lim}F \to FX\}_{X\in\mathbf{J}_0}$. We claim that $\{R\ell_X : R\mathrm{lim}F \to RFX\}_{\mathbf{J}_0}$ is the limit cone of $R \circ F$. For any other cone making (210) commute for any $f : X \to Y \in \mathbf{J}_1$, we can apply transposition to the c_X's to obtain (211) which commutes by (206).[504]

[504] In (206), putting $g := Ff$ and $f := c_X$, we obtain
$$c_Y{}^{\mathsf{t}} = (RFf \circ c_X)^{\mathsf{t}} = Ff \circ c_X{}^{\mathsf{t}}.$$

$$\begin{tikzcd}
 & C \arrow[ddl, bend right, "c_X"'] \arrow[ddr, bend left, "c_Y"] & \\
 & R\lim F \arrow[dl, "R\ell_X"'] \arrow[dr, "R\ell_Y"] & \\
RFX \arrow[rr, "RFf"'] & & RFY
\end{tikzcd} \tag{210}$$

$$\begin{tikzcd}
 & LC \arrow[ddl, bend right, "c_X{}^{\mathsf{t}}"'] \arrow[ddr, bend left, "c_Y{}^{\mathsf{t}}"] & \\
 & \lim F \arrow[dl, "\ell_X"'] \arrow[dr, "\ell_Y"] & \\
FX \arrow[rr, "Ff"'] & & FY
\end{tikzcd} \tag{211}$$

By the universal property of $\lim F$, there is a unique mediating morphism $! : LC \to \lim F$ making (212) commute. Transposing ! yields a mediating morphism making (213) commutes by (207).[505]

[505] In (207), putting $g := \ell_X$ and $f := !$, we obtain
$$c_X = (c_X{}^{\mathsf{t}})^{\mathsf{t}} = (\ell_X \circ !)^{\mathsf{t}} = R\ell_X \circ !^{\mathsf{t}}.$$
Symmetrically, we have
$$c_Y = (c_Y{}^{\mathsf{t}})^{\mathsf{t}} = (\ell_Y \circ !)^{\mathsf{t}} = R\ell_Y \circ !^{\mathsf{t}}.$$

$$\begin{tikzcd}
 & LC \arrow[ddl, bend right, "c_X{}^{\mathsf{t}}"'] \arrow[ddr, bend left, "c_Y{}^{\mathsf{t}}"] \arrow[d, dashed, "!"] & \\
 & \lim F \arrow[dl, "\ell_X"'] \arrow[dr, "\ell_Y"] & \\
FX \arrow[rr, "Ff"'] & & FY
\end{tikzcd} \tag{212}$$

$$\begin{tikzcd}
 & C \arrow[ddl, bend right, "c_X"'] \arrow[ddr, bend left, "c_Y"] \arrow[d, dashed, "!^{\mathsf{t}}"] & \\
 & R\lim F \arrow[dl, "R\ell_X"'] \arrow[dr, "R\ell_Y"] & \\
RFX \arrow[rr, "RFf"'] & & RFY
\end{tikzcd} \tag{213}$$

Finally, $!^{\mathsf{t}}$ is the only mediating morphism that fits in (213) because if $m : C \to R\lim F$ fits, then $m^{\mathsf{t}} : LC \to \lim F$ fits in (212)[506] and by uniqueness of !, $m^{\mathsf{t}} = !$ which further implies $m = !^{\mathsf{t}}$. □

[506] Suppose $R\ell_X \circ m = c_X$, then we use (206) to conclude
$$c_X{}^{\mathsf{t}} = (R\ell_X \circ m)^{\mathsf{t}} = \ell_X \circ m^{\mathsf{t}},$$
and similarly for Y.

Corollary H.26 (Dual)**.** *Left adjoints are cocontinuous.*

Theorem H.27. *If* $\mathbf{C} : L \dashv R : \mathbf{D}$ *and* $\mathbf{D} : L' \dashv R' : \mathbf{E}$ *are two adjunctions, then* $\mathbf{C} : L'L \dashv RR' : \mathbf{E}$ *is an adjunction.*[507]

[507] This theorem is often referred to as *adjunctions can be composed*.

