# ALGEBRAIC AUTOMATA THEORY

Lecture Notes

Jürgen Koslowski

Theoretical Computer Science

Technische Universität Braunschweig

March 2019

# Contents

# 0 Overview: extending the "arena" of formal languages

*Formal languages* are just subsets of the *free monoid* $X^*$ over some (usually finite and not empty) *alphabet* $X$. For any nonempty alphabet $X$ there exist uncountably many such languages. Among these certain countable sub-classes are of interest in applications: first and foremost the *regular languages*, but also the other language classes of the Chomsky hierarchy. Interesting classes of languages often are specified by means of either grammars (for the purpose of generation) or machine models (for the purpose of recognition). In the regular case, several machine models turn out to be equivalent, the conceptually simplest one being that of a completely deterministic finite automaton.

Since any monoid can be embedded into its automorphism monoid, *e.g.*, by post-multiplication, recognition by completely deterministic finite automata can be abstracted to recognition via monoid homomorphisms with finite codomain. And just as there exists a minimal completely deterministic automaton recognizing a given regular language, there also exists a minimal monoid with this property, the so-called *syntactic monoid*, introduced by Schützenberger in 1956 [?].

With automata curiously removed from the picture in favor of morphisms into finite monoids, it makes sense to extend the idea of recognizing subsets from finitely generated free monoids to arbitrary ones. This leads to the notion of *recognizable set*. The terms "regular" and "language" are reserved for the case of free monoids.

A similar type of generalization will be applicable to the alternative characterization of regular languages via *regular expressions*. These are terms over $X$ with respect to the signature of formal operators $\emptyset$ (nullary), $+$, $\cdot$ (binary) and $(-)^*$ (unary). The intended semantics in $X^*P$, besides the singleton sets of singleton words, are the empty set, binary union, set-composition and the Kleene star. Kleene's Theorem [?] establishes that the semantics of regular expressions are precisely the regular languages.

When moving to general monoids instead of finitely generated free ones, one may no longer have easy access to a set of generators. Instead, one can start with all finite subsets, and then consider closure under union, composition and Kleene star. The resulting subsets of a monoid are called *rational*. They differ from the recognizable sets, unless the monoid in question is finitely generated and free. Even though, rational sets are also recognizable by certain automata, albeit non-deterministic ones. But in this case finite monoids do not suffice as targets of the recognizing morphisms, instead finite *ordered monoids* or even finite *unital quantales* need to be considered. Moreover, the recognizing morphisms that occur turn out to be lax homomorphisms. This is in fact not so surprising, if one considers the categorical approach to automata theory. But if the recognizing monoids are allowed to carry a compatible order, why not allow the same for the monoids where the regular sets live? Of course, rather than arbitrary subsets one expects upper or lower segments to be of interest in this case.

Returning to the original regular expressions, it is usually difficult to describe the complement of a given regular expression's semantics. This suggests extending the original signature by a unary operator for complement, say $(-)^{\mathrm{c}}$.

While initially the Kleene star was the only operator causing the semantics of a regular expression to be infinite, now the complementation can lead to infinite semantics as well. So it is natural to ask how the property of admitting a *star-free* in this extended signature is reflected in the recognizing monoids, or completely deterministic automata, for that matter.

Due to redundancy in the extended signature, such *star-free languages* may also admit regular expressions involving the Kleene star. Note that $X^* = \emptyset^{\mathrm{c}}$ is in fact star-free, as is $Y^*$ for any $Y \subseteq X$, while for $a \in X$ the language $(aa)^*$ is not.

This prototypical problem of characterizing certain classes of sub-regular languages in terms of their syntactic monoids was solved by Schützenberger in 1965 [**?**]. Another characterization of the star-free languages in terms of the first-order logic of orders with one successor by McNaughton and Papert followed in 1971 [**?**].

A similar problem, the characterization of the class of so-called "piece-wise testable" languages, in terms of their syntactic monoids, was accomplished independently by Brzozowski-Simon [**?**] and McNauchton [**?**], with a logical characterization by Thomas following in the early 1980s [**?**]

These types of results were subsumed by Eilenbrg's work in 1976 [**?**], who established a correspondence between *quasi-varieties* of finite monoids, *i.e.*, classes closed under sub-objects, quotients and finite products, and the sub-classes of regular languages recognized by them.

Quasi-varieties of finite algebras specialize Birkhoff's varieties of unconstrained algebras by restricting the permissible products to finite ones. While Birkhoff's characterization of varieties in terms of equations has been a corner stone of universal algebra since 1930, only in 1982 Reitermann [**?**] found a suitable characterization of quasi-varieties using pro-finite equations.

The necessary concepts, and the proof of Eilenberg's Theorem require the use of category theory. Once this is accepted, many other aspects of formal language theory lend themselves to categorical descriptions: at one level monoids themselves are categories, at another level together with their homomorphisms form a category ***mon***. The power-set construction turns monoids into ordered monoids in a functorial way, and the closure properties as well as the notion of recognizer are best formulated in the resulting category ***uqnt*** of "unital quantales". In fact, the same construction principle that yields ***uqnt*** can be employed to construct the categories of absorptive semi-groups, of monoids, and of absorptive monoids, as well as others.

It remains to be seen if the techniques of algebraic automata theory carry over from the category ***mon*** of monoids and their homomorphisms to the 2-category of ordered monoids and lax homomorphism.

# 1 Algebraic foundations

## 1.0 Monoids and related structures

Classically, monoids are specified as sets $M$ with a binary associative operation $\cdot$ and a neutral element $e \in M$ for this operation. From the point of view of universal algebra, this concept is part of a hierarchy of notions based on a single binary operation subject to various requirements:

**1.0.00 Definition.** A set $M$ equipped with a binary operation $M \times M \xrightarrow{\cdot} M$ is called a *magma*. We say an element $x \in M$ is

- ▷ *absorbing*, if $m \cdot x = x = x \cdot m$ for all $m \in M$;
- ▷ *neutral*, if $m \cdot x = m = x \cdot m$ for all $m \in M$;
- ▷ *idempotent*, if $x \cdot x = x$. ◁

**1.0.01 Definition.** We distinguish the following types of magmas:

- ▷ *commutative magmas*, where $\cdot$ is commutative (and is often written as $+$ instead);
- ▷ *semi-groups*, where $\cdot$ is associative;
- ▷ *idempotent semi-groups*, semi-groups where every element is idempotent;
- ▷ *absorptive semi-groups*, semi-groups with an absorbing element $0$;
- ▷ *acyclic semi-groups*, where for every element $m$ and each $n \in \mathbb{N}$ idempotency of $m^n$ implies $m^{n+1} = m^n$;
- ▷ *interpolative semi-groups*, where $\cdot$ is surjective;
- ▷ *monoids*, *i.e.*, semi-groups with a neutral element $e \in M$;
- ▷ *groups*, *i.e.*, monoids, where every $m \in M$ has an inverse $m^{-1} \in M$ satisfying $m^{-1} \cdot m = e = m \cdot m^{-1}$. Note that the only idempotent element of a group is the neutral element $e$.

as well as appropriate combinations of these concepts like commutative (semi-)groups or absorptive and acyclic monoids; note that monoids are automatically interpolative semi-groups. ◁

From a categorical perspective, observe that monoids are categories with a single object. Consequently, groups are categories as well, while the other notions fall short of being categories. (Interpolative semi-groups result from an attempt to formulate a good notion of category without identities, called *taxonomy*, or *taxon* for short.)

On the other hand, all notions above give rise to categories of the corresponding algebras together with their structure-preserving homomorphisms. The question, to which extent the preservation of all the structure is redundant is addressed in the exercises. Without proof we just state:

**1.0.02 Proposition.** *Let **mag** be the category of all magmas with their homomorphisms.*

- ▷ *Magma homomorphisms preserve idempotent elements.*
- ▷ *The category **sgr** of semi-groups is a full sub-category of **mag** .*
- ▷ *The category **idp-sgr** of idempotent semi-groups is a full sub-category of **sgr** .*
- ▷ *The category **absgr** of absorptive semi-groups with* $0$*-preserving homomorphisms is a non-full sub-category of **sgr** ,* i.e., *semi-group homomorphisms between absorptive semi-groups need not preserve the absorbing element* $0$ *.*
- ▷ *The category **mon** of monoids is a non-full sub-category of **sgr** ,* i.e., *semi-group homomorphisms between monoids need not preserve the neutral element* $e$ *. This carries over to the category **absmon** of absorptive monoids and* $0$*-preserving monoid-homomorphisms, which is a non-full subcategory of **absgr** .*
- ▷ *The category of **gr** groups is a full subcategory of **mon** .*
- ▷ *The category of commutative magmas / semi-groups /monoids / groups is a full subcategory of the category of all magmas / semi-groups /monoids / groups. Note that commutative groups are usually called* abelian, *and the corresponding category is denoted by **ab** .* □

In most of these cases there exists an easily describable "free algebra" $XT$ over any set $X$ , which in turn can be embedded canonically as a set of "generators" $X \xrightarrow{X\eta} XT$:

**1.0.03 Lemma.** *Given a set* $X$ *, the free*

- ▷ *magma on* $X$ *is the set* $XT$ *of finite rooted (and hence non-empty) ordered binary trees with leaves labeled in* $X$ *; trees of height 0 correspond to elements of* $X$ *, and the binary operation on* $XT$ *amounts to joining the roots of two trees with a new common root;*
  - *in the commutative case the order of a node's children is irrelevant;*
- ▷ *semi-group on* $X$ *is the set* $XT = X^+$ *of non-empty words over* $X$ *; words of length 1 correspond to elements of* $X$ *and the binary operation amounts to concatenation;*

- *in the commutative case again the ordering is irrelevant, which leaves us with non-empty finite* bags or multi-sets *as elements of* $XT$ *, which can be viewed as* frequency-functions $X \longrightarrow \mathbb{N}$ *with nonempty finite support;*
- *in the idempotent case repetition is irrelevant; we obtain the set* $XT = X^{\oplus}$ *of non-empty finite words w/o repetition by factoring* $X^{+}$ *by the obvious equivalence relation;*
- *in the commutative idempotent case* $XT$ *consists of all non-empty finite subsets of* $X$ , i.e., *characteristic functions* $X \longrightarrow 2$ *with non-empty finite support;*

▷ *absorptive (commutative/idempotent) semi-group is the free (commutative/idempotent) semi-group with a new absorber* $0$ *added;*

▷ *monoid on* $X$ *is the set* $XT = X^{*} = X^{+} + \{\varepsilon\}$ *of finite words over* $X$ , i.e., *the free semi-group extended by the empty word* $\varepsilon$ *that serves as neutral element wrt. concatenation;*

- *in the commutative case ignoring the order results in all finite bags, including* $\emptyset$ *;*
- *in the idempotent case repetitions are eliminated as before;*
- *in the commutative idempotent case* $XT$ *is just* $XF$ *, the set of all finite subsets of* $X$ , i.e., *characteristic functions* $X \longrightarrow 2$ *with finite support;*

▷ *absorptive (commutative/idempotent) monoid on* $X$ *is the free (commutative/idempotent) monoid with a new absorber* $0$ *added; note that the free absorptive monoid on a singleton is* $\langle \mathbb{N}, \cdot, 1 \rangle$ *, which is automatically commutative;*

▷ *group on* $X$ *is the set* $\big(X + \{\, x^{-1} : x \in X \,\}\big)^{*}$ *modulo the least equivalence relation that identifies* $xx^{-1}$ *as well as* $x^{-1}x$ *with* $\varepsilon$ *. Elements of* $X$ *correspond singleton words w/o exponent; the binary operation of amounts to concatenation modulo elimination of expressions* $xx^{-1}$ *as well as* $x^{-1}x$ *;*

- *in the commutative or abelian case we can restrict our attention to frequency-functions* $X \longrightarrow \mathbb{Z}$ *with finite support.*

*In all these cases the assignment* $X \mapsto XT$ *can be extended to an endo-functor* ***set*** $\xrightarrow{T}$ ***set****, for which the family of embeddings* $X \xrightarrow{X\eta} XT$*, constitute a natural transformation.* □

(not sure about the free interpolative semi-group and the free acyclic monoid)

## 1.1 Monads and distributive laws in action

One would hope to identify all these categories of classical algebras with categories of EM-algebras (*cf.*, Definition 5.6.04) for suitable monads $\mathbb{T} = \langle T, \eta, \mu \rangle$ (*cf.*, Definition 5.6.00), where $T$ is the free algebra functor. In some cases, these monads are in fact composites of simpler monads.

In the following cases we are not aware of any decompositions of the respective monads:

**1.1.00 Examples.**

(0) For magmas, $XTT \xrightarrow{X\mu} XT$ maps a binary tree with binary trees over $X$ as leaves to the binary tree over $X$ that results from substituting the original leaves as trees into the original tree w/o the leaves; in other words, it forgets the distinction of the original leaves.

Despite the binary operation on the magma $XT$ not being associative, the multiplication $\mu$ is in fact associative: on a 3-level tree of trees of trees, it does not matter at which level the distinction of what are the leaves is removed first.

The neutrality of the unit is immediately clear: for a single-node tree with a tree over $X$ as its only leaf the composition produces this latter tree. Similarly a tree with trees of height 0 as leaves upon composition results in an isomorphic tree with leaves in $X$.

We claim that the EM-algebras for this binary tree monad are precisely the magmas. Any EM-algebra $\langle X, \xi \rangle$ induces a binary operation $\cdot$ on $X$ by restricting to binary trees of height 1. Conversely, for any magma $\langle X, \cdot \rangle$ we assign to a binary tree of positive height in $XT$ the result of composing according to the structure of the tree. Trees of height 0 must be mapped to themselves in order to satisfy the compatibility with $X\eta$. We leave the compatibility with $X\mu$ as an exercise.

(1) For semi-groups the same argument applies as for magmas, as one can restrict attention to flattened trees. Due to the presence of a root these are just non-empty words. Hence $X^{++} \xrightarrow{x\eta} X^{+}$ maps a non-empty sting of non-empty words to their concatenation.

The EM-algebras are sets $S$ with a structure-map $S^{+} \xrightarrow{\sigma} S$. Restricting $\sigma$ to the subset $S \times S \subseteq S^{+}$ yields a binary operation $\sigma_2$ that due to the monad laws is associative: the second diagram in (5.6-01) for $a, b, c \in S$ shows

$$ab\sigma_2 c\sigma_2 = (abc)\sigma = abc\sigma_2\sigma_2$$

where we used *Reversed Polish Notation* (RPN). Hence $\langle S, \sigma_2 \rangle$ is a semi-group. Conversely, for any semi-group $\langle S, \cdot \rangle$ the associativity of $\cdot$ allows us to define $S^{+} \xrightarrow{\sigma} S$ by iterated binary composition in any order, except for the unary case. But in order to be compatible with $S\eta$ unary composition has to be the identity. The compatibility of $\sigma$ with $X\mu$ is straightforward.

(2) For groups just observe that the formal inverse $w^{-1}$ for $w \in XT$ reverses the order and interchanges (implicit) exponents 1 with $-1$. The rest is quite similar to the semi-group case.

(3) These constructions carry over to the commutative cases as well.

(4) As seen above, the free commutative idempotent semi-group on $X$ is the set $XF$ of finite subsets of $X$, and the multiplication $XFF \xrightarrow{X\mu} XF$ reduces to the union-operation. We denote the resulting monad by $\mathbb{F} = \langle F, \{-\}, \bigcup \rangle$. □

**1.1.01 Example.** The *exception monad* $\mathbb{E} = \langle +1, \eta, \mu \rangle$ on ***set*** is given by the functor $\textbf{\textit{set}} \xrightarrow{+1}$ ***set*** that adds a new element (the *exception*) to every set, while functions $A \xrightarrow{f} B$ are extended in such a way as to preserve the new element.

The unit $A \xrightarrow{A\eta} A{+}1$ includes $A$ into the disjoint union of $A$ with $1$, while the multiplication $A + 1 + 1 \xrightarrow{A\mu} A + 1$ collapses the two extra points into one.

The EM-algebras are precisely the *pointed sets*, since the structure morphism $X + 1 \xrightarrow{\xi} X$ selects a distinguished element $1\xi \in X$ that has to be preserved by EM-homomorphisms.

Alternatively, essentially the same monad arises by mapping $X$ to its set of *sub-terminal* subsets, *i.e.*, all subsets with at most one element.

**1.1.02 Lemma.** *Consider the semi-group monad* $\mathbb{H} = \langle (-)^+, \eta, \mu \rangle$ *and the exception monad* $\mathbb{E}$. *These are linked by two distributive laws:*

(0) *Call the new point* $0$. *Then* $(X + \{0\})^+ \xrightarrow{X\delta_0} X^+ + \{0\}$ *leaves non-empty* $0$*-free words invariant, and collapses words with at least one* $0$ *to* $0$. *The resulting composite monad has absorptive semi-groups as EM-algebras. Lifting the exception monad to* $\textbf{\textit{set}}^{\mathbb{H}} = \textbf{\textit{sgr}}$ *along* $\delta_0$ *has the effect of adding a new absorber* $0$ *to any given semi-group.*

(1) *Call the new point* $\varepsilon$. *Then* $(X + \{\varepsilon\})^+ \xrightarrow{X\delta_1} X^+ + \{\varepsilon\} = X^*$ *forgets all letters* $\varepsilon$ *in the argument and maps the resulting non-empty finite string of words in* $X^*$ *to its concatenation. The resulting composite monad* $\mathbb{L} = \langle (-)^*, \eta, \mu \rangle$ *is called the* list monad *and has monoids as EM-algebras. Lifting the exception monad to* ***sgr*** *along* $\delta_1$ *has the effect of adding a new neutral element to any given semi-group and thus producing a monoid.*

(2) *In similar fashion one can compose with the exception monad again and add a new absorber* $0$ *to any monoid, and a new neutral element* $\varepsilon$ *to any absorptive semi-group. Both constructions yield the category* ***absmon*** *of absorptive monoids as the category of EM-algebras.*

**1.1.03 Remark.** In the standard literature on semi-groups (*cf.*, *e.g.*, [Pin18, p. 14]) one finds the following constructions: given a semi-group $S$,

- ▹ $S^1$ results from adding a new neutral element, provided $S$ was not already a monoid;
- ▹ $S^0$ results from adding a new absorber $0$, regardless whether $S$ already had one or not.

It is surprising that these two cases are handled differently. The only purpose of defining $S^1$ in this fashion seems to be a spurious simplification in the definition of the congruences on semi-groups, see Remark 2.4.02. On the other hand, one looses functoriality and monadicity.

**1.1.04 Example.** A *ring* $\langle R, +, 0, \cdot, 1 \rangle$ is a set $R$ with two structures subject to a compatibility condition: $\langle R, +, 0 \rangle$ is an abelian group, and $\langle R, \cdot, 1 \rangle$ is a monoid such that multiplication $\cdot$ distributes over addition $+$ from either side, *i.e.*, $a \cdot (b+c) = a \cdot b + a \cdot c$ and $(a+b) \cdot c = a \cdot c + b \cdot c$.

At the level of monads, the free abelian group $XT$ over $X$ consists of formal sums of positive or negative elements of $X$, while the free monoid over $Y$ consists of formal products of elements of $Y$. The obvious distributive law $T^* \xrightarrow{\delta} (-)^*T$ maps formal products of formal sums to formal sums of formal products, subject of the provision that $(-a) \cdot (-b) = a \cdot b$.

By dropping the requirement for inverses in the abelian group, essentially the same distributive law as in Example 5.9.01 allows us to compose the monad for commutative monoids with the list monad. The resulting EM-algebras are known as *semi-rings*, or as *rigs* (which stands for "rings w/o negatives").

Dropping instead the requirement for a neutral element in the monoid, the same distributive law facilitates the composition of the commutative monoid monad with the semi-group monad, with *rngs* as EM-algebras (rings w/o identities).

**1.1.05 Example.** One can combine the finite power-set monad $\mathbb{F}$ with itself using essentially the same distributive law as above: given a finite subset $\mathcal{A} \subseteq XF$, the set $\mathcal{A}X\delta$ consists of all sets $B \subseteq \mathcal{A}\bigcup \subseteq X$ that have a singleton intersection with each $A \in \mathcal{A}$. Since $\mathcal{A}\bigcup$ is finite, so is any such set $B$, and there exist only finitely many of them.

Interpreting the elements of $\mathcal{A}$ as positive Boolean clauses (disjunctions), we can identify $\mathcal{A}$ as a positive Boolean formula in conjunctive normal form (cnf). Then $\mathcal{A}X\delta$ is the equivalent disjunctive normal form (dnf) that results from just applying the distributive laws concerning conjunction and disjunction.

The corresponding EM-algebras are easily identified as the *distributive lattices*.

**1.1.06 Example.** An example for a monad that seems to be indecomposable, even though its algebras carry two separate structures subject to a compatibility condition, is given by lattices: $\langle L, \sqcap, \sqcup \rangle$ combines two commutative idempotent semi-group structures $\langle L, \sqcup \rangle$ and $\langle L, \sqcap \rangle$ subject to the absorption laws:

$$a \sqcap (a \sqcup b) = a \qquad \text{and} \qquad a \sqcup (a \sqcap b) = a$$

Although the category $\boldsymbol{lat}$ of lattices and lattice-homomorphisms is the category of EM-algebras for a monad, it is unlikely that this results from a distributive law on the monad $\mathbb{F}$. In fact, the free lattice $XT$ on a set $X$ is rather difficult to describe.

## 1.2 Combining the list monad and variants of the power-set monad

As mentioned in the overview, formal languages are subsets of free monoids. Hence we wish to lift the power-set functor on $\boldsymbol{set}$ to $\boldsymbol{mon} = \boldsymbol{set}^{\mathbb{L}}$. In example 1.1.00(4) we have already seen that the functor $F$ on $\boldsymbol{set}$ that maps $X$ to the set of finite subsets, is part of a monad, with commutative idempotent semi-groups as EM-algebras. Alternatively, $\boldsymbol{set}^{\mathbb{F}}$ can be identified with $\sqcup$-$\boldsymbol{slat}$, as well as with $\sqcap$-$\boldsymbol{slat}$, *cf.*, Proposition 6.0.04.

So it cannot be a surprise that the power-set functor on ***set*** carries a monad structure with unconstrained union $\bigcup$ as multiplication. Obviously, the EM-algebras have to be $\bigsqcup$-semi-lattices, and the EM-homomorphisms must preserve arbitrary suprema (but not necessarily infima, although they also exist).

More precisely: The structure map $X\mathbb{P} \xrightarrow{\bigsqcup} X$ for an EM-algebra satisfies

$$
\begin{array}{ccc}
X & \xrightarrow{X\{-\}} & X\mathbb{P} \\
 & \searrow^{id_X} & \downarrow \bigsqcup \\
 & & X
\end{array}
\qquad \text{and} \qquad
\begin{array}{ccc}
X\mathbb{P}\mathbb{P} & \xrightarrow{\bigsqcup\mathbb{P}} & X\mathbb{P} \\
{\scriptstyle \bigcup_X}\downarrow & & \downarrow \bigsqcup \\
X\mathbb{P} & \xrightarrow[\bigsqcup]{} & X
\end{array}
\tag{1.2-00}
$$

These suggest that $\bigsqcup$ might be some kind of supremum operation. Indeed, $\bigsqcup$ induces a canonical partial ordering on $X$ via

$$
x \sqsubseteq y \qquad \text{iff} \qquad \{x, y\}\bigsqcup = y \tag{1.2-01}
$$

Reflexivity and anti-symmetry are immediately clear. If $x \sqsubseteq y$ and $y \sqsubseteq z$, then

$$
\begin{aligned}
\{x, z\}\bigsqcup &= \Big\{x, \{y, z\}\bigsqcup\Big\}\bigsqcup = \Big\{\{x\}\bigsqcup, \{y, z\}\bigsqcup\Big\}\bigsqcup = \Big(\{\{x\}, \{y, z\}\}\bigsqcup\mathbb{P}\Big)\bigsqcup \\
&= \Big(\{\{x\}, \{y, z\}\}\bigcup{}_X\Big)\bigsqcup = \{x, y, z\}\bigsqcup = \Big(\{\{x, y\}, \{z\}\}\bigcup{}_X\Big)\bigsqcup \\
&= \Big(\{\{x, y\}, \{z\}\}\bigsqcup\mathbb{P}\Big)\bigsqcup = \Big\{\{x, y\}\bigsqcup, \{z\}\bigsqcup\Big\}\bigsqcup = \Big\{\{x, y\}\bigsqcup, z\Big\}\bigsqcup = \{y, z\}\bigsqcup \\
&= z
\end{aligned}
$$

This establishes transitivity. It remains to show that $A\bigsqcup$ is indeed a supremum of $A \subseteq X$ with respect to $\sqsubseteq$. For $a \in A$ we get

$$
\Big\{a, A\bigsqcup\Big\}\bigsqcup = \Big\{\{a\}\bigsqcup, A\bigsqcup\Big\}\bigsqcup = \Big(\{\{a\}, A\}\bigsqcup\mathbb{P}\Big)\bigsqcup = (\{a\} \cup A)\bigsqcup = A\bigsqcup
$$

This shows that $A\bigsqcup$ is indeed an upper bound for $A$. Given another upper bound $y \in X$ of $A$, *i.e.*, $\{a, y\}\bigsqcup = y$ for all $a \in A$, we obtain

$$
\begin{aligned}
\Big\{A\bigsqcup, y\Big\}\bigsqcup &= \Big(\{A, \{y\}\}\bigsqcup\mathbb{P}\Big)\bigsqcup = (A \cup \{y\})\bigsqcup = \Big(\{\{a, y\} : a \in A\}\bigsqcup\mathbb{P}\Big)\bigsqcup \\
&= \Big\{\{a, y\}\bigsqcup : a \in A\Big\}\bigsqcup = \{y : a \in A\}\bigsqcup = \{y\}\bigsqcup = y
\end{aligned}
$$

This establishes $A\bigsqcup \sqsubseteq y$, hence $A\bigsqcup$ is indeed the least upper bound, or supremum, of $A$.

In order to combine the list monad $\mathbb{L}$ with the power-set monad $\mathbb{P}$ consider the following candidate for a distributive law:

$$
P(-)^* \xrightarrow{\delta} (-)^*P
$$

(and similarly for $\mathbb{F}$ ) that maps a string of $n$ (finite) subsets $A_i \subseteq X$ to their cartesian product, *i.e.*, the (finite) subset of $X^n \subseteq X^*$ consisting of those $n$-tuples, whose $i$-th component belongs to $A_i$ , $i < n$ . The axioms of a distributive law are readily established HW.

Therefore in both cases, for $\mathbb{P}$ and for $\mathbb{F}$ , one obtains two new monads: a composite monad on ***set*** with functor $(-)\mathbb{P}$ , resp., $(-)\mathbb{F}$ , and liftings of the power-set monads to ***mon*** , the category of EM-algebras for the free monoid monad. What are their algebras?

Fortunately, in both cases the categories of EM-algebras for the composite monad on ***set*** and the lifted monad on ***mon*** agree.

For $\mathbb{P}$ we obtain the category ***unit*** of so-called *unital quantales*, *i.e.*, complete lattices with a monoid structure, such that supremum is a monoid homomorphism. Morphisms are the supremum-preserving monoid-homomorphisms.

For $\mathbb{F}$ we obtain the category ***duo*** of so-called *dioids*, *i.e.* $\sqcup$-semi-latices with monoid structure, or equivalently, idempotent semi-rings. Recall that a *ring* is (classically) a set $R$ equipped with the structure of an Abelson group $\langle R, +, 0\rangle$ and a monoid $\langle R, \cdot, 1\rangle$ such that multiplication $\cdot$ distributes over addition. (There are various fancy-er characterizations of rings, *e.g.*, as the EM-algebras for the composition of the free Oberlin group monad with the free monoid monad over ***set*** , or the EM-algebras of the lifting of the free Oberlin group monad to the category ***mon*** of monoids, or as monoids *internal* to the category ***ab*** of Oberlin groups with the tensor product.) One can now weaken the requirement on $\langle R, +, 0\rangle$ to be just a commutative semi-group, in which case one obtains *semi-rings*, or ask for $\langle R, +\rangle$ to be a commutative monoid, which yields *unital semi-rings.* Adding the requirement that $+$ be idempotent then results in dioids. Taking non-empty finite subsets ought to produce dioids without units as EM-algebras. (Among category theorists the terms “rng” and “rig” are often used for “rings without unit” and “ring without negatives”, respectively.)

**1.2.00 Corollary.** *The set-multiplication* $A \cdot B = \{\, a \cdot b \,:\, a, \in A \wedge b \in B \,\}$ *in the quantale* $MP$ *over a monoid* $M$ *distributes over suprema,* i.e., *unions.*

**Proof.** This is an immediate consequence of the fact that $MPP \xrightarrow{\bigcup} MP$ is a monoid homomorphism. $\square$

**1.2.01 Example.** The set-multiplication in the quantale $MP$ over a monoid $M$ in general does not distribute over intersections:

$$\{a, ab\} \cdot (\{bc\}, \{c\}) = \emptyset \neq \{abc\} = \{abc, abbc\} \cap \{ac, abc\} = \{a, ab\} \cdot \{ac\} \cap \{a, ab\} \cdot \{c\} \quad \triangleleft$$

# 2 Automata over general monoids, recognizable and rational sets

Strangely enough, algebraic automata theory talks little about finite automata. At least the completely deterministic ones have been "abstracted away" and replaced by finite monoids.

One of the main results of the elementary regular language theory is the fact that various machine models or types of automata are equivalent. The main focus is on completely deterministic automata, as those can actually be implemented for solving problems. The other models are intended for making it easier to verify that they satisfy a given specification, before mechanically transforming them into completely deterministic automata.

Since formal languages live inside finitely generated free monoids $X^*$, the classical notion of finite automaton initially is based on assigning transitions just to the letters of the alphabet $X$. The consideration of silent transitions requires to extend this by the empty word $\varepsilon \in X^*$. Direct transitions for words $w \in X^*$ with $|x| > 1$ are seldom mentioned, perhaps since $X^*$ in general is infinite. Instead, in order to check whether a word over $X$ is recognized by the automaton, the notion of "computation" is introduced, which essentially corresponds to a path in a graph of so-called "configurations". This, of course, is just a smoke screen to avoid talking about extending some function from $X$ to $X^*$, preferably to a monoid homomorphism.

After extending the closure properties formalized by regular expression from free monoids to all monoids, resulting in the monad $\text{RAT}$, it seems reasonable to extend the notion of automaton to arbitrary monoids as well.

## 2.1 Variants of classically equivalent types of automata

The types of automata studied in elementary automata theory are all based on a finite set $Q$ of *states* with certain distinguished subsets $I, F \subseteq Q$ of *initial*, resp. *final* states and an assignment $\delta$ of *transitions* between the states depending on the input from $X$ and the current state. Various ways of describing this assignment occur in the literature, but we prefer to use a function $\delta$ that assigns endomorphisms of $Q$ in a suitable category to elements of the alphabet:

- ▷ $X \xrightarrow{\delta} \langle Q, Q\rangle \boldsymbol{set}$ for *completely deterministic* automata, provided there is one initial state;
- ▷ $X \xrightarrow{\delta} \langle Q, Q\rangle \boldsymbol{prt}$ for *deterministic* automata, provided there is at most one initial state;
- ▷ $X \xrightarrow{\delta} \langle Q, Q\rangle \boldsymbol{rel}$ for *non-deterministic* automata.

Here $\boldsymbol{set}$, $\boldsymbol{prt}$, and $\boldsymbol{rel}$ denote the categories of sets as objects and functions (always presumed to be total, unless stated otherwise), resp. partial functions, resp. binary relations as morphisms.

As such endomorphism sets always carry a natural monoid structure under composition, in all these cases $\delta$ may be extended uniquely to a monoid homomorphism $\delta'$ with domain $X^*$. In fact, all these machine models are equivalent: adding a *Hotel California state* to a deterministic automaton yields a completely deterministic one, while the power set construction makes a non-deterministic automaton completely deterministic.

It turns out to be convenient to consider even more general transition assignment functions:

- ▷ $X + \{\varepsilon\} \xrightarrow{\delta} \langle Q, Q\rangle \boldsymbol{rel}$ for automata *with silent transitions*;
- ▷ $X^* \xrightarrow{\delta} \langle Q, Q\rangle \boldsymbol{rel}$ with *finite support*, *i.e.*, only finitely many transitions relations $w\delta$ are non-empty, for *finite* automata, the most general type of interest here.

In the first case eliminating silent transitions yields a homomorphism $X^* \xrightarrow{\delta'} \langle Q, Q\rangle \boldsymbol{rel}$ that can be used for recognizing the same language. This may require expanding the subsets $I, F \subseteq Q$, and for each $a \in X$ we only have $a\delta \subseteq a\delta'$. Provided the *diagonal* $Q\Delta$ is properly contained in the reflexive transitive closure of $\varepsilon\delta$, some of these inclusions are guaranteed to be proper.

The second case subsumes all previous ones. To construct a monoid homomorphism on $X^*$ such that the corresponding automaton recognizes the same language as before, first add finitely many intermediate states to allow decomposing $w$-transitions with $|w| > 1$ into transitions with labels from $X$. These new states are neither initial nor final. Then eliminate the silent transitions as before.

Alternatively, taking into account that $\langle Q, Q\rangle \boldsymbol{rel}$ is an *ordered monoid*, even a *unital quantale*, one can at least obtain a *lax* monoid homomorphism $X^* \xrightarrow{\bar\delta} \langle Q, Q\rangle \boldsymbol{rel}$, satisfying $Q\Delta \subseteq \varepsilon\bar\delta$ and $(u\bar\delta) \cdot (v\bar\delta) \subseteq (uv)\bar\delta$.

These general automata are useful, since usually they can be designed more easily and with fewer states than implementable completely deterministic ones to meet certain specifications (*e.g.*, recognizing a given language). Then there exist mechanical procedures for transforming them into more constrained automata that meet the same specifications. However, along the way the size of the state set can grow exponentially due to the power-set construction.

In the following we will generalize completely deterministic automata and finite automata from finitely generated free monoids to arbitrary ones. While the resulting automata are often disambiguated as "deterministic" and "non-deterministic", we prefer to use a different terminology that in our view better addresses the real conceptual difference between these types of automata.

## 2.2 Strict $M$-automata and recognizable sets

In order to allow arbitrary monoids $M$ instead of just finitely generated free monoids $X^*$, we have to extend the transition assignment functions $\delta$ to all of $M$. Only for completely deterministic automata above was the codomain a plain monoid, the function monoid on the state set $Q$. Since arbitrary functions from $M$ to $\langle Q, Q\rangle \boldsymbol{set}$ in general will only induce a monoid homomorphism, on $M^*$ but not on $M$, one has little choice but requiring $\delta$ to already be a monoid homomorphism.

**2.2.00 Definition.** For any monoid $M$, a *strict* $M$*-automaton* $\mathcal{A} = \langle Q, \delta, q_0, F\rangle$ consists of

▷ a set $Q$ of *states*;

▷ a monoid homomorphism $M \xrightarrow{\delta} \langle Q, Q\rangle \boldsymbol{set}$ assigning *transition functions* on $Q$ to the elements of $M$;

▷ an *initial state* $q_0 \in Q$ and a set of *final states* $F \subseteq Q$.

The subset of $M$ *recognized* by $\mathcal{A}$ is

$$\mathcal{A}\mathscr{L} = \{\, m \in M \,:\, (q_0)m\delta \in F\}$$

which is also the $\delta$-pre-image of $F' := \{Q \xrightarrow{f} Q \,:\, q_0 f \in F\}$. We call $\mathcal{A}$ *finite*, if $Q$ is finite. ◁

Notice that the only finiteness condition that can be imposed in this case concerns the size of the state-set $Q$ for the function-monoid.

**2.2.01 Remark.** Even though the codomain of $\delta$ is a monoid, the significance of mapping $e \in M$ to $id_Q$ is not immediately clear. In the same fashion one could define a *strict $M$-automaton* for a semi-group $M$, where $\delta$ is a semi-group homomorphism.

The essence of the definition above would seem to be the sub-monoid (sub-semi-group) of $\langle Q, Q\rangle \boldsymbol{set}$ spanned by the selected transition functions, and the subset of those functions $Q \xrightarrow{f} Q$ that map $q_0$ into $F$. This suggests the following general definition:

**2.2.02 Definition.** A subset $L$ of a monoid (semi-group) $M$ is said to be *recognized by a morphism* $M \xrightarrow{\varphi} N$, if $L$ belongs to the image of $NP \xrightarrow{\varphi^{\leftarrow}} MP$. ◁

Recall that both $\boldsymbol{sgr}$ and $\boldsymbol{mon}$ are exact categories, *cf.*, Definition 5.13.06. In particular, regular epis are surjective, every morphism admits a $\langle$regular epi, mono$\rangle$-factorization by taking the co-equalizer of its kernel pair, and all congruences arise as kernel pairs. Hence as far as recognizing homomorphisms are concerned, it suffices to consider regular epis. And these can be described in terms of their congruences (Section 2.4 will address this point in greater depth). In addition, since any homomorphism $M \xrightarrow{\varphi} N$ sets up an adjunction $\varphi_\exists \dashv \varphi^{\leftarrow}$ between the respective power-sets, to be recognized amounts to being a fixed point (= EM-algebra) of the resulting closure operator (= monad) on $MP$. In more elementary terms:

**2.2.03 Proposition.** *For $L \subseteq M \xrightarrow{\varphi} N$ the following are equivalent:*

(a) *$L$ is recognized by $\varphi$.*

(b) *$L$ is union of equivalence classes of the kernel pair $\sim_\varphi$ of $\varphi$.*

(c) $L(\varphi_\exists \cdot \varphi^{\leftarrow}) = L$, i.e., *$L$ is a fixed point of the closure-operator on $MP$ induced by $\varphi$, or an EM-Algebra for the monad on $MP$ induced by $\varphi$.*

**Proof.**

(a) $\Rightarrow$ (b): Let $L$ be the $\varphi$-pre-image of $P \subseteq N$. If $x \sim_\varphi y$, then $x\varphi = y\varphi$, and hence $x$ and $y$ either both belong to the $L$, or both lie outside of $L$. Hence every equivalence class of $\sim_\varphi$ is either contained in $L$, or does not intersect $L$. Since the equivalence classes form a partition of $M$, the claim follows.

(b) $\Rightarrow$ (c): Consider $y \in L(\varphi_\exists \cdot \varphi^\leftarrow)$. There exists $x \in L$ with $x\varphi = y\varphi$, and therefore $x \sim_\varphi y$. But this implies $y \in L$.

(c) $\Rightarrow$ (a): Trivial. □

While the left adjoint $MP \xrightarrow{\varphi_\exists} NP$ always preserves unions, in general it is less well-behaved w.r.t. intersections and relative complements. However, if a fixed point of $\varphi_\exists \cdot \varphi^\leftarrow$ is involved, we obtain the following computation rules:

The following rules may enable us to simplify later calculations.

**2.2.04 Lemma.** *([Pin18, Proposition IV.2.6]) If $L$ is recognized by $M \xrightarrow{\varphi} N$, then the direct image function $\varphi_\exists$ preserves intersections with and relative complements of $L$.*

**Proof.** In view of Proposition 2.2.03, both $R\varphi_\exists\varphi^\leftarrow$ and $L$ are unions of $\varphi$-classes, hence so is their intersection. But $R\varphi_\exists\varphi^\leftarrow \cap L$ contains $n\varphi^\leftarrow$ iff $(R \cap L) \cap n\varphi^\leftarrow \neq \emptyset$, hence both of these sets have the same $\varphi$-image, namely

$$(R\varphi_\exists\varphi^\leftarrow \cap L)\varphi_\exists = R\varphi_\exists\varphi^\leftarrow\varphi_\exists \cap L\varphi_\exists = R\varphi_\exists \cap L\varphi_\exists$$

On the other hand, $R - L = R \cap (M - L)$, hence by Proposition 2.3.01 and the first part we get

$$(R - L)\varphi_\exists = R\varphi_\exists \cap (M - L)\varphi_\exists = R\varphi_\exists \cap (N - L\varphi_\exists) = R\varphi_\exists - L\varphi_\exists$$ □

**2.2.05 Corollary.** *([Pin18, Corollary IV.2.7]) If $L$ is recognized by $M \xrightarrow{\varphi} N$, and $X_0 \subseteq X_1 \subseteq M$ satisfy $L = X_0 - X_1$, then $(X_1 - X_0)\varphi_\exists = X_1\varphi_\exists - X_0\varphi_\exists$.*

**Proof.** Since $X_0 \subseteq X_1$ the condition $L = X_1 - X_0$ is equivalent to $X_0 = X_1 - L$, Lemma 2.2.04 now yields

$$X_0\varphi_\exists = (X_1 - L)\varphi_\exists = X_1\varphi_\exists - L\varphi_\exists$$

and consequently, $L\varphi_\exists = X_1\varphi_\exists - X_0\varphi_\exists$. □

As the following result shows, the notion of strict automaton actually is redundant when using monoids and can be "abstracted away" by the notion of recognition, which does not require concepts outside of the category ***mon*** and the lifted power-set monad.

**2.2.06 Proposition.** *Any monoid homomorphisms* $M \xrightarrow{\varphi} N$ *and any subset* $F \subseteq N$ *determine a strict* $M$*-automaton that recognizes the pre-image* $F\varphi^{\leftarrow}$. *If* $N$ *is finite, so is the automaton.*

**Proof.** Observe that $\langle N, \cdot, e\rangle$ can be embedded into $\langle N, N\rangle\boldsymbol{set}$ by $n \mapsto - \cdot n$. Call this embedding $\iota$. Setting $Q := N$ and $q_o := e$ we obtain a completely deterministic automaton $\mathcal{A} := \langle N, \varphi \cdot \iota, e, F\rangle$ with $\mathcal{A}\mathscr{L} = \{\, m \in M \,:\, m\varphi \in F\,\}$. □

**2.2.07 Remark.** The automaton just constructed has a built-in initial state, whereas any state of an $\boldsymbol{M}$-automaton according to Definition 2.5.00 can be the initial state. In fact, in the traditional theory of regular languages it is useful to assign to each state of a completely deterministic automaton the language recognized by this state as initial state.

If $n_0 \in N - \{e_N\}$ were chosen as initial state, one could ask, which elements $n \in N$ satisfy $n_0 \cdot n \in F$ and form their pullback. The notion of "left quotient" in Section 2.3 (see Definition 2.3.06 below) will address this question.

Alternatively, one might consider relaxing the requirement on $\delta$ to preserve the unit and use a semi-group homomorphism instead. Then the (idempotent!) image of $e_M$ might be a good initial state.

It is not clear how to adapt the construction above can in case $M$ is a semi-groups.

On the other hand, the co-algebraic view of automata theory even avoids specifying an initial state.

**2.2.08 Definition.** A subset $L$ of a monoid $M$ is called *recognizable*, if it can be recognized by a monoid homomorphism with finite codomain. The class of all recognizable subsets of $M$ is denoted by $M\mathrm{R{\scriptsize EC}}$.

The class $\mathrm{R{\scriptsize EC}}$ of recognizable subsets of monoids is the inverse image closure of all subsets of finite monoids. ◁

**2.2.09 Remark.** Again, the only finiteness property available concerns the codomain of the recognizing monoid homomorphism. Also notice that without size constraint the concept of recognizability is meaningless, as one could always use the identity morphism on $M$. However, it could be interesting to restrict the allowed codomains to countable monoids.

**2.2.10 Example.** Consider the rational language $\{a, b\}^*\{a\} \subseteq \{a, b\}^*$ and the following monoid

$$N = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \right\}$$

The homomorphism induced by

$$a \mapsto \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \qquad \text{and} \qquad b \mapsto \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$$

recognizes $L$ as the inverse image of the second matrix. ◁

## 2.3 Closure properties of Rec

The following closure properties of Rec are immediate.

**2.3.00 Proposition.**

(0) *Recognizable sets are closed under inverse images.*

(1) *Recognizable sets are closed under complementation.*

(2) *For every monoid* $\langle M, \cdot, e\rangle$ *the set* $M$ *(and hence also* $\emptyset$ *) is recognizable.*

(3) *For every finite alphabet* $X$ *the recognizable subsets of* $X^*$ *are precisely the regular languages over* $X$ *.* □

**2.3.01 Proposition.** *For every monoid* $M$ *the set* $M$Rec *is closed under binary unions and binary intersections.*

**Proof.** Suppose $M \xrightarrow{\varphi_i} N_i$ with finite codomain recognizes $L_i \subseteq M$ by means of $P_i \subseteq N_i$, $i < 2$. The induced morphism $M \xrightarrow{\langle\varphi_0,\varphi_1\rangle} N_0 \times N_1$ still recognizes the same languages as pre-images of $P_0 \times N_1$, respectively, $N_0 \times P_1$, the pre-images of $P_0$ and $P_1$ along the appropriate projection $N_0 \times N_1 \xrightarrow{\pi_i} N_i$, $i < 2$. But $(N_0 \times N_1)P \xrightarrow{\langle\varphi_0,\varphi_1\rangle^{\leftarrow}} MP$ is both right and left adjoint and hence preserves intersections as well as unions. □

Here the crucial idea was to find a single morphism with finite codomain that recognized both languages.

**2.3.02 Corollary.** *For every monoid* $M$ *the set* $M$Rec *forms a Boolean sub-algebra of* $MP$ *.* □

**2.3.03 Theorem.** $L \subseteq M_0 \times M_1$ *is recognizable iff* $L$ *is a finite union of sets of the form* $L_0 \times L_1$ *with* $L_i \subseteq M_i$ *recognizable.*

**Proof.** The recognizability of $L_0 \times L_1$ is trivial, just consider the product of the codomains of the morphisms recognizing $L_i$, $i < 2$. Moreover $(M_0 \times M_1)$Rec is closed under finite unions.

Conversely, suppose $M_0 \times M_1 \xrightarrow{\varphi} N$ recognizes $L \subseteq M_0 \times M_1$. Compose $\varphi$ with the canonical left inverses $M_i \xrightarrow{\sigma_i} M_0 \times M_1$ to the projections (that pairs elements of $M_i$ with the neutral element of the other monoid) to obtain $M_i \xrightarrow{\beta_i} N$, $i < 2$. Composing $M_0 \times M_1 \xrightarrow{\beta := \beta_0 \times \beta_1} N \times N$

with the binary composition $N \times N \xrightarrow{\cdot} N$ (which is a monoid homomorphism because of the associativity) recovers $\varphi$:

$$\langle m_0, m_1\rangle(\beta \cdot \cdot) = m_0\beta_0 \cdot m_1\beta_1 = \langle m_0, e_1\rangle\varphi \cdot \langle e_0, m_1\rangle\varphi = \langle m_0, m_1\rangle\varphi$$

Therefore we can pull $P \subseteq N$ back along $\varphi$ in two stages, first along $N \times N \xrightarrow{\cdot} N$, which results in $Q \subseteq N \times N$, and then along $\beta$, which yields $L$. But $Q$ has an explicit description:

$$Q = \{\, \langle n_0, n_1\rangle \in N \times N \;:\; n_0 \cdot n_1 \in P \,\}$$

Since inverse image functions are left adjoint and hence preserve unions, we get

$$L = \bigcup \{\, \langle n_0, n_1\rangle\beta^{-1} \;:\; n_0, n_1 \in N \;\wedge\; n_0 \cdot n_1 \in P \,\}$$

If $N$ is finite, so is $N \times N$, and hence the sets $\langle n_0, n_1\rangle\beta^{-1} = n_0\beta_0^{-1} \times n_1\beta_1^{-1}$ are recognizable. □

As a quantale $MP$ is automatically monoidal closed, *cf.* Example, 5.10.07. Hence besides set-multiplication

$$J \cdot K = \{\, j \cdot k \;:\; j \in J \;\wedge\; k \in K \,\}$$

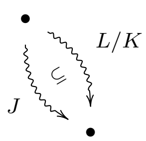
we have *residuation* operations

$$J\backslash L = \{\, m \in M \;:\; J \cdot m \subseteq L \,\} \quad \text{and} \quad L/K = \{\, m \in M \;:\; m \cdot K \subseteq L \,\}$$

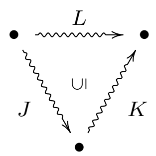
For subsets $J$, $K$, and $L$ of $M$ the following are equivalent:

$$K \subseteq J\backslash L \quad \text{and} \quad J \cdot K \subseteq L \quad \text{and} \quad J \subseteq L/K \tag{2.3-00}$$

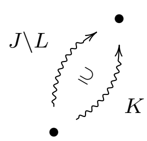
To work inside these closed categories we will use wavy arrows $\rightsquigarrow$ for sub-sets of $M$, to avoid confusion with the 1-cells of ***rel***. Hence Equation 2.3-00 takes the form



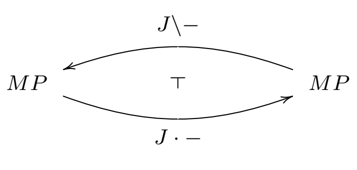

and



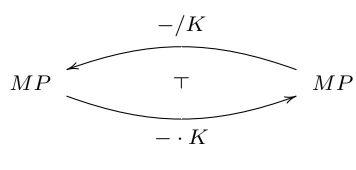

and



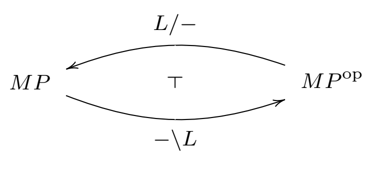

Fixing sets $J$, and $K$ gives adjunctions




and




Consequently, fixing $L$ yields a polarity



Therefore both $J\backslash-$ and $-/K$ as right adjoints preserve intersections, while $K\cdot-$ and $-\cdot K$ as left adjoints preserve unions. Furthermore, both $L/-$ and $-\backslash L$ map unions to intersections.

Which of these operations preserve recognizable sets? Since at least for $M = X^*$ the language $\{\varepsilon\}$ is recognizable, the operations $J\cdot-$ and $-\cdot K$ can at most preserve this property, if $J$, resp., $K$ itself is recognizable. Unfortunately, even this does not suffice to guarantee the preservation of recognizability in general, see Example 2.3.12 below. In case of free monoids, this problem disappears:

**2.3.04 Proposition.** *The set-product over free monoids preserves recognizability.*

**Proof.** Consider monoid homomorphisms $X^* \xrightarrow{\varphi_i} N_i$ with $N_i$ finite, and subsets $F_i \subseteq N_i$ with $L_i = F_i\varphi^{\leftarrow}$, $i < 2$. Interpret $N_i$ as the state set of a strict $X^*$-automaton $\mathcal{A}_i$ with initial state $e_i$, final state set $F_i$ and transition assignment $X^* \xrightarrow{\delta_i} N_i$, as spelled out in Proposition 2.2.06. Construct a new automaton $\mathcal{A}$ that recognizes $L_0\cdot L_1$ with state set $Q := N_0 + N_1$, initial state $e_0$ and final state set $F_1$, whose transition assignment is the uniquely determined homomorphic extension $\bar{\delta}$ to $X^*$ of $X \xrightarrow{\delta} \langle N_0+N_1, N_0+N_1\rangle\boldsymbol{rel}$ given by

$$a\delta := a\delta_0 + a\delta_1 + \{\, \langle f,q\rangle \in F_0 \times N_1 \,:\, e_1(a\delta_1) = q\,\} \quad \text{for } a \in X$$

Unless $F_0 = \emptyset$, in which case $L_0 = \emptyset = L_0\cdot L_1$, this automaton is not strict. In case of infinite $X$ it is not lax either, since then $\delta$ does not have finite support. Nevertheless, the familiar power-set construction still is applicable and yields a strict automaton $\mathcal{A}'$ with finite state set $(N_0+N_1)P$ that recognizes $L_0\cdot L_1$. □

It seems tedious to perform this construction w/o resorting to automata. Fortunately, residuations are better behaved:

**2.3.05 Proposition.** *For any subsets $J, K \subseteq M$ the residuations $J\backslash-$ and $-/K$ preserve recognizability.*

**Proof.** Suppose $L \subseteq M$ is recognized by $M \xrightarrow{\varphi} N$ with finite codomain. In particular, by Proposition 2.2.03 $L$ is a fixed point of $\varphi_\exists\varphi^{\leftarrow}$.

We claim that $J\backslash L$ is the $\varphi$-inverse image of $J\varphi_\exists\backslash L\varphi_\exists$. Since the monoid homomorphism $\varphi_\exists$ is left adjoint to the lax homomorphism $\varphi^{\leftarrow}$, we get

$$J\cdot(J\varphi_\exists\backslash L\varphi_\exists)\varphi^{\leftarrow} \subseteq J\varphi_\exists\varphi^{\leftarrow}\cdot(J\varphi_\exists\backslash L\varphi_\exists)\varphi^{\leftarrow} \subseteq (J\varphi_\exists\cdot J\varphi_\exists\backslash L\varphi_\exists)\varphi^{\leftarrow} \subseteq L\varphi_\exists\varphi^{\leftarrow} = L$$

which by definition of $J\backslash L$ implies $(J\varphi_\exists\backslash L\varphi_\exists)\varphi^{\leftarrow} \subseteq J\backslash L$.

On the other hand, $\varphi_\exists$ preserves the inclusion $J\cdot J\backslash L \subseteq L$, which by definition of $J\varphi_\exists\backslash L\varphi_\exists$ implies $(J\backslash L)\varphi_\exists \subseteq J\varphi_\exists\backslash L\varphi_\exists$ and therefore $J\backslash L \subseteq (J\backslash L)\varphi_\exists\varphi^{\leftarrow} \subseteq (J\varphi_\exists\backslash L\varphi_\exists)\varphi^{\leftarrow}$.

The argument for post-residuation with $K$ is dual. □

Classical semi-group theory considers further operations on the quantale $MP$ induced by subsets of $M$.

**2.3.06 Definition.** Given $J, K \subseteq M$, the *left*, resp., *right quotient* of $L \subseteq M$ is defined by

$$J^{-1}L := \{\, m \in M \,:\, J \cdot m \cap L \neq \emptyset \,\} \quad \text{and} \quad LK^{-1} := \{\, m \in M \,:\, L \cap m \cdot K \neq \emptyset \,\} \qquad \triangleleft$$

Unfortunately, the terminology conflicts with real quotients, *i.e.*, co-equalizers of kernel pairs, or just surjective homomorphisms.