Proof. Let η and ε be the unit and counit of the first adjunction and η' and ε' be the unit and counit of the second one. We define the following unit and counit for the composite adjunction:

$$\widehat{\eta} = R\eta' L \cdot \eta : \mathrm{id}_{\mathbf{C}} \Rightarrow RR'L'L$$
$$\widehat{\varepsilon} = \varepsilon' \cdot L'\varepsilon R' : L'LRR' \Rightarrow \mathrm{id}_{\mathbf{E}}.$$

The following diagrams show the triangle identities.

$$\begin{tikzcd}
L'L \arrow[r, "L'L\eta"] \arrow[rr, bend left, "L'L\widehat{\eta}"] \arrow[dr, "\mathbb{1}_{L'L}"'] \arrow[ddrr, bend right, "\mathbb{1}_{L'L}"'] & L'LRL \arrow[r, "L'LR\eta' L"] \arrow[d, "L'\varepsilon L"] \arrow[dl, phantom, "\text{(a)}"] \arrow[dr, phantom, "\text{(b)}"] & L'LRR'L'L \arrow[d, "L'\varepsilon R'L'L"] \arrow[dd, bend left, "\widehat{\varepsilon}L'L"] \\
 & L'L \arrow[r, "L'\eta' L"] \arrow[dr, "\mathbb{1}_{L'L}"'] \arrow[dr, phantom, "\text{(c)}", shift left=3] & L'R'L'L \arrow[d, "\varepsilon' L'L"] \\
 & & L'L
\end{tikzcd} \tag{214}$$

Showing (214) commutes:

(a) Apply $L'(-)$ to the left triangle identity of η and ε.

(b) Apply $L'(-)L$ to $\mathsf{HOR}(\varepsilon, \eta')$.

(c) Apply $(-)L$ to the left triangle identity of η' and ε'.

$$\begin{tikzcd}
RR'L'LRR' \arrow[d, "RR'L'\varepsilon R'"] \arrow[dd, bend right=60, "RR'\widehat{\varepsilon}"'] \arrow[dr, phantom, "\text{(a)}"] & RLRR' \arrow[l, "R\eta' LRR'"'] \arrow[d, "R\varepsilon R'"] \arrow[dr, phantom, "\text{(b)}"] & RR' \arrow[l, "\eta RR'"'] \arrow[dl, "\mathbb{1}_{RR'}"] \arrow[ll, bend right, "\widehat{\eta}RR'"'] \arrow[ddll, bend left, "\mathbb{1}_{RR'}"] \\
RR'L'R' \arrow[d, "RR'\varepsilon'"] \arrow[r, phantom, "\text{(c)}", near end, shift right=6ex] & RR' \arrow[l, "R\eta' R'"'] \arrow[dl, "\mathbb{1}_{RR'}"] \\
RR'
\end{tikzcd} \tag{215}$$

Showing (215) commutes:
(a) Apply $R(-)R'$ to $\mathsf{HOR}(\eta', \varepsilon)$.
(b) Apply $(-)R'$ to the right triangle identity of η and ε.
(c) Apply $R(-)$ to the right triangle identity of η' and ε'.

$\square$

Proposition H.28. *If* $\mathbf{D} : L \dashv R : \mathbf{E}$ *is an adjunction, then there is an adjunction* $[\mathbf{C}, \mathbf{D}] : (L \circ -) \dashv (R \circ -) : [\mathbf{C}, \mathbf{E}]$.

Proof. We simplify the notation a little bit by writing $L-$ and $R-$ instead of $L \circ -$ and $R \circ -$ respectively. First, we can see that $L-$ and $R-$ are functors by Exercise F.34,[508] they send a natural transformation $\phi : F \Rightarrow G$ to $L\phi$ and $R\phi$ respectively. Composing them yields $RL- : [\mathbf{C}, \mathbf{D}] \rightsquigarrow [\mathbf{C}, \mathbf{D}]$ and $LR- : [\mathbf{C}, \mathbf{E}] \rightsquigarrow [\mathbf{C}, \mathbf{E}]$. Let $\eta : \mathrm{id}_{\mathbf{D}} \Rightarrow RL$ and $\varepsilon : LR \Rightarrow \mathrm{id}_{\mathbf{E}}$ be the unit and counit of $L \dashv R$. We claim that $\eta- = F \mapsto \eta F$ and $\varepsilon- = G \mapsto \varepsilon G$ are the unit and counit of an adjunction $L- \dashv R-$.

To see that $\eta-$ and $\varepsilon-$ are natural transformations of the right type, we can recognize them in the image of $\Lambda(- \circ -)$ (noting that $\mathrm{id}_{\mathbf{D}}- = \mathrm{id}_{[\mathbf{C},\mathbf{D}]}$ and $\mathrm{id}_{\mathbf{E}}- = \mathrm{id}_{[\mathbf{C},\mathbf{E}]}$):

$$\begin{aligned}
\eta- &= \Lambda(- \circ -)(\eta) : \mathrm{id}_{[\mathbf{C},\mathbf{D}]} \Rightarrow RL- \\
\varepsilon- &= \Lambda(- \circ -)(\varepsilon) : LR- \Rightarrow \mathrm{id}_{[\mathbf{C},\mathbf{E}]}.
\end{aligned}$$

It is left to show the triangle identities hold assuming they hold for η and ε. In the following derivations, we use three simple facts:[509]

- the biaction of $F-$ and $G-$ on $\phi-$ yields $(F\phi G)-$,
- $(\phi-) \cdot (\phi'-) = (\phi \cdot \phi')-$, and
- $(\mathbb{1}_F)- = \mathbb{1}_{F-}$.

Now, the triangle identities hold by:

$$\begin{aligned}
(\varepsilon-)(L-) \cdot (L-)(\eta-) = (\varepsilon L-) \cdot (L\eta-) = (\varepsilon L \cdot L\eta)- = (\mathbb{1}_L)- = \mathbb{1}_{L-} \\
(R-)(\varepsilon-) \cdot (\eta-)(R-) = (R\varepsilon-) \cdot (\eta R-) = (R\varepsilon \cdot \eta R)- = (\mathbb{1}_R)- = \mathbb{1}_{R-}.
\end{aligned}$$

$\square$

Corollary H.29 (Dual)**.** *If* $\mathbf{D} : L \dashv R : \mathbf{E}$ *is an adjunction, then there is an adjunction* $[\mathbf{C}, \mathbf{D}] : -L \dashv -R : [\mathbf{C}, \mathbf{E}]$.

[508] They are compositions:

$$\begin{aligned}
L- &= (- \circ -) \circ (\Delta(L) \times \mathrm{id}_{[\mathbf{C},\mathbf{D}]}) \\
R- &= (- \circ -) \circ (\Delta(R) \times \mathrm{id}_{[\mathbf{C},\mathbf{E}]}).
\end{aligned}$$

Alternatively, we can use Example F.36.5 where we described currying for functors. In that setting, we have

$$\begin{aligned}
L- &= \Lambda(- \circ -)(L) \\
R- &= \Lambda(- \circ -)(R).
\end{aligned}$$

[509] They can be shown by proving the equality at each component.

Theorem H.30. *Let* $\mathbf{D}$ *be a category with all limits of shape* $\mathbf{J}$. *For any category* $\mathbf{C}$, *the functor category* $[\mathbf{C}, \mathbf{D}]$ *has all limits of shape* $\mathbf{J}$ *and the limit of any diagram* $F : \mathbf{J} \rightsquigarrow [\mathbf{C}, \mathbf{D}]$ *satisfies for any* $X \in \mathbf{C}_0$, $(\lim_{\mathbf{J}} F)(X) = \lim_{\mathbf{J}}(F(-)(X))$.[510]

[510] This means limits in functor categories are taken pointwise, just like we proved in Theorem F.16

Proof. From previous results, we have the following chain of adjunctions.

$$\begin{tikzcd}[column sep=large]
{[\mathbf{C},\mathbf{D}]} \arrow[r, shift left=2, "\Delta^{\mathbf{J}}_{\mathbf{D}}\circ -"] \arrow[r, phantom, "\perp"] & {[\mathbf{C},[\mathbf{J},\mathbf{D}]]} \arrow[l, shift left=2, "\lim_{\mathbf{J}}\circ -"] \arrow[r, shift left=2, "\Lambda^{-1}"] \arrow[r, phantom, "\perp"] & {[\mathbf{C}\times\mathbf{J},\mathbf{D}]} \arrow[l, shift left=2, "\Lambda"] \arrow[r, shift left=2, "-\circ\mathsf{swap}"] \arrow[r, phantom, "\perp"] & {[\mathbf{J}\times\mathbf{C},\mathbf{D}]} \arrow[l, shift left=2, "-\circ\mathsf{swap}^{-1}"] \arrow[r, shift left=2, "\Lambda"] \arrow[r, phantom, "\perp"] & {[\mathbf{J},[\mathbf{C},\mathbf{D}]]} \arrow[l, shift left=2, "\Lambda^{-1}"]
\end{tikzcd} \tag{216}$$