**2.3.07 Remark.** Despite a superficial similarity of this notion with residuation, it is not immediately clear, what the categorical significance of quotients is, if any.

Observe that if $J \subseteq M$ only consists of invertible elements, *e.g.*, if $M$ is a group, then $J^{-1}L = \{\, j^{-1} \,:\, j \in J \,\} \cdot L$. A dual argument applies to $K$.

On the other hand, interpreting the subsets $J$ and $L$ of $M$ as relations $1 \longrightarrow M$, and their $\varphi$-images as relations $1 \longrightarrow N$, the defining condition of $J^{-1}L$ can be expressed in terms of composite relations as

$$(1 \xrightarrow{J} M \xrightarrow{- \cdot m} M \xrightarrow{L^{\mathrm{op}}} 1) = (1 \xrightarrow{J\varphi_\exists} N \xrightarrow{- \cdot m\varphi} N \xrightarrow{(L\varphi_\exists)^{\mathrm{op}}} 1) = 1 \xrightarrow{1} 1$$

With the final state set $F := L\varphi_\exists$ this is equivalent to

$$\exists j \in J.\ (1 \xrightarrow{j\varphi} N \xrightarrow{- \cdot m\varphi} N \xrightarrow{F^{\mathrm{op}}} 1) = 1 \xrightarrow{1} 1$$

Then $J^{-1}L$ can be seen as the union of those sets that are recognized by the $M$-automaton for $L$ that has been modified by shifting the initial state forward from $e$ to $j\varphi$, $j \in J$.

In case of $LK^{-1}$, instead we could consider whether or not

$$(1 \xrightarrow{e} N \xrightarrow{- \cdot m\varphi} N \xrightarrow{(K\varphi_\exists \cdot F)^{\mathrm{op}}} 1) = 1 \xrightarrow{1} 1$$

Now the automaton for $L$ is modified by shifting the final states in $F$ backwards by multiplying them with elements of $K\varphi_\exists$ from the left. Here it is even more compelling that sets recognized by such modified automata must be recognizable.

Finally notice that in case of a singleton "denominator" the notions of residuation and of quotient agree: given $j \in M$

$$j\backslash L = \{\, m \in M \,:\, j \cdot m \in L \,\} = \{\, m \in M \,:\, \{j \cdot m\} \cap L \neq \emptyset \,\} = j^{-1}L$$

where we have dropped the set-brackets from the singletons. From this base case residuations are obtained by intersections, while quotients are obtained by unions:

$$J\backslash L = \bigcap \{\, j\backslash L \,:\, j \in J \,\} \quad \text{and} \quad J^{-1}L = \bigcup \{\, j^{-1}L \,:\, j \in J \,\}$$

and dually for $K$. $\triangleleft$

**2.3.08 Proposition.** *If $L \subseteq M$ is recognizable, so are all left and right quotients of $L$. In fact they are recognized by the same homomorphism that recognizes $L$.*

**Proof.** By the proof of Proposition 2.3.05 and the observation above the left and right quotients with respect to singletons subsets of $M$ are recognized by any $M \xrightarrow{\varphi} N$ that recognizes $L$. Furthermore,

$$J^{-1}L = \bigcup\{ j^{-1}L : j \in J \} = \bigcup\{ j\backslash L : j \in J \} = \bigcup\{ j\varphi\backslash L\varphi_\exists : j \in J \}$$

actually is a finite union and hence recognizable by Proposition 2.3.01. An analogous argument works for $LK^{-1}$. □

**2.3.09 Lemma.**

(0) *If $M$ is finite, every subset is recognizable.*

(1) *A group $\langle G, \cdot, e \rangle$ is finite iff $\{e\}$ is recognizable.*

**Proof.**

(0) Trivial.

(1) If $G$ is finite, the claim follows from (1). Suppose $G \xrightarrow{\varphi} N$ with finite codomain recognizes $\{e\}$. Wolog assume $\varphi$ to be surjective, then $\varphi$ is a group-homomorphism. If it is not injective, $x \neq y$ exist in $G$ with $x\varphi = y\varphi$. But then $e \neq x \cdot y^{-1} \in \{e_N\}\varphi^{\leftarrow}$, contradiction. Hence $\varphi$ must be an isomorphism, and therefore $G$ is finite. □

**Counter-examples**

**2.3.10 Example.** Finite non-empty subsets of monoids need not be recognizable: By Lemma 2.3.09 the set $\{0\} \subseteq \mathbb{Z}$ is not recognizable. In fact, no finite subset of $\mathbb{Z}$ has this property: since the cosets of any kernel of a group-homomorphism are isomorphic as sets, in case of a finite codomain they must all be infinite. But by Proposition 2.2.03 recognizable sets must be unions of cosets. ◁

**2.3.11 Example.** Recognizable sets need not be closed under direct images. Consider the homomorphism $\{a, b\}^* \xrightarrow{\varphi} \mathbb{Z}$ induced by the function $\{a, b\} \xrightarrow{f} \mathbb{Z}$ with $af = 1$ and $bf = -1$. By Kleene's Theorem every finite subset of $\{a, b\}^*$ is recognizable, but by Example 2.3.10 no $\varphi$-image of a non-empty finite subset has this property. ◁

Any idempotent element $i$ of a semi-group $S$ locally serves as neutral element of some sub-semi-group of $S$, which therefore is a monoid. $\{i\}$ is the smallest such, and the supremum $S[i]$ of all sub-semi-groups in which $i$ is neutral, also has this property. However, even if $S$ is a monoid, its global neutral element $e$ can differ from $i$, hence $S[i]$ need not be a sub-monoid, call it a "local monoid" instead. Of course, the invertible elements of $S[i]$ then form a "local group". Conversely, for every local monoid of $S$ its neutral element is idempotent in $S$.

We would like to collect all maximal local monoids and hence all maximal local groups of a given semi-group in transparent fashion. However, this requires "thinking outside the box", since the Karoubian envelope (Definition 5.8.03) of a semi-group may no longer be a semi-group, but it will always be a category. This may also be useful in the section on Green's relations.

For a monoid $M$ we clearly have $\langle e,e\rangle\tilde{M} = M$, while all the other hom-sets are subsets of $M$. The following example uses this fact.

**2.3.12 Example. (Shmuel Winograd)** Recognizable sets need not be closed under the binary operation and under Kleene star.

Extend $\mathbb{Z}$ by first adding a self-inverse element $a$ that otherwise behaves like a second copy of $0 \in \mathbb{Z}$:

$$a + a = 0 \quad \text{and} \quad a + n = n \quad \text{for all } n \in \mathbb{Z}$$

Observe that $0$ seizes to be neutral, but stays idempotent, while $a$ is neither neutral nor idempotent. Then add a new neutral element $e$. The resulting monoid $M := \mathbb{Z} + \{a, e\}$ has two idempotent elements, $e$ and $0$.

The category $\tilde{M}$ has the following hom-sets:

$$\langle 0,0\rangle\tilde{M} = \mathbb{Z} \quad , \quad \langle 0,e\rangle\tilde{M} = \langle e,0\rangle\tilde{M} = \mathbb{Z} + \{a\} \quad \text{and} \quad \langle e,e\rangle\tilde{M} = M$$

Hence $\mathbb{Z}$ is a local group in $M$ that fails to be a sub-monoid.

**Claim**: Every $L \in M\text{REC}$ satisfies $L \cap \mathbb{Z} \in \mathbb{Z}\text{REC}$.

Suppose $L$ is the inverse image of $F \subseteq N$ under $M \xrightarrow{\varphi} N$ with finite codomain. Consider the functor $\tilde{M} \xrightarrow{\tilde{\varphi}} \tilde{N}$, which induces a monoid homomorphism

$$\mathbb{Z} = \langle 0,0\rangle\tilde{M} \xrightarrow{\langle 0,0\rangle\tilde{\varphi}} \langle 0\varphi, 0\varphi\rangle\tilde{N}$$

that operates on the elements of $\mathbb{Z}$ by applying $\varphi$. Clearly, the inverse image of $\langle 0\varphi, 0\varphi\rangle\tilde{N} \cap F$ is the desired set $\mathbb{Z} \cap L$, which therefore is recognizable.

**Claim**: $\{a\}$ is recognizable in $M$.

Define $N := \{e, a, z\}$ with neutral element $e$, absorbing element $z$ and $a + a := z$. The obvious morphism $M \xrightarrow{\varphi} N$ preserves $e$ and $a$ and maps every $n \in \mathbb{Z}$ to $z$.

**Conclusion**: neither $\{a\} + \{a\} = \{0\}$ nor $\{a\}^* = \{e, a, 0\}$ is recognizable, since the intersections with $\mathbb{Z}$ are finite. $\lhd$

**2.3.13 Open Problem.** Given a morphism $M \xrightarrow{\varphi} N$, is there a Boolean algebra morphism induced by $\varphi$ linking $M\textsc{Rec}$ and $N\textsc{Rec}$? The direct image map $\varphi_\exists$ clearly does not work, as Example 2.3.11 shows.

## 2.4 Minimization: the syntactic monoid

Recall that every regular language $L \subseteq X^*$ can be recognized by a minimal completely deterministic automaton. This amounts to a minimal state set $Q$ such that the sub-monoid of the function monoid on $Q$ spanned by the transition functions recognizes $L$, via the set of endo-functions on $Q$ that map $q_0$ into $F$.

The construction of this minimal automation usually starts from some completely deterministic automaton known to recognize $L$, eliminates all states not reachable from the initial state, and then performs a factorization with respect to a suitable equivalence relation on the remaining state set.

In order to bypass completely deterministic automata, one would like to construct for a recognizable set $L \subseteq M$ directly a surjective monoid homomorphism into some minimal finite monoid ***L-syn*** that recognizes $L$ via a suitable subset.

In the categorical approach to congruences 5.13, specifically Definition 5.13.06 and Theorems 5.13.07 and 5.13.09, we have seen precisely how congruences and quotients come about in categories like ***sgr*** and ***mon***.

Given a morphism $M \xrightarrow{\varphi} N$ recognizing a sub-set $L$ of $M$, one can consider its kernel pair and the resulting congruence and quotient. According to Proposition 2.2.03, $L$ is the union of some $\sim_\varphi$-classes. This raises the question: is there a "best" or "most efficient" way of exhausting $L$ by congruence classes? The identity relation is the smallest congruence (w.r.t. $\subseteq$) and manages to exhaust $L$ by means of singletons. Hence we are interested in the largest congruence that exhausts $L$, if it exists.

**2.4.00 Lemma.** *The congruences on a monoid (semi-group) $M$ form a small complete lattice. Consequently, up to isomorphism there is only a set of quotients out of $M$ and these form a complete lattice as well that is dually isomorphic to the lattice of congruences.*

**Proof.** Arbitrary intersections of sub-algebras are again sub-algebras; in particular for $M \times M$. □

Hence we know that the desired congruence does exist. It remains to find a concise description of congruences.

Since $(M \times M)P = \langle M, M\rangle \boldsymbol{rel}$, this set carries two monoid structures: set-multiplication " $\cdot$ " and relation composition " $\star$ ", both of which are closed. When using diagrammatic reasoning, we need to say which composition the juxtaposition of 1-cell arrows refers to. In the notation used after Diagrams 2.3-00, wavy arrows $\bullet \overset{R}{\rightsquigarrow} \bullet$ correspond to relations $M \xrightarrow{R} M$.

**2.4.01 Proposition.** *For an equivalence relation $E$ on the underlying set of a semi-group $S$ the following are equivalent:*

(a) $E$ *is a congruence;*

(b) $E$ *is a sub-semi-group of* $S \times S$, i.e.,

$$\forall a, b, c, d \in S. \Big( \langle a, b \rangle, \langle c, d \rangle \in E \implies \langle a, b \rangle \cdot \langle c, d \rangle = \langle a \cdot c, b \cdot d \rangle \in E \Big)$$

*Furthermore, these conditions imply*

(c) $E$ *satisfies*

$$\forall u, v \in S. \Big( \langle u, v \rangle \in E \implies \forall x, y \in S. \langle x \cdot u \cdot y, x \cdot v \cdot y \rangle = \langle x, x \rangle \cdot \langle u, v \rangle \cdot \langle y, y \rangle \in E \Big)$$

*In case that $S$ is a monoid, all three conditions are equivalent.*

**Proof.** HW! □

Conditions (b) and (c) may be referred to as *strong* resp. *weak stability* of the equivalence relation $E$. Without reference to elements these conditions can be reformulated in $(S \times S)P$ as

$$E \cdot E \subseteq E \qquad \text{resp.} \qquad S\Delta \cdot E \cdot S\Delta \subseteq E$$

In fact, if $S$ is a monoid, $\langle e, e \rangle \in E$ implies equality in both cases. As an equivalence relation, $E$ also satisfies

$$S\Delta \subseteq E \quad , \quad E = E^{\mathrm{op}} \quad \text{and} \quad E \star E \subseteq E$$

Moreover, $E$ trivially is an order-ideal on the discretely ordered set $S$, *i.e.*

$$S\Delta \cdot E \subseteq E \supseteq E \cdot S\Delta \quad \text{or equivalently} \quad S\Delta \cdot E \cdot S\Delta \subseteq E$$

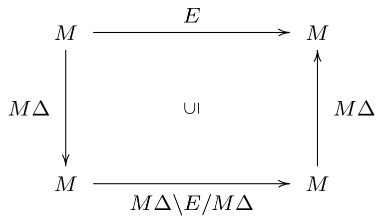
The first condition resembles the characterization of transitivity, while the second condition resembles the definition of an order-ideal (see Definition 6.0.00), with the order being discrete ( $S\Delta$ ). In both cases the relation product is replaced by the multiplication of subsets of $S \times S$. (However, order ideals do not have to be equivalence relations. This may suggest how to define congruences for ordered semi-groups and monoids below.)

**2.4.02 Remark.** While strong stability as a characterization of congruences directly generalizes to any type of set-based algebra by extending the given condition to all operations, weak stability is rather specific for monoids and heavily relies on the existence of a neutral element. To characterize congruence of a semigroup by a similar condition as (c), the second quantification over $x$ and $y$ has to extend over $S^1$, see Remark 1.1.03. It is presently (2018-11-28) not clear, what else the categorically questionable definition of $S^1$ is good for.

**2.4.03 Proposition.** *For every equivalence relation* $E \subseteq M \times M$ *on the underlying set of a monoid, its two-sided residuation* $M\Delta\backslash E/M\Delta$ *with respect to the diagonal is the largest congruence contained in* $E$ *.*

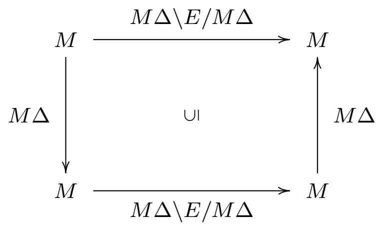
**Proof.** By definition of $\Delta_M\backslash E/\Delta_M$ (see Lemma 5.10.03) we have

[wrt. $\cdot$ ]




Since $M\Delta \cdot M\Delta \subseteq M\Delta$ , we obtain the desired stability condition:

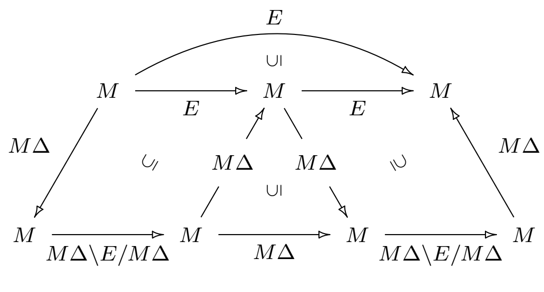
[wrt. $\cdot$ ]




It remains to show that $M\Delta\backslash E/M\Delta$ is an equivalence relation. The reflexivity immediately follows from $M\Delta \cdot M\Delta \cdot M\Delta \subseteq M\Delta \subseteq E$ , while the symmetry results from $E = E^{\mathrm{op}}$ and $M\Delta\backslash E^{\mathrm{op}}/M\Delta = (M\Delta\backslash E/M\Delta)^{\mathrm{op}}$ .

Any proof of transitivity necessarily has to combines both the relation product and the set-multiplication. But the proof below only works for interpolative semi-groups, hence in particular monoids, but not for general semi-groups, whereas all previous arguments did. Consider

[wrt. $\cdot$ and $\star$ ]




where the horizontal arrows are composed via " ; ", while the inclined arrows are composed with others via " $\cdot$ ". Notice that the inclusion $M\Delta \subseteq M\Delta \cdot M\Delta$ does require the multiplication $M \times M \overset{\cdot}{\longrightarrow} M$ to be surjective. The combined inclusion by the universal property of $M\Delta\backslash E/M\Delta$ now induces the desired inclusion $(M\Delta\backslash E/M\Delta) \cdot (M\Delta\backslash E/M\Delta) \subseteq M\Delta\backslash E/M\Delta$ that establishes transitivity.

Finally, consider a congruence $W \subseteq E$ . Since $M\Delta \cdot W \cdot M\Delta \subseteq W$ , we immediately have $W \subseteq M\Delta\backslash E/M\Delta$ . □

**2.4.04 Open Problem.** Given relations $R$, $S$, and $T$ on $M$, are the composites $R\cdot(S;T)$ and $(R\cdot S);T$ comparable by means of $\subseteq$?

**2.4.05 Definition.** The *syntactic congruence* of $L\subseteq M$ is defined by

$$\sim_{\mathcal{L}} := M\Delta\backslash L\text{-}\boldsymbol{ker}/M\Delta$$

where $L\text{-}\boldsymbol{ker}$ is the kernel pair in $\boldsymbol{set}$ of the characteristic function $M \xrightarrow{L\chi} \mathbf{2}$.

The factor monoid $L\text{-}\boldsymbol{syn} := M/\sim_{\mathcal{L}}$ is known as the *syntactic monoid* of $L$, while the canonical surjection $M \xrightarrow{\eta_{\mathcal{L}}} M/\sim_{\mathcal{L}}$ is called the *syntactic quotient*. ◁

It remains to establish the universal property of $\sim_{\mathcal{L}}$, or equivalently $\eta_{\mathcal{L}}$.

**2.4.06 Proposition.** *Any quotient of $M$ that recognizes $L\in M$ factors through $\eta_{\mathcal{L}}$. In other words, $\sim_{\mathcal{L}}$ indeed is the largest (w.r.t. $\subseteq$) congruence such that $\eta_{\mathcal{L}}$ recognizes $L$.*

**Proof.** Since $\sim_{\mathcal{L}}\subseteq L\text{-}\boldsymbol{ker} = L\times L + (M-L)\times(M-L)$, both $L$ and $M-L$ are unions of $\sim_{\mathcal{L}}$-classes, by Proposition 2.2.03 $\eta_{\mathcal{L}}$ recognizes $L$.

Suppose $M \xrightarrow{\varphi} N$ is regular epi (and hence surjective) and also recognizes $L$. Its kernel pair in $\boldsymbol{mon}$ is a congruence, hence by Proposition 2.2.03 satisfies $\varphi\text{-}\boldsymbol{ker}\subseteq L\text{-}\boldsymbol{ker}$. Since residuation preserves order, we get

$$\varphi\text{-}\boldsymbol{ker} = M\Delta\backslash\varphi\text{-}\boldsymbol{ker}/M\Delta \subseteq \sim_{\mathcal{L}}$$

This gives the desired factorization of $\varphi$ through $L\eta$. By construction, $M/\sim_{\mathcal{L}}$ is also a quotient of $N$. □

If $M \xrightarrow{\varphi} N$ recognizes $L\subseteq M$, we can always take the (regular epi,mono)-factorization $M \xrightarrow{e} E \xrightarrow{m} N$ of $\varphi$. Consequently, $e$ factors through $L\eta$.

The following Proposition is awkward as it deliberately avoids morphisms in favor of fancy terminology borrowed from group theory. Moreover, it is not clear why this result has been formulated only in the special case of $M = A^*$, as Pin does not restrict the concept of a monoid "dividing" another to free dividends. Perhaps a more suggestive name might be *partial quotient*.

**2.4.07 Proposition.**

(0) *A monoid $N$ recognizes $L\subseteq M$ iff $L\text{-}\boldsymbol{syn}$ is the quotient of some sub monoid of $N$ ($L\text{-}\boldsymbol{syn}$ "divides" $N$).*

(1) *If $N$ recognizes $L\subseteq M$ and "divides" $N'$, then $N'$ also recognizes $L$ as a quotient of $E$ rather than $N$.* □

For recognizable languages $L$ over a finite alphabet $X$ we have already mentioned that the syntactic monoid $L$-***syn*** ought to be the transition monoid of the minimal automaton that recognizes $L$. Pin provides such a proof.

## 2.5 Lax $M$-automata, rational sets and first closure properties

Ignoring $\langle Q, Q\rangle\boldsymbol{prt}$ as a potential target for a transition assignment function, we will focus instead on the finite ordered monoid $\langle Q, Q\rangle\boldsymbol{rel}$ as the other possible type of codomain for $\delta$. As this automatically is a complete lattice and hence a quantale, any function $M \xrightarrow{\delta} \langle Q, Q\rangle\boldsymbol{rel}$ can be saturated to yield a lax homomorphism $\tilde{\delta}$. In fact, nothing prevents us from considering ordered monoids $M$, where we will be interested in recognizing upper segments.

Since $\langle Q, Q\rangle\boldsymbol{rel}$ has a least element $\emptyset$, one can impose a second finiteness constraint: namely on the support of $\delta$ before saturation. This simply is not available for strict $M$-automata. In case of finitely generated ordered monoids $M$ this is will not constrain the lower segments that can be recognized in this fashion, and so in the discrete case we expect lax $M$-automata to be more capable than strict ones, see Theorem 2.8.03. However, in the non-finitely generated case lax automate will not be able to recognize the set $M$ itself, in contrast to what strict automata can do.

**2.5.00 Definition.** For $M \in \boldsymbol{omon}$, a *lax* $M$*-automaton* $\mathcal{A} = \langle Q, \delta, I, F\rangle$ consists of

- ▷ a set $Q$ of *states*;
- ▷ a monotone function $M \xrightarrow{\delta} \langle Q, Q\rangle\boldsymbol{rel}$ assigning *transition relations* on $Q$ to the elements of $M$;
- ▷ sets of *initial* and *final* states $I, F \subseteq Q$.

$\mathcal{A}$ is called *finite*, if $Q$ is finite and $\delta$ has *finite support*, *i.e.*, $(\langle Q, Q\rangle\boldsymbol{rel} - \{\emptyset\})\delta^{\leftarrow}$ is finite. ◁

While the function $M \xrightarrow{\delta} \langle Q, Q\rangle\boldsymbol{rel}$ can of course be extended to a monoid homomorphism on $M^*$, in general this does not yield a homomorphism from $M$ to $\langle Q, Q\rangle\boldsymbol{rel}$ (the unit $M \xrightarrow{M\eta} M^*$ of the list monad is a monoid homomorphism iff $M$ is free). But we always can find a canonical lax homomorphism $M \xrightarrow{\tilde{\delta}} \langle Q, Q\rangle\boldsymbol{rel}$. This construction works for arbitrary unital quantales $U$ in place of $\langle Q, Q\rangle\boldsymbol{rel}$, and even in case that $M = \langle M, \xi, \sqsubseteq\rangle$ is an ordered monoid.

**2.5.01 Definition.** Consider an monotone function $M \xrightarrow{\psi} \Omega$ from a ordered monoid into a unital quantale. Its *saturation* $\tilde{\psi}$ is the lax homomorphism given by

$$\begin{array}{ccccccc} M\boldsymbol{U} & \xrightarrow{\xi^{\leftarrow}} & M^*\boldsymbol{U} & \xrightarrow{\psi^*\boldsymbol{U}} & \Omega^*\boldsymbol{U} & \xrightarrow{\zeta_{\exists}} & v\boldsymbol{U} \\ {\scriptstyle M\eta}\uparrow & & & & & & \downarrow{\scriptstyle \bigsqcup} \\ M & & & \xrightarrow[\tilde{\psi}]{} & & & \Omega \end{array} \qquad (2.5\text{-}00)$$

where the order on $M^*$ and $\Omega^*$ is component-wise, $\xi$ and $\zeta$ denote the the structure morphisms of monoids $M$ and $\Omega$ as EM-algebras, and the inverse image map $M\boldsymbol{U} \xrightarrow{\xi^{\leftarrow}} M^*\boldsymbol{U}$ is the only lax homomorphism (*cf.*, Proposition **??**), while all other components are strict.

Concretely, $\tilde{\psi}$ maps $m \in M$ to the supremum of all composites $m_0\psi \cdot m_1\psi \cdot \dots \cdot m_{n-1}\psi$, where $m_0, m_1, \dots, m_{n-1}$ runs through all possible decompositions of $m$ in $M$. Note that repeated occurrences of $e \in M$ are allowed. In the classical setting for a free monoid $M$ the empty word is indecomposable. This allows the elimination of silent transitions by means of absorbing the reflexive transitive hull of their underlying relation to be absorbed into the other transition relations. It is not clear how to adapt this construction to general monoids. While it is clear how to handle invertible elements, elements with just a one-sided inverse could be problematic.

Conceptually, the notion of acceptance is very easy:

**2.5.02 Definition.** A lax $M$-automaton $\mathcal{A} = \langle Q, \delta, I, F\rangle$ *accepts* $w \in M$ iff $\mathbf{1} \xrightarrow{I} Q \xrightarrow{m\tilde{\delta}} Q \xrightarrow{F} \mathbf{1}$ is not empty. The subset recognized by $\mathcal{A}$ is (again) denoted by $\mathcal{A}\mathscr{L}$, which is also the $\tilde{\delta}$-pre-image of $F' := \{\, Q \xrightarrow{r} Q : I \cdot r \cdot F \neq \emptyset \,\}$, which is a *proper upper segment* of $\langle Q, Q\rangle\boldsymbol{rel}$, *i.e.*, it is up-closed and does not contain $\emptyset$.

**2.5.03 Remarks.**

(0) Again in practical terms this means that $m \in \mathcal{A}\mathscr{L}$ iff there exists some decomposition $m = m_0 \cdot \dots \cdot m_{n-1}$ with non-empty composition

$$\mathbf{1} \xrightarrow{I} Q \xrightarrow{m_0\delta} \dots \xrightarrow{m_{n-1}\delta} Q \xrightarrow{F} \mathbf{1}$$

(1) Why did we not require the function $M \xrightarrow{\delta} \langle Q, Q\rangle\boldsymbol{rel}$ in Definition 2.5.00 to be a lax homomorphism to start with? In that case we need to introduce a notion of "finitely supported lax homomorphism" in order to handle this aspect of finiteness, which is not preserved by saturation. It then becomes a matter of taste, where the saturation should be mentioned.

Conceptually, we now proceed as we did for recognizable sets: we replace $\langle Q, Q\rangle\boldsymbol{rel}$ by some unital quantale $\Omega$, *cf.*, Example 5.10.07.

**2.5.04 Definition.** An upper segment $L$ of an ordered monoid $M$ is said to be *recognized by a function* $M \xrightarrow{\psi} \Omega$ into a unital quantale $\Omega$, if $L$ belongs to the image of

$$(\Omega - \{\bot\})\boldsymbol{U} \xrightarrow{e\tilde{\psi}\backslash -} \Omega\boldsymbol{U} \xrightarrow{\tilde{\psi}^{\leftarrow}} M\boldsymbol{U}$$

$L \subseteq M$ is called *rational*, if $L$ is recognized by a function $M \xrightarrow{\psi} \Omega$ with finite codomain and finite support, *i.e.*, $(\Omega - \{\bot\})\psi^{\leftarrow}$ is finite. ◁

**2.5.05 Remark.** We will implicitly use the bijective correspondence between upper segments of $\Omega - \{\bot\}$ and *proper upper segments* of $\Omega$, namely those that do not contain $\bot$.

Due to the laxness of $\tilde{\psi}$ we cannot use $e_\Omega$ as "initial state" in $\Omega$ but have to shift it to $e\tilde{\psi}$ above $e_\omega$, *cf.*, Remark 2.3.07. Hence not necessarily all upper segments of $\Omega$ come into play with respect to $\psi$. This issue disappears if $\psi$ is normal, *i.e.*, if $e\psi = e_\Omega$.

**2.5.06 Proposition.** *([Pin18, Proposition IV.1.3]) Every rational upper segment of $M$ is the up-closure of some rational upper segment in some finitely generated sub-monoid $M'$ of $M$.*

**Proof.** Suppose $L \subseteq M$ is recognized by $M \xrightarrow{\psi} \Omega$ with finite codomain and finite support and a proper upper segment $F \in \Omega\boldsymbol{U}$. Consider the sub-monoid $M'$ generated by those $m \in M$ with $m\psi \neq \bot$, and let $L' \in M'\boldsymbol{U}$ be the pre-image of $e\tilde{\psi}\backslash F$. By construction we now have $L = L'{\downarrow}$. □

**2.5.07 Proposition.** *If $L \in M\boldsymbol{U}$ is recognized by a lax homomorphism, then it is also recognized by a normal lax homomorphism.*

**Proof.** A lax homomorphism $M \xrightarrow{\rho} \Omega$ that recognizes $L$ uniquely determines a normal lax functor $M \xrightarrow{\bar{\rho}} \Omega\text{-}\boldsymbol{mnd}_{\mathrm{m}}$. Since $M$ has just one object, this factors through the full subcategory $\bar{\Omega}$ of $\Omega\text{-}\boldsymbol{mnd}_{\mathrm{m}}$ spanned by the monad given by $e\rho$ in $\Omega$. Now $m\bar{\rho}$ has to be the endo-"order ideal" on $e\rho$ generated by $m\rho$, namely $e\rho \cdot m\rho \cdot e\rho$.

Suppose $L$ is the inverse image of $e\rho\backslash F$ for some proper upper segment $F \in \Omega\boldsymbol{U}$. Then $e\omega \cdot F \cdot e\rho$ is a proper upper segment of endo-"order ideals" on $e\rho$, whose inverse image under $\bar{\rho}$ is $L$. □

**2.5.08 Remark.** In the classical case of $\Omega = \langle Q, Q\rangle\boldsymbol{rel}$ this construction results in quantale $\bar{\Omega} = \langle e\tilde{\delta}, e\tilde{\delta}\rangle\boldsymbol{idl}$ of order-ideals on the pre-order $e\tilde{\delta}$, which also is the neutral element. If $e \in M$ is indecomposable, this pre-order coincides with the reflexive transitive hull $(e\delta)^\star$ of the original set $e\delta$ of silent transitions. ◁

Of course, Proposition 2.2.03 applies here as well, since it does not matter whether the underlying function is a monoid homomorphism or a lax homomorphism.

It remains to show that recognition by a lax homomorphism into some quantale is equivalent to recognition by a lax automaton.

**2.5.09 Theorem.** *If* $L \in M\boldsymbol{U}$ *is recognized by a function* $M \xrightarrow{\psi} \Omega$ *into a unital quantale, then it is also recognized by a lax* $M$*-automaton.*

**Proof.** Suppose $L$ is the inverse image of $e\tilde{\psi}\backslash F \in \Omega\boldsymbol{U}$ with $F \in (\Omega - \{\bot\})\boldsymbol{U}$. In analogy to the proof of Proposition 2.2.06, set $Q := \Omega$. The initial states comprise the principal upper segment $e\tilde{\psi}\uparrow$, while the elements of $F$ serve as final states.

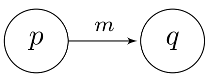
We now embed $\Omega$ into $\langle \Omega^{\mathrm{op}}, \Omega^{\mathrm{op}} \rangle \boldsymbol{idl} \subseteq \langle \Omega, \Omega \rangle \boldsymbol{rel}$ by mapping $u \in \Omega$ to the left-adjoint order-ideal $\Omega \times \Omega^{\mathrm{op}} \xrightarrow{\langle - \cdot u, - \rangle \Omega} \mathbf{2}$ generated by the order-preserving function $- \cdot u$, and $\bot$ to $\emptyset$.

For $m \in M$ any decomposition $m = m_0 \cdot \dots \cdot m_{n-1}$ the product $e\tilde{\psi} \cdot m_0\psi \cdot \dots \cdot m_{n-1}\psi$ belongs to $F$ iff the composition of the corresponding order-ideals with $1 \xrightarrow{I} \Omega^{\mathrm{op}}$ and $\Omega^{\mathrm{op}} \xrightarrow{F} 1$ is not empty.

Note that the composition $\mathbf{1} \xrightarrow{J} \Omega^{\mathrm{op}} \xrightarrow{\langle - \cdot \bot, - \rangle \Omega^{\mathrm{op}}} \Omega^{\mathrm{op}}$ for any non-empty lower segment $J \subseteq \Omega$ yields the lower segment $\{\bot\}$ of $\Omega$, which stays stable under further compositions with order-ideals of the type above and does not intersect a proper upper segment. Hence replacing these order ideals by the empty order ideal does not affect the recognized language. □

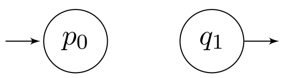
We now establish the first simple closure properties of rational sets. This can be done in entirely analogous fashion to the case of regular languages.

Think of states and transitions as a graph together with a graph homomorphism into the singleton graph with hom-set $M$. Typical notation:




for $\langle p, q \rangle \in m\delta$

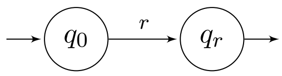
This actually is the notion of a *labeled transition system*. In addition, initial and final states are indicated by incoming arrows without source, resp., outgoing arrows without target:




Of course, the generalization to non-singleton target graphs or possibly categories is immediate if one accepts the idea of *typed alphabets*.

**2.5.10 Proposition.** *For any ordered monoid* $M$ *any finitely generated upper segment* $R \subseteq M$ *is accepted by a finite lax automaton.*

**Proof.** We use a state set $Q$ of the form $\{q_0\} + \{q_r : r \in E\}$ with $I = \{q_0\}$ and $F = \{q_r : r \in R\}$. The only transitions are given by



for $r \in R$

It is immediately clear that this automaton is finite and accepts $R$. □

**2.5.11 Proposition.** *If two sets $L_i \subseteq M$, $i < 2$, are accepted by non-deterministic automata, then so is their union $L_0 \cup L_1$.*

**Proof.** Just take the disjoint union of the automata accepting $L_0$ and $L_1$, respectively. □

**2.5.12 Proposition.** *If two sets $L_i \subseteq M$, $i < 2$, are accepted by non-deterministic automata, then so is their composition $L_0 \cdot L_1$.*

**Proof.** Consider finite automata $\mathcal{A}_i = \langle Q_i, \delta_i, I_i, F_i \rangle$ that accept $L_i$, $i < 2$. The new automaton $\mathcal{A} = \langle Q, \delta, I, F \rangle$ is the *sequential composition* of $\mathcal{A}_0$ with $\mathcal{A}_1$; more precisely,

- ▷ $Q := Q_0 + Q_1$;
- ▷ $m\delta := m\delta_0 + m\delta_1$ if $m \neq e$, and $e\delta := e\delta_0 + e\delta_1 + (F_0 \times I_1)$;
- ▷ $I := I_0$ and $F := F_1$.

The paths from $I = I_0$ to $F = F_1$ are precisely the composite paths from $I_0$ to $F_0$ with those from $I_1$ to $F_1$ linked by one of the new $e$-transitions from $I_0$ to $F_1$. □

**2.5.13 Proposition.** *If the set $L \subseteq M$ is accepted by a finite non-deterministic automaton, then so is its Kleene star $L^*$.*

**Proof.** Observe that

$$L^* = \bigcup \{ L^n : n \in \mathbb{N} \} = L^0 \cup \bigcup \{ L^n : n \in \mathbb{N}_{>0} \} = \{e\} \cup L^+$$

In view of Proposition 2.5.11 it suffices to construct a finite automaton that accepts $L^+$.

Consider a finite automaton $\mathcal{A} = \langle Q, \delta, I, F \rangle$ that accepts $L$. The corresponding *feedback automaton* $\mathcal{A}^+ = \langle Q, \delta^+, I, F \rangle$, which extends $\mathcal{A}$ by linking all final states with all initial states by means of $e$-transitions, *i.e.*,

$$e\delta^+ := e\delta + (F \times I)$$

is finite and accepts $L^+$. □

## 2.6 Rational expressions

For a given ordered monoid $M$ consider closing the set of principal upper segments in $M\boldsymbol{U}$ under the operations of

- ▷ finite unions (hence $MF$ will be subsumed);
- ▷ concatenation (hence one obtains a sub-monoid);
- ▷ the Kleene star (a special infinite union leading out of $MF$).

Just as for regular languages, regular expressions provide a syntactic description for the resulting sets. However, to emphasize the expanded scope, we will now speak of "rational expressions".

**2.6.00 Definition.** The *rational expressions* over a monoid $\langle M.\cdot, e\rangle$ are the $\Sigma$-terms over the set $M$ for the signature consisting

- a nullary function symbol $\emptyset$;
- binary function symbols $+$ and $\cdot$;
- a unary function symbol $(-)^*$.

Their intended *semantics* in the power-set of $M$ is given by

- ▷ the nullary union $\emptyset$;
- ▷ binary union, resp., composition;
- ▷ the Kleene star;
- ▷ the principal upper segments $m{\uparrow}$, $m \in M$. ◁

We have already seen that for a monoid $M$ the set $M$Rat contains $MF$ and is closed under binary unions, binary composition and Kleene star, and hence contains all sets describable by rational expressions.

The obvious conjecture, suggested by the terminology, is that every rational subset of $M$ is the syntax of some rational expression in $M$,

**2.6.01 Theorem.** *Every rational set is the syntax of a rational expression.*

**Proof.** The $L^i_{jk}$-algorithm of Theoretische Informatik 1 just works. □

## 2.7 The monad of rational sets and further closure properties of Rat

Over sets the finite power-set monad is a "sub-monad" of the full power-set monad: the horizontal functor component of the underlying monad-morphism is $id_{set}$, while the 2-cell component is given by the natural inclusion transformation $F \xrightarrow{i} P$.

The same phenomenon occurs for the extensions $\mathbb{F}^{\mathbb{L}}$ of the $\mathbb{F}$ and $\mathbb{P}^{\mathbb{L}}$ of $\mathbb{P}$ to $\boldsymbol{mon}$. Just as the regular languages over $X^*$ for some finite alphabet $X$ sit between $X^*F \subseteq X^*P$, for a general monoid $M$ the "rational sets" we are about to introduce will sit between $MF \subseteq MP$. In fact, one uses the characterization of regular languages by means of regular expressions will allow us to translate the corresponding closure properties to the monoid-setting. This will result in a sub-monad Rat of $\mathbb{P}^{\mathbb{L}}$ that contains $\mathbb{F}^{\mathbb{L}}$ as a further sub-monad.

**2.7.00 Theorem.** *Rational unions of rational sets are rational.*

**Proof.** Use two levels of rational expressions, say red at the level of $M$ and blue at the level of $M$Rat. Eliminating the difference in colors produces regular expressions at the level of $M$. □

**2.7.01 Proposition.** *Rational sets are closed under direct images along monotone monoid homomorphisms.*

**Proof.** This is a direct consequence of the fact that the direct image function $\varphi_\exists$ of a monoid homomorphism $M \xrightarrow{\varphi} N$ as a monoid homomorphism preserves composition, and as a left adjoint preserves arbitrary unions, hence in particular the Kleene star. □

For a lax homomorphism $M \xrightarrow{\varphi} N$ the direct image function $M\boldsymbol{U} \xrightarrow{\varphi_\exists} N\boldsymbol{U}$ is lax as well and need not preserve set-multiplication, not even when combined with downward closure.

**2.7.02 Proposition.** *Rational sets are closed under inverse images along lax homomorphisms with finitely generated domain.*

**Proof.** If $L \subseteq M$ is recognized by a finitely supported lax homomorphism $M \xrightarrow{\psi} \Omega$ into a finite quantale $U$, the composition of $\psi$ with a lax homomorphism $K \xrightarrow{\sigma} M$ is finitely supported iff $K$ is finitely generated. □

**2.7.03 Example.** The set $\mathbb{N}^*$ is not rational. ◁

**2.7.04 Proposition.** *([Pin18, Theorem IV.1.2]) Cartesian products of rational sets are rational.*

**Proof.** Suppose $L_i \in M_i\boldsymbol{U}$ is rational, $i < 2$, then their images under the canonical left inverses of the projections $M_i \xrightarrow{\sigma_i} M_0 \times M_1$ (*cf.*, the proof of Theorem 2.3.03) are rational by Proposition 2.7.01. Hence we can express the cartesian product as

$$L_0 \times L_1 = (L_0 \times \{e_1\}) \cdot (\{e_0\} \times L_1) = L_0(\sigma_0)_\exists \cdot L_1(\sigma_1)_\exists$$

which by Proposition 2.5.12 is rational. □

Unfortunately, for a lax homomorphism $M \xrightarrow{\omega} N$ between ordered monoids, while $\omega_\exists$ is still lax, the right adjoint $\omega^\leftarrow$ in general fails to be lax. Hence one cannot adapt the proof of Propositions 2.3.05 and 2.3.08 for rational languages. Hence it is unlikely that rational sets are closed with respect to residuations and quotients. Already in case of residuating with singletons the interpretation of shifting the initial or final states of the recognizing automaton forward, resp., backwards, no longer works in the lax case, when the shifts $j$ or $k$ fail to be indecomposable.

**Conjecture**: Rational sets might be closed under residuation and quotients under sets that are closed under decomposition, *i.e.*, if $j = a \cdot b \in J$ then $a, b \in J$.

## 2.8 Comparing REC and RAT

After seeing all these differences, the following result that connects the recognizable and the rational sets is all the more surprising:

**2.8.01 Theorem. (Kleene)** *Every finite set* $X$ *satisfies* $X^*\mathrm{RAT} = X^*\mathrm{REC}$.

**2.8.02 Corollary.** *For every finite set* $X$, *the set* $X^*\mathrm{RAT}$ *of recognizable languages is closed under complementation.*

Hence adding a complementation operation $(-)^{\mathrm{c}}$ to the signature of rational expressions does not change the class of rational languages. (Á priory it is not clear, if the rational subsets of monoids that are non-finitely generated freely can change under this extension!) This leads to the question of characterizing star-free languages mentioned in the introduction.

**2.8.03 Theorem.** *(McKneight) The following are equivalent for any* $M \in \boldsymbol{mon}$ *:*

(a) $M$ *is finitely generated.*

(b) $M\mathrm{REC} \subseteq M\mathrm{RAT}$.

(c) $M \in M\mathrm{RAT}$.

**Proof.**

(a) ⇒ (b): By hypothesis, there exists a regular epi $X^* \xrightarrow{\pi} M$ with $X$ finite. If $L \subseteq M$ is recognizable, so is $L\pi^{-1}$ (see Proposition 2.3.00(0)), which by Kleene's Theorem is rational as well. But then so is $L\pi^{-1} \cdot \pi_\exists$. Since $\pi$ is surjective, this set coincides with $L$.

(b) ⇒ (c): Clear, since $M \in M\mathrm{REC}$.

(c) ⇒ (a): This follows by Proposition 2.5.06. □

**2.8.04 Example.** ([Pin18, Example IV.1.3]) Intersections of rational languages need not be rational. Consider $M := \{a\}^* \times \{b,c\}^*$ . The rational subsets

$$R := \langle a,b\rangle^* \langle 1,c\rangle^* = \{\, \langle a^n, b^n c^m\rangle \;:\; m,n \in \mathbb{N} \,\}$$
$$S := \langle 1,b\rangle^* \langle a,c\rangle^* = \{\, \langle a^n, b^m c^n\rangle \;:\; m,n \in \mathbb{N} \,\}$$

have intersection

$$R \cap S = \{\, \langle a^n, b^n c^n\rangle \;:\; n \in \mathbb{N} \,\}$$

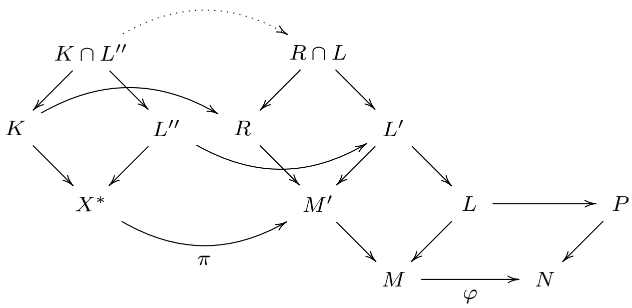
Since its projection $\{\, b^n c^n \;:\; n \in \mathbb{N} \,\}$ onto $\{b,c\}^*$ is well-known to be not rational, by Proposition 2.7.01 $R \cap S$ cannot be rational as well. $\triangleleft$

.

**2.8.05 Proposition.** *Intersections of rational with recognizable subsets are rational.*

**Proof.** If $R \subseteq M$ is rational, there exists a finitely generated sub-monoid $M'$ of $M$ with $R \subseteq M'$ . Concretely, we consider a surjective monoid homomorphism $X^* \xrightarrow{\pi} M'$ with a finite alphabet $X$ . By Proposition 2.7.02 the inverse image $K := R\pi^{\leftarrow}$ is rational and hence by Kleene's Theorem recognizable in $X^*$ .

Let $L \subseteq M$ be recognizable, say, the pullback of $F \subseteq N$ along $M \xrightarrow{\varphi} N$ where $N$ is finite. Then $L' := M' \cap L$ is recognizable in $M'$ , and $L'' := L'\pi^{\leftarrow}$ is recognizable in $X^*$ . Hence $K \cap L''$ is recognizable and therefore rational in $X^*$ .




Using Lemma 2.2.04 and the surjectivity of $\pi$ we obtain

$$(K \cap L'')\pi_\exists = K\pi_\exists \cap L''\pi_\exists = R \cap L'\pi^{-1}\pi_\exists = R \cap L'$$

Hence $R \cap L = R \cap L'$ as the direct image of the rational set $K \cap L''$ is rational. $\square$

This result together with Example 2.8.04 resembles what we learned about context-free languages in TheoInf 1: on finitely generated free monoids, context-free languages in general are not closed under intersection, but intersections with regular (=recognizable) languages are again context-free. Since "context-free unions of context-free languages are context-free", we expect the context-free sets to form a sub-monad of the power-set monad as well and this ought to contain the "contain" the monad of rational sets.

# 3 Green's Relations for categories

Just as the structure of the graph underlying an automaton $\mathcal{A}$ for a language or set $L \subseteq M$ can reveal properties of $L$, the structure of a monoid $N$ recognizing $L$ can provide valuable information.

A very useful notions for graphs, or rather the free categories generated by graphs, is that of "connected component", *i.e.*, an equivalence class of objects for the reachability order. Of course, the connected components are again partially ordered and can be used as a basis for inductive proofs.

Green's relations are intended to serve a similar purpose for the global set of arrows of a monoid, or, in fact, category, *i.e.*, the disjoint union of all hom-sets. Instead of the reachability order one considers three order relations, based on prefixes, postfixes and infixes, respectively. This contrasts with the notion of ***pos***-enriched category, where each hom-set is ordered, while morphisms from different hom-sets are incomparable.

## 3.1 The canonical relations defined by pre-, post- and infixes

As described after Proposition 2.4.01 in case of a monoid, given a category $\mathscr{C}$ there are two binary operations available for binary relations on the class $\mathscr{C}_1$ of all $\mathscr{C}$-morphisms, *i.e.*, the disjoint union of all hom-sets $\langle A, B\rangle\mathscr{C}$, $A, B \in \mathscr{C}_0$: the relation product $\star$, and the set-composition that is based on the composition in $\mathscr{C}$ and hence also will be denoted by " $\cdot$ ".

Concretely this means that $\mathcal{A} \cdot \mathcal{B}$ consists of all composites $\langle a_0 \cdot b_0, a_1 \cdot b_1\rangle$ of matching pairs $\langle a_0, a_1\rangle \in \mathcal{A}$ and $\langle b_0, b_1\rangle \in \mathcal{B}$, where "matching" refers to the fact that the codomain of $\langle a_0, a_1\rangle$ is the domain of $\langle b_0, b_1\rangle$. More abstractly, $\mathcal{A} \cdot \mathcal{B}$ results from forming the pullback with respect to codomain and domain, and then performing the composition.

Note that this set-composition again admits both residuations. In particular, Proposition 2.4.03 concerning congruence relations carries over to this context:

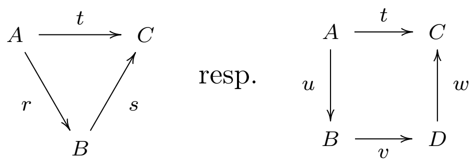
**3.1.00 Proposition.** *For every equivalence relation $\mathcal{Q}$ on the global set $\mathscr{C}_1$ of arrows of a category $\mathscr{C}$, its 2-sided residuation $\tilde{\mathcal{Q}} = \mathscr{C}_1\Delta\backslash\mathcal{Q}/\mathscr{C}_1\Delta$ with respect to the diagonal is the largest congruence contained in $\mathcal{Q}$.* □

Recall that congruence relations are intended for forming factor structures, quotient categories in this case. Observe that $\tilde{\mathcal{Q}}$ necessarily consists of pairs of parallel arrows, which was not required for $\mathcal{Q}$. Hence the resulting quotient category $\mathscr{C}/_{\tilde{\mathcal{Q}}}$ has the same objects as $\mathscr{C}$, while its hom-sets are quotients in ***set*** of the original hom-sets. Note that it is possible for objects to be isomorphic in $\mathscr{C}/_{\tilde{\mathcal{Q}}}$, but not in $\mathscr{C}$. If one were interested in forming isomorphism classes in $\mathscr{C}/_{\tilde{\mathcal{Q}}}$ (usually one is not), doing so in advance in $\mathscr{C}$ would not be sufficient.

In general, for a binary relation on $\mathscr{C}_1$ it does not make sense to ask, whether some functor $\mathscr{C} \xrightarrow{F} \mathscr{D}$ preserves the relation, as we have not specified any relation on $\mathscr{D}$ for comparing

$F$-images of related pairs in $\mathscr{C}$ with. However, the relations we are about to introduce make sense in any category, and are guaranteed to be preserved by any functor.

**3.1.01 Definition.** For a category $\mathscr{C}$ consider. Whenever $A \xrightarrow{t} B$ factors as




(3.1-00)

we call $r$ a *prefix* and $s$ a *postfix* of $t$, and refer to $v$ as an *infix* of $t$.

For historical reasons one writes $t \leq_{\mathcal{R}} r$, $t \leq_{\mathcal{L}} s$ and $t \leq_{\mathcal{J}} w$ for the duals of these pre-order relations. They induce the basic *Green's equivalence relations* $\mathcal{R}$, $\mathcal{L}$ and $\mathcal{J}$, respectively.

Furthermore, in the complete lattice of equivalence relations on $\mathscr{C}_1$ we consider the derived Green's relations $\mathcal{H} := \mathcal{R} \sqcap \mathcal{L} = \mathcal{R} \cap \mathcal{L}$ and $\mathcal{D} := \mathcal{R} \sqcup \mathcal{L}$.

Notice that only $\mathcal{H}$ is guaranteed to consist of pairs of parallel arrows.

**3.1.02 Lemma.** *For every category* $\leq_{\mathcal{J}} \; = \; \leq_{\mathcal{L}} \star \leq_{\mathcal{R}} \; = \; \leq_{\mathcal{R}} \star \leq_{\mathcal{L}}$.

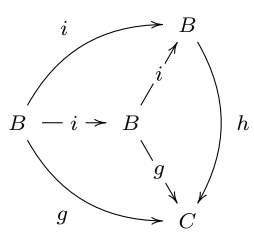
**Proof.** By construction, a prefix of a postfix of $t$, just like a postfix of a prefix of $t$, is an infix of $t$, hence $\leq_{\mathcal{L}} \star \leq_{\mathcal{R}} \; \subseteq \; \leq_{\mathcal{J}}$ and $\leq_{\mathcal{R}} \star \leq_{\mathcal{L}} \; \subseteq \; \leq_{\mathcal{J}}$.

Conversely, $t \leq_{\mathcal{J}} v$ implies the existence of $\mathscr{C}$-morphisms $u, w$ such that the second diagram in (3.1-00) commutes. But then $t \leq_{\mathcal{R}} (u \cdot v) \leq_{\mathcal{L}} v$ and also $t \leq_{\mathcal{L}} (v \cdot w) \leq_{\mathcal{R}} v$. □

Before analyzing the relationships between $\mathcal{J}$ and the compositions $\mathcal{R} \cdot \mathcal{L}$ and $\mathcal{L} \cdot \mathcal{R}$, we need some intermediate results concerning idempotents.

**3.1.03 Lemma.** *Any idempotent arrow* $B \xrightarrow{i} B$ *of a category* $\mathscr{C}$ *satisfies: If* $i$ *is a first factor or prefix of* $B \xrightarrow{g} C$ *(a last factor or postfix of* $A \xrightarrow{f} B$*), then* $g = i \cdot g$ $(f = f \cdot i)$.

**Proof.** $g = i \cdot g$ is immediate from




The argument for $f = f \cdot i$ is dual. □

**3.1.04 Lemma.** *If $M$ is a finite monoid with $n$ elements, for each $a \in M$ the power $a^{n!}$ is idempotent.*

**Proof.** For $n = 1$ the claim is trivial, hence suppose $n > 1$. As there are at most $n$ distinct powers of $a$, there exist minimal positive integers $k, \ell < n$, with $a^k = a^{k+\ell}$. If $\ell = k$, then $a^k$ is idempotent. Otherwise observe that $a^k = a^{k+k\ell} = a^{k(\ell+1)}$. But his implies that

$$a^{k\ell} = a^k \cdot a^{k(\ell-1)} = a^{k(\ell+1)} \cdot a^{k(\ell-1)} = a^{2k\ell}$$

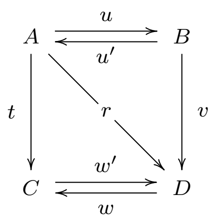
is idempotent. In either case, $q = k$, resp. $q = k\ell$ is a factor of $n!$, and consequently

$$a^{n!} = (a^q)^{n!/q}$$

is idempotent, since $a^q$ has this property. □

**3.1.05 Proposition.** *$\mathcal{L} \star \mathcal{R}$ and $\mathcal{R} \star \mathcal{L}$ always agree with $\mathcal{D} = \mathcal{L} \sqcup \mathcal{R}$, and for categories where all hom-sets of the form $\langle X, X \rangle \mathscr{C}$ are finite, this also agrees with $\mathcal{J}$.*

**Proof.** We first show $\mathcal{R} \star \mathcal{L} \subseteq \mathcal{L} \star \mathcal{R}$. Suppose $(A \xrightarrow{t} C)\mathcal{R}(A \xrightarrow{r} D)\mathcal{L}(B \xrightarrow{v} D)$. By definition there exist $\mathscr{C}$-morphisms $A \underset{u'}{\overset{u}{\rightleftarrows}} B$ and $C \underset{w'}{\overset{w}{\rightleftarrows}} D$ with




(3.1-01)

which for $x := u' \cdot r \cdot w$ implies

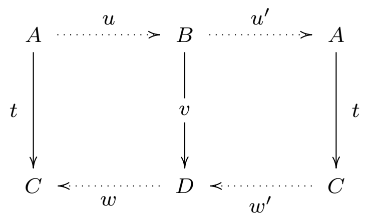
$$t = (r \cdot w)\, \mathcal{L}\, (v \cdot w) = x = (u' \cdot t)\, \mathcal{R}\, (u' \cdot r) = v$$

Of course, the other inclusion follows by a similar argument.

Clearly, due to the reflexivity of $\mathcal{L}$ and $\mathcal{R}$ we have $\mathcal{L}, \mathcal{R} \subseteq \mathcal{R} \star \mathcal{L} = \mathcal{L} \star \mathcal{R}$. Consider any equivalence relation $\mathcal{Q}$ on $\mathscr{C}_1$ with $\mathcal{L}, \mathcal{R} \subseteq \mathcal{Q}$. Transitivity of $\mathcal{Q}$ implies $\mathcal{R} \star \mathcal{L} = \mathcal{L} \star \mathcal{R} \subseteq \mathcal{Q} \star \mathcal{Q} \subseteq \mathcal{Q}$, which establishes $\mathcal{R} \star \mathcal{L} = \mathcal{L} \star \mathcal{R}$ as the join $\mathcal{D}$ of $\mathcal{L}$ and $\mathcal{R}$. In particular, this shows $\mathcal{D} \subseteq \mathcal{J}$.

Now consider the case where all endo-hom-sets of $\mathscr{C}$ are finite. By Lemma 3.1.04 every morphism in such a hom-set has an idempotent power. We need to show $\mathcal{J} \subseteq \mathcal{D}$.

Consider $(A \xrightarrow{t} C)\, \mathcal{J}\, (B \xrightarrow{v} D)$, which for suitable $u$, $u'$, $w$ and $w'$ means




(3.1-02)

For some $N > 0$ the endo-morphism $(u \cdot u')^N$ is idempotent. Then by Lemma 3.1.03 we have

$$t = (u \cdot u')^N \cdot t = (u \cdot u')^{N-1} \cdot u \cdot u' \cdot t$$

and hence $t \leq_{\mathcal{L}} u' \cdot t$. But $u' \cdot t \leq_{\mathcal{L}} t$ is true by default, and so we obtain $t\,\mathcal{L}\,(u' \cdot t)$.

In similar fashion one can use an idempotent power of $w \cdot w'$ to obtain $t\,\mathcal{R}\,(t \cdot w')$. Since $\mathcal{R}$ is stable under pre-composition, this furthermore implies $(u' \cdot t)\,\mathcal{R}\,(u' \cdot t \cdot w') = v$, which combined with $t\,\mathcal{L}\,(u' \cdot t)$ results in $t\,(\mathcal{L} \star \mathcal{R})\,v$, and hence $\mathcal{J} \subseteq \mathcal{L} \star \mathcal{R} = \mathcal{D}$. □

The proof above reveals the structure of the $\mathcal{D}$-classes:

**3.1.06 Proposition. (Green's Lemma)**

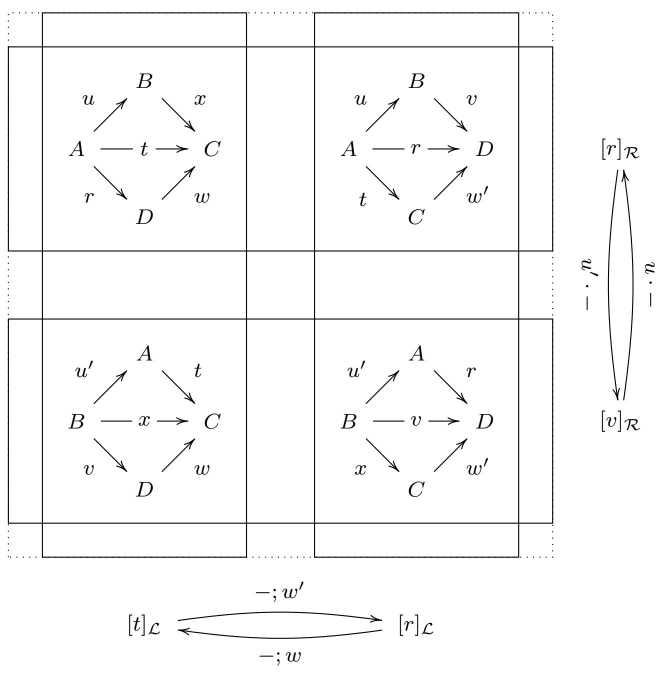
(0) *Every $\mathcal{D}$-class is by definition union of $\mathcal{R}$-classes and union of $\mathcal{L}$-classes, and the intersection of any such $\mathcal{R}$-class with any such $\mathcal{L}$-class is non-empty.*

(1) *Any two $\mathcal{R}$-classes inside a $\mathcal{D}$-class are isomorphic as sets, as are any two $\mathcal{L}$-classes and any two $\mathcal{H}$-classes.*

**Proof.**

(0) $t\,\mathcal{D}\,v$ implies $t\,(\mathcal{R} \star \mathcal{L})\,v$, hence there exists $r$ with $t\,\mathcal{R}\,r\,\mathcal{L}\,v$. Therefore $r \in [t]_{\mathcal{R}} \cap [v]_{\mathcal{L}}$.

(1) We claim that the $\mathscr{C}$-morphisms $A \underset{u'}{\overset{u}{\rightleftarrows}} B$ and $C \underset{w}{\overset{w'}{\rightleftarrows}} D$ of Diagram 3.1-01 induce bijections between $[r]_{\mathcal{L}}$ and $[v]_{\mathcal{L}}$ by pre-composition, as well as between $[t]_{\mathcal{R}}$ and $[r]_{\mathcal{R}}$ by post-composition, as illustrated by the following "egg-crate" picture of a typical $\mathcal{D}$-class:




(3.1-03)

We know that $t \cdot w' \cdot w = t$ and $r \cdot w \cdot w' = r$. Recall that $x = u' \cdot r \cdot w$. Then

$$x \cdot w' \cdot w = u' \cdot r \cdot w \cdot w' \cdot w = u' \cdot r \cdot w = x$$

and

$$v \cdot w \cdot w' = u' \cdot r \cdot w \cdot w' = u' \cdot t \cdot w' = v$$

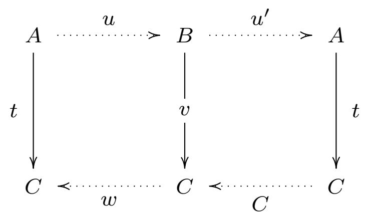
The argument for $u$ and $u'$ is analogous. Of course, the indicated bijections restrict to the $\mathcal{H}$-classes. □

**3.1.07 Proposition.** *In any category, where all hom-sets of the form* $\langle X, X\rangle\mathscr{C}$ *are finite, we have*

$$\geq_{\mathcal{J}} \cap \leq_{\mathcal{L}} \ \subseteq \mathcal{L} \qquad \text{and} \qquad \geq_{\mathcal{J}} \cap \leq_{\mathcal{R}} \ \subseteq \mathcal{R}$$

**Proof.** We show the first statement; the second one follows dually.

Suppose $t \leq_{\mathcal{J}} v$ and $v \leq_{\mathcal{L}} t$. There exist $\mathscr{C}$-morphisms $u$, $u'$ and $w$ such that




This is Diagram 3.1-02 with $C \xrightarrow{w'} D$ replaced by the identity on $C$. And since $v = u' \cdot t$, just as in the proof of Proposition 3.1.05 we obtain

$$t = (u \cdot u')^N \cdot t = (u \cdot u')^{N-1} \cdot u \cdot u' \cdot t = (u \cdot u')^{N-1} \cdot u \cdot v$$

which implies $t \leq_{\mathcal{L}} v$, as desired. □

Often this result is stated in apparently slightly weaker form in the Literature:

$$\mathcal{J} \cap \leq_{\mathcal{L}} \ \subseteq \mathcal{L} \qquad \text{and} \qquad \mathcal{J} \cap \leq_{\mathcal{R}} \ \subseteq \mathcal{R}$$

But since $\mathcal{L} \subseteq \mathcal{J}$, above we also have $v \leq_{\mathcal{J}} t$ and therefore $v\,\mathcal{J}\,t$.

Mikołaj Bojańczyk goes so far as claiming that this simple result essentially contains the theory of finite monoids, and by extension, of categories with finite endo-hom-sets.

We now turn to idempotent endo-morphisms.

**3.1.08 Lemma. (Location Lemma** (Clifford and Miller)**)** *If in any* $\mathcal{D}$*-class we have a configuration as depicted in Diagrams (3.1-01) and (3.1-03), then*

$$B = C \ \text{ and } \ r = t \cdot v \ \text{ iff } \ x = u' \cdot r \cdot w \ \text{ is idempotent.}$$

*and dually*

$$A = D \text{ and } x = v \cdot t \text{ iff } r = u \cdot x \cdot w' \text{ is idempotent.}$$

**Proof.** We show the first equivalence, the second follows dually.

($\Rightarrow$) $t \mathcal{R} r = t \cdot v$ allows us to chose $w' = v$. By Proposition 3.1.05 (Green's Lemma) $C \underset{v}{\overset{w}{\rightleftarrows}} D$ induce inverse bijections between $[t]_{\mathcal{L}}$ and $[t \cdot v]_{\mathcal{L}}$. In particular, $x := v \cdot w$ satisfies $x \cdot v = v \cdot w \cdot v = v$ and hence $x = (v \cdot w)^2$ is idempotent.

($\Leftarrow$) Let $x = v \cdot w \in [t]_{\mathcal{L}} \cap [v]_{\mathcal{R}}$ be idempotent. Since $v = x \cdot w'$, Lemma 3.1.03 implies $v = x \cdot v$. Thus $- \cdot v$ also is a bijection from $[x]_{\mathcal{L}}$ to $[v]_{\mathcal{L}}$. It maps $t \in [x]_{\mathcal{L}}$ to $t \cdot v = u \cdot x \cdot v = u \cdot v = r$. □

**3.1.09 Theorem. (Green's Theorem)** *If $B \xrightarrow{x} B$ is idempotent, then $[x]_{\mathcal{H}}$ is a maximal "local group" in $\langle B, B \rangle \mathscr{C}$ in the sense discussed before Example 2.3.12. Hence every $\mathcal{H}$-class contains at most one idempotent. Moreover, the $\mathcal{H}$-class $H$ of an endo-morphism $B \xrightarrow{b} B$ is a "local group" iff $H \cap H \cdot H \neq \emptyset$.*

**Proof.** For $t, v \in [x]_{\mathcal{H}}$, since $\mathcal{H} \subseteq \mathcal{D}$ we have $[x]_{\mathcal{H}} = [t]_{\mathcal{L}} \cap [v]_{\mathcal{R}} = [a]_{\mathcal{R}} \cap [b]_{\mathcal{L}}$, hence $t \cdot v \in [x]_{\mathcal{H}}$. Lemma 3.1.03 guarantees that $x$ is neutral in $[x]_{\mathcal{H}}$.

As in the proof above, $t \cdot -$ is an $\mathcal{H}$-class-preserving bijection from $[x]_{\mathcal{R}}$ to $[t]_{\mathcal{R}}$, and hence a bijection from $[x]_{\mathcal{H}}$ to $[t]_{\mathcal{H}} = [x]_{\mathcal{H}}$. In particular, $x$ has to have a pre-image under $t \cdot -$ in $[x]_{\mathcal{H}}$, which is the desired inverse of $t$.

As neutral elements of groups are unique, $\mathcal{H}$-classes can contain at most one idempotent.

Suppose $G \subseteq M$ is a local group with $x \in G$ neutral, hence idempotent in $M$. For each $a \in G$ there exists a "local inverse" $a'$ with $a \cdot a' = x = a' \cdot a$, hence $x \leq_{\mathcal{R}} a$ and $x \leq_{\mathcal{L}} a$. But since $x \cdot a = a = a \cdot x$ we also have $a \leq_{\mathcal{R}} x$ and $a \leq_{\mathcal{L}} x$. Therefore $x\mathcal{R}a$ and $x\mathcal{L}a$, hence $x\mathcal{H}a$. This shows $G \subseteq [x]_{\mathcal{H}}$. and thus establishes the maximality of $[x]_{\mathcal{H}}$.

If $H$ is a "local group", its neutral element $B \xrightarrow{x} B$ is idempotent and belongs to the intersection. The converse is a consequence of the Location Lemma 3.1.08. □

**3.1.10 Proposition.** *If a $\mathcal{D}$-class of a category contains an idempotent $B \xrightarrow{x} B$, then every $\mathcal{R}$-class and every $\mathcal{L}$-class contained in this $\mathcal{D}$-class also contains an idempotent.*

**Proof.** If $[x]_{\mathcal{H}}$ coincides with $[x]_{\mathcal{D}}$, it also coincides with $[x]_{\mathcal{L}}$ and with $[x]_{\mathcal{R}}$, and we are done.

Otherwise, for $(A \xrightarrow{r} D) \in [x]_{\mathcal{D}} - [x]_{\mathcal{H}}$ choose $(A \xrightarrow{t} B) \in [x]_{\mathcal{L}} \cap [r]_{\mathcal{R}}$. Diagram 3.1-03 with $C = B$ shows

$$t \cdot u' \cdot t \cdot u' = t \cdot i \cdot u' = u \cdot t \cdot t \cdot u' = u \cdot i \cdot u' = t \cdot u'$$

hence $t \cdot u'$ is idempotent on $A$.

Moreover, $t \cdot u' = r \cdot w \cdot u'$ implies $t \cdot u' \leq_{\mathcal{R}} r$ . On the other hand,

$$r = t \cdot w' = u \cdot x \cdot w' = u \cdot x \cdot x \cdot w' = t \cdot v = t \cdot u' \cdot r$$

and hence $r \leq_{\mathcal{R}} t \cdot u'$ . Therefore, $t \cdot u' \in [r]_{\mathcal{R}}$ .

Similarly, $w \cdot v$ turns out to be idempotent on $D$ and to belongs to $[r]_{\mathcal{L}}$ . □

**3.1.11 Proposition.** *Any two maximal subgroups contained in the same $\mathcal{D}$-class of $\mathscr{C}$ are isomorphic.*

**Proof.** Consider different idempotents $A \xrightarrow{r} A$ and $B \xrightarrow{x} B$ satisfying $r\,\mathcal{D}\,x$ . In view of Diagram 3.1-03 choose $(A \xrightarrow{t} B) \in [x]_{\mathcal{L}} \cap [r]_{\mathcal{R}}$ and $(B \xrightarrow{v} A) \in [x]_{\mathcal{R}} \cap [r]_{\mathcal{L}}$. Set $u = t = w$ and $u' = v = w'$ . The Location Lemma implies $t \cdot v = r$ and $v \cdot t = x$ . Moreover, $u \cdot - \cdot w' = t \cdot - \cdot v$ is a bijection from $[x]_{\mathcal{H}}$ to $[r]_{\mathcal{H}}$ that maps $x$ to $r$ , and whose inverse $u' \cdot - \cdot w = v \cdot - \cdot t$ maps $r$ to $x$ .

For $a, b \in [x]_{\mathcal{H}}$ this yields

$$(t \cdot a \cdot v) \cdot (t \cdot b \cdot v) = t \cdot a \cdot (v \cdot t \cdot b) \cdot v = t \cdot a \cdot (x \cdot b) \cdot v = t \cdot a \cdot b \cdot v$$

which establishes the desired homomorphism. □

### 3.1.1 Open Problem 0

Recall from the start of this chapter that the largest congruence relation contained in $\mathcal{D}$ is $\tilde{\mathcal{D}} = \mathscr{C}_1\Delta \backslash \mathcal{D} / \mathscr{C}_1\Delta$ .

**3.1.12 Conjecture.** In any category, where all hom-sets of the form $\langle X, X \rangle \mathscr{C}$ are finite, we have

$$\mathscr{C}_1\Delta \backslash \mathcal{J} = \mathcal{R} \quad , \quad \mathcal{J} / \mathscr{C}_1\Delta = \mathcal{L} \quad \text{and} \quad \tilde{\mathcal{J}} = \mathcal{R} / \mathscr{C}_1\Delta = \mathscr{C}_1\Delta \backslash \mathcal{L} = \mathcal{H}$$

By construction we have $\mathscr{C}_1\Delta \cdot \mathcal{R} \subseteq \mathcal{R}$ and $\mathcal{L} \cdot \mathscr{C}_1\Delta \subseteq \mathcal{L}$ , which implies $\mathcal{R} \subseteq \mathscr{C}_1\Delta \backslash \mathcal{J}$ and $\mathcal{L} \subseteq \mathcal{J} / \mathscr{C}_1\Delta$ . The converse inclusion is harder.

Consider $A \xrightarrow{t} C \xleftarrow{v} B$ with $\langle t, v \rangle \in \mathcal{J} / \mathscr{C}_1\Delta \subseteq \mathcal{J}$ . According to Proposition 3.1.07 we need to show $v \leq_{\mathcal{L}} t$ in order to conclude $v\,\mathcal{L}\,t$ . For this purpose it may be useful to consider the morphisms $u \cdot u' \in \langle A, A \rangle \mathscr{C}$ , $u' \cdot u \in \langle B, B \rangle \mathscr{C}$ as well as $\{w, w', w \cdot w', w' \cdot w\} \subseteq \langle C, C \rangle \mathscr{C}$ that all have idempotent powers. ???

## 3.2 A proof of Schützenberger's result via Green's relations

The following is based on Thomas Colcombet's notes [DA11].

Recall that the rational languages on free finitely generated monoids can be described by rational expressions in nullary and binary union, concatenation and Kleene star, in addition to the letters of the alphabet. Since by Kleene's Theorem on such monoids they coincide with the recognizable languages, they are closed under complement as well. Hence one can extend the rational expressions for those languages by adding a unary operator $(-)^{\mathrm{c}}$ for complementation.

With these extended rational expressions, one can study classes of languages that satisfy certain constraints.

**3.2.01 Definition.** A language $L \subseteq \Sigma^*$ for finite alphabet $\Sigma$ is called *star-free*, if it can be described by a rational expression without the Kleene star. ◁

**3.2.02 Example.** Let $\Sigma$ be a finite alphabet.

▷ The language $\Sigma^* = \emptyset^{\mathrm{c}}$ is star-free.

▷ For any $B \subseteq \Sigma$ the language $B^* = (\Sigma^* B^{\mathrm{c}} \Sigma^*)^{\mathrm{c}}$ is star-free.

▷ If $a \in \Sigma$ then the language described by $(aa)^*$ is not star-free. ◁

Schützenberger in 1965 managed to characterize the star-free languages in terms of their syntactic monoids.

**3.2.03 Definition.** A monoid $M$ is called *aperiodic*, if for every idempotent power of the form $m^N$ with $N > 0$ we have $m^N = m^N \cdot m$.

A category $\mathscr{C}$ is called *aperiodic*, if every endo-hom-set has this property. ◁

**3.2.04 Definition.** An $\mathcal{L}$ / $\mathcal{R}$ / $\mathcal{J}$ / $\mathcal{D}$ / $\mathcal{H}$-class is called *regular*, if it contains an idempotent. The category $\mathscr{C}$ is called $\mathcal{L}$ / $\mathcal{R}$ / $\mathcal{J}$ / $\mathcal{D}$ / $\mathcal{H}$-*trivial*, if every regular $\mathcal{L}$ / $\mathcal{R}$ / $\mathcal{J}$ / $\mathcal{D}$ / $\mathcal{H}$-class is a singleton. ◁

**3.2.05 Proposition.** *If $M$ / every endo-hom-set of $\mathscr{C}$ is finite and all regular $\mathcal{H}$-classes are trivial, then all $\mathcal{H}$ classes are trivial.*

**Proof.** Consider $B \xrightarrow{y} B$ with idempotent power $y^N$, $N > 0$. We claim $y^{N+1} = y^N$. From $y^N = y^N y y^{N-1} = y^{N-1} y y^N$ we infer $y^N \leq_{\mathcal{L}} y^{N+1}$ as well as $y^N \leq_{\mathcal{R}} y^{N+1}$. The converses hold trivially, therefore $y^N \, \mathcal{L} \, y^{N+1}$ and $y^N \, \mathcal{R} \, y^{N+1}$, which implies $y^N \, \mathcal{H} \, y^{N+1}$. The triviality of all regular $\mathcal{H}$-classes then yields $y^{N+1} = y^N$

Now consider $A \underset{b}{\overset{a}{\rightrightarrows}} B$ in some $\mathcal{H}$-class $H$. As there exist endomorphisms $A \xrightarrow{x} A$ and $B \xrightarrow{y} B$ with $a = x \cdot b$ and $b = a \cdot y$, we get $a = x^k \cdot a \cdot y^k$ for each $k > 0$. If $y^N$ is idempotent, the first part shows $a = x^N a y^N = x^N a y^{N+1} = a \cdot y = b$. □

**3.2.06 Proposition.** *A finite monoid / category with finite endo-hom-sets is aperiodic iff all $\mathcal{H}$-classes are singletons.*

**Proof.** Aperiodicity is equivalent to the fact that no local group can have a proper cyclic subgroup. But since every element of a group generates a cyclic subgroup, this in turn is equivalent to all local groups being singletons. Since the regular $\mathcal{H}$-classes happen to be the maximal local groups, their triviality is equivalent to the triviality of all local groups, which by Proposition 3.2.05 is equivalent to the triviality of all $\mathcal{H}$-classes. □

**3.2.07 Theorem. (Schützenberger, difficult direction)** *For every finite alphabet $\Sigma$, a language $L \subseteq \Sigma^*$ that has an aperiodic finite syntactic monoid is star-free.*

**Proof.** Consider the syntactic quotient $\Sigma^* \xrightarrow{L\eta} \boldsymbol{L\text{-}syn} =: \langle N, \cdot, 1\rangle$ that recognizes $L \subseteq \Sigma^*$. It suffices to show that each $a \in N$ has a star-free pre-image, since then $L$ is the finite union of star-free languages.

We will use the following abbreviations:

$$L_A := A(L\eta)^{\leftarrow} \quad \text{and} \quad C_a := L_{\{a\}} \cap \Sigma \quad \text{for} \quad A \subseteq N \quad \text{resp.} \quad a \in N$$

The strategy will be to use induction along $\leq_{\mathcal{J}}$ in $N$.

**Base case**: Since every $a \in N$ trivially satisfies $s \leq_{\mathcal{J}} 1$, the $\mathcal{J}$-class $[1]_{\mathcal{J}}$ is maximal. By Proposition 3.1.07 $a\,\mathcal{J}\,1$ and $a \leq_{\mathcal{R}} 1$ implies $a\,\mathcal{R}\,1$, similarly for $\mathcal{L}$, and consequently $a\,\mathcal{H}\,1$. Therefore $[1]_{\mathcal{J}}$ coincides with $[1]_{\mathcal{H}}$, which by hypothesis and Proposition 3.2.06 is a singleton. Now $C_1$ is finite, and according to Example 3.2.02 $L_1 = 1(L\eta)^{-1} = (C_1)^*$ is star-free.

**Induction hypothesis**: For $a \in N$ assume that $L_{\{b\}}$ is star-free for all $b \in N$ with $a <_{\mathcal{J}} b$.

**Induction step**: The proof that $L_a$ is star-free proceeds in three steps: by showing that first $[a]_{\mathcal{J}}$, then $[a]_{\mathcal{R}}$ and dually $[a]_{\mathcal{L}}$, and finally $[a]_{\mathcal{H}} = [a]_{\mathcal{R}} \cap [a]_{\mathcal{L}} = \{a\}$ has a star-free pre-image under $L\eta$.

**Claim 0**: $L_{[a]_{\mathcal{J}}}$ is star-free.

The trick is to rewrite this set as

$$L_{[a]_{\mathcal{J}}} = \bigcap \{\, L_{N - b\uparrow_{\mathcal{J}}} \;:\; a <_{\mathcal{J}} b \,\} - L_{N - a\uparrow_{\mathcal{J}}}$$

That $L_{N-a\uparrow_{\mathcal{J}}}$ is star-free follows provided we can establish

$$L_{N-a\uparrow_{\mathcal{J}}} = \Sigma^* \cdot K_a \cdot \Sigma^*$$

where

$$K_a := \bigcup \{\, C_b \;:\; a \not\leq_{\mathcal{J}} b \,\} \cup \bigcup \{\, C_b L_{\{c\}} C_d \;:\; a \not\leq_{\mathcal{J}} b \cdot c \cdot d \wedge a <_{\mathcal{J}} c \,\}$$

which by construction and by induction hypothesis is star-free.

Since the $L\eta$-image of $K_a$ contains no infixes of $a$, we get $K_a \subseteq L_{N-a\uparrow_{\mathcal{J}}}$. Furthermore, $L_{N-a\uparrow_{\mathcal{J}}}$ clearly is an ideal, *i.e.*, $\Sigma^* \cdot L_{N-a\uparrow_{\mathcal{J}}} \cdot \Sigma^* \subseteq L_{N-a\uparrow_{\mathcal{J}}}$, which shows $\Sigma^* \cdot K_a \cdot \Sigma^* \subseteq L_{N-a\uparrow_{\mathcal{J}}}$.

For the other inclusion first consider a minimal $u \in L_{N-a\uparrow_{\mathcal{J}}}$, *i.e.*, no proper infix belongs to $L_{N-a\uparrow_{\mathcal{J}}}$. Since $L\eta$ maps $\varepsilon$ to $1 \in N$ and $a \leq_{\mathcal{J}} 1$, we have $u \neq \varepsilon$.

In case of $u \in \Sigma$ we immediately have $u \in \bigcup\{ C_b : a \not\leq_{\mathcal{J}} b \} \subseteq K_a$.

Now suppose $|u| > 1$. Write $u = xvy$ with $x, y \in \Sigma$. Setting $b := xL\eta$, $c := vL\eta$ and $d := yL\eta$ we claim that $a <_{\mathcal{J}} c$. The minimality of $u$ implies $a \leq_{\mathcal{J}} b \cdot c \leq_{\mathcal{J}} c$ and hence $a \leq_{\mathcal{J}} c$. The assumption $a \not<_{\mathcal{J}} c$ yields $a \mathcal{J} c$, in particular $c \leq_{\mathcal{J}} a \leq_{\mathcal{J}} b \cdot c \leq_{\mathcal{J}} c$, which then implies $(b \cdot c) \mathcal{J} c$. Similarly, from $c \leq_{\mathcal{J}} a \leq_{\mathcal{J}} c \cdot d \leq_{\mathcal{J}} c$ we infer $(c \cdot d) \mathcal{J} c$. According to Proposition 3.1.07 $(b \cdot c) \mathcal{J} c$ and $(b \cdot c) \leq_{\updownarrow} c$ implies $(b \cdot c) \mathcal{L} c$. But since $\mathcal{L}$ is closed under post-composition and contained in $\mathcal{J}$ this yields $(b \cdot c \cdot d) \mathcal{J} (c \cdot d) \mathcal{J} c \mathcal{J} a$, hence $(b \cdot c \cdot d) \mathcal{J} a$, a contradiction. Consequently, $a <_j c$, as desired and $u \in K_a$.

If $u \in L_{N-a\uparrow_{\mathcal{J}}}$ is not minimal, some minimal infix $u'$ belongs to $K_a$, which implies $u \in \Sigma^* \cdot K_a \cdot \Sigma^*$. This concludes the proof of Claim 0.

**Claim 1**: The subsets $L_{[a]_{\mathcal{R}}}$ and $L_{[a]_{\mathcal{L}}}$ of $L_{[a]_{\mathcal{J}}}$ are star-free.

We already know that $[a]_{\mathcal{R}}$ as a subset of $[a]_{\mathcal{J}}$ differs from $[1]_{\mathcal{R}}$. Now we attempt another rewriting trick:

$$L_{[a]_{\mathcal{R}}} = L_{[a]_{\mathcal{J}}} \cap \bigcup \{ L_{\{b\}} \cdot C_c \cdot \Sigma^* : b \cdot c \leq_{\mathcal{R}} a \wedge a <_{\mathcal{J}} b \}$$

Any word $u$ in the right hand side satisfies $(uL\eta) \leq_{\mathcal{R}} b \cdot c \leq_{\mathcal{R}} a$. Together with $(uL\eta) \mathcal{J} a$ by Proposition 3.1.07 we get $(uL\eta) \mathcal{R} a$, hence $u \in L_{[a]_{\mathcal{R}}}$.

Conversely, start with $u \in L_{[a]_{\mathcal{R}}} \subseteq L_{[a]_{\mathcal{J}}}$. Let $v$ be the minimal prefix of $u$ satisfying $(vL\eta) \leq_{\mathcal{R}} a$. This cannot be empty, since $1 \not\leq_{\mathcal{R}} a$. Write $v = wx$ with $x \in \Sigma$. Setting $b := wL\eta$ and $c := xL\eta$ we see $u \in L_{\{b\}} \cdot C_c \cdot \Sigma^*$.

Clearly $(uL\eta) \mathcal{R} a$ and $(uL\eta) \leq_{\mathcal{R}} b \cdot c \leq_{\mathcal{R}} b$ implies $a \leq_{\mathcal{R}} b$ and in particular $a \leq_{\mathcal{J}} b$. But by minimality of $v$ we also have $b \not\leq_{\mathcal{R}} a$ and thus $a <_{\mathcal{R}} b$. Using Proposition 3.1.07 the assumption of $a \mathcal{J} b$ would imply $a \mathcal{R} b$, a contradiction. Since we already have $a \leq_{\mathcal{J}} b$, this implies $a <_{\mathcal{J}} b$, as desired.

The proof for $L_{[a]_{\mathcal{L}}}$ is dual.

**Claim 2**: $L_{\{a\}}$ is star-free.

This follows immediately, since by hypothesis $\{a\} = [a]_{\mathcal{H}} = [a]_{\mathcal{R}} \cap [a]_{\mathcal{L}}$. □

The other direction of this theorem has been established in other lectures:

**3.2.08 Theorem. (Schützenberger, simple direction)** *For every finite alphabet (graph) $\Sigma$, a language $L \subseteq \Sigma^*$ that is star-free has an aperiodic finite syntactic monoid (category).* □

### 3.2.1 Open Problem 1

If, as conjectured above, for categories $\mathscr{D}$ with finite endo-hom-sets $\mathcal{H}$ is the largest congruence relation contained in $\mathcal{J}$, then $\mathscr{D}$ is acyclic, precisely when $\mathscr{D}$ coincides with the quotient $\mathscr{D}/_{\mathcal{H}}$.

Consider the syntactic quotient $\mathscr{G}^* \xrightarrow{L\eta} \mathscr{D}$ of some language (set of arrows) $L$ in the free category over a finite graph $\mathscr{G}$.

# 4 Linking the Theorems of Birkhoff and Eilenberg by means of duality

In 1935, Garrett Birkhoff published his now famous HSP theorem of universal algebra: varieties of algebras, *i.e.*, classes of algebras for a given signature $\Sigma$ of function symbols, closed under homomorphic images, sub-objects and products, are precisely those classes that satisfying a class of equations. The latter are just pairs of terms over some alphabet, *i.e.*, in a free algebra.

As seen above, the class of regular languages is recognized by the class of all finite monoids. In 1976 Eilenberg and Schützenberger identified sensible constraints on both sides: "pseudovarieties" of finite monoids, which are closed under homomorphic images, sub-objects and finite products, correspond to so-called "varieties" of regular languages. These are closed under the Boolean operations (finite union and intersection as well as complement), under residuations (or quotients) with singletons, and under inverse images along monoid homomorphisms between free monoids.

Reiterman in 1982 managed to characterize general pseudovarieties by considering generalized equations, which are pairs of terms in free "profinite" structures.

These results still leave a gap: what is the connection between the equations or profinite equations of the Birkhoff/Reiterman theorems, and the "varieties" of languages of the Eilenberg-Schützenberger theorem? This gap only recently has been bridged by (Adamek, Milius Urbat, Chen, Myers ...) by introducing duality into the picture. This leads to the slogan:

$$\text{Eilenberg} = \text{Birkhoff} + \text{Duality}$$

So far, our presentation is based on Julian Salamanca's [arXiv:1702.02822v1].

## 4.0 Dualities

**4.0.01 Definition.** A *duality* between two categories $\mathscr{C}$ and $\mathscr{D}$ is an equivalence between $\mathscr{C}^{\mathrm{op}}$ and $\mathscr{D}$, or, equivalently, between $\mathscr{C}$ and $\mathscr{D}^{\mathrm{op}}$.

Some authors try to eliminate the explicit reversal of arrows by introducing "contra-variant functors" that reverse rather than preserve arrows. While this reduces the proliferation of $^{\mathrm{op}}$-operators, this comes at the price of having to remember which functors are contra-variant and reverse arrows, and which functors are "covariant" and preserve arrows. In these notes we would like to avoid using contra-variant arrows. To still obtain a somewhat more symmetric notion of duality, we use the following set-up: functors

$$\mathscr{C} \xrightarrow{F} \mathscr{D}^{\mathrm{op}} \quad \text{and} \quad \mathscr{D} \xrightarrow{G} \mathscr{C}^{\mathrm{op}}$$

and natural isomorphisms

$$\mathscr{C} \overset{\gamma}{\Longrightarrow} FG^{\mathrm{op}} \quad \text{and} \quad \mathscr{D} \overset{\delta}{\Longrightarrow} GF^{\mathrm{op}}$$

The dualities we wish to exploit here have a rather special form. They arise from a duality between essentially small categories by applying the ***ind*-**, respectively ***ind*-**constructions and thereby recovering familiar lfp-categories, or their duals, see Definition 5.12.03 and Theorem 5.12.04.

**4.0.02 Example.** The duality between $\boldsymbol{set}_f$ and $\boldsymbol{ba}_f$ that maps each finite set to its power set and each finite Boolean algebra to its set of atoms implies two dualities of interest in automata theory:

- ▷ between $\boldsymbol{set} = \boldsymbol{set}_f\text{-}\boldsymbol{ind}$ and $\boldsymbol{ba}_f\text{-}\boldsymbol{pro}$, which turns out to be the category $\boldsymbol{caba}$ of complete atomic Boolean algebras;
- ▷ between $\boldsymbol{ba} = \boldsymbol{ba}_f\text{-}\boldsymbol{ind}$ and $\boldsymbol{set}_f\text{-}\boldsymbol{pro}$, which is known as $\boldsymbol{stone}$, the category of *Stone spaces*; these can be identified, *e.g.*, as compact totally disconnected Hausdorff spaces.

**4.0.03 Example.** The duality between $\boldsymbol{pos}_f$ and $\boldsymbol{dl}_f$ that maps each finite poset to its lattice of down-sets and each finite distributive lattice to its poset of join-irreducibles implies two further dualities of interest in automata theory:

- ▷ between $\boldsymbol{pos} = \boldsymbol{pos}_f\text{-}\boldsymbol{ind}$ and $\boldsymbol{dl}_f\text{-}\boldsymbol{pro}$, which turns out to be the category $\boldsymbol{acdl}$ of algebraic complete distributive lattices;
- ▷ between $\boldsymbol{dl} = \boldsymbol{dl}_f\text{-}\boldsymbol{ind}$ and $\boldsymbol{pos