From left to right. The first adjunction is induced by Proposition H.28 and the adjunction $\Delta^{\mathbf{J}}_{\mathbf{D}} \dashv \lim_{\mathbf{J}}$ given by completeness of $\mathbf{D}$. The second adjunction is obtained from Proposition H.19 and the fact that Λ and Λ^{-1} are inverses. The third adjunction is induced by Corollary H.29 and the canonical isomorphism $\mathsf{swap} : \mathbf{C} \times \mathbf{J} \rightsquigarrow \mathbf{J} \times \mathbf{C}$.[511] The fourth adjunction is similar to the second one.

[511] One could also see that $- \circ \mathsf{swap}$ and $- \circ \mathsf{swap}^{-1}$ are inverses.

There is a simpler way to describe the composition of the three rightmost adjunctions. If we view a functor $F : \mathbf{C} \rightsquigarrow [\mathbf{J}, \mathbf{D}]$ as taking two arguments and write it $F(-_1)(-_2)$, the composition $\Lambda \circ (- \circ \mathsf{swap}) \circ \Lambda^{-1}$ (the top path) swaps the order of the arguments to yield the functor $F(-_2)(-_1) : \mathbf{J} \rightsquigarrow [\mathbf{C}, \mathbf{D}]$. The bottom path swaps back the arguments.

Next, we show that the composition of the top path is $\Delta^{\mathbf{J}}_{[\mathbf{C},\mathbf{D}]}$. Starting with a functor $F : \mathbf{C} \rightsquigarrow \mathbf{D}$, the first left adjoint sends it to $\Delta^{\mathbf{J}}_{\mathbf{D}} \circ F$ which sends $X \in \mathbf{C}_0$ to the constant functor at FX and $f : X \to Y \in \mathbf{C}_1$ to the natural transformation whose components are all $Ff : FX \to FY$. Applying the three other left adjoints, we obtain a functor which sends any $j \in \mathbf{J}_0$ to the functor F and any $m : j \to j' \in \mathbf{J}_1$ to $\mathbb{1}_F$. We conclude that the top path sends F to the constant functor at F.

We obtain a right adjoint to $\Delta^{\mathbf{J}}_{[\mathbf{C},\mathbf{D}]}$ by composing all the right adjoins in (216) with Theorem H.27 and thus $[\mathbf{C}, \mathbf{D}]$ has all limits of shape $\mathbf{J}$. To compute them, we can compose the right adjoints in (216) to find $(\lim_{\mathbf{J}} F)(X) = \lim_{\mathbf{J}}(F(-)(X))$. □

Corollary H.31 (Dual)**.** *Let* $\mathbf{D}$ *be a category with all colimits of shape* $\mathbf{J}$. *For any category* $\mathbf{C}$, *the functor category* $[\mathbf{C}, \mathbf{D}]$ *has all colimits of shape* $\mathbf{J}$ *and the colimit of any diagram* $F : \mathbf{J} \rightsquigarrow [\mathbf{C}, \mathbf{D}]$ *satisfies for any* $X \in \mathbf{C}_0$, $(\operatorname{colim}_{\mathbf{J}} F)(X) = \operatorname{colim}_{\mathbf{J}}(F(-)(X))$.[512]

[512] In other words, colimits are taken pointwise. You can use Exercise F.15 or draw a similar chain of adjunctions as in (216).

Corollary H.32. *If a category* $\mathbf{D}$ *is (finitely) complete or cocomplete, then so is* $[\mathbf{C}, \mathbf{D}]$ *for any category* $\mathbf{C}$.

OL Exercise H.33. Let $\mathbf{C}$ have all limits of shape $\mathbf{J}$ and $\mathbf{C} : L \dashv R : \mathbf{D}$ be an adjunction. Using Theorem H.14, Corollary H.5, Theorem H.27 and Proposition H.28, show that R preserves all limits of shape $\mathbf{J}$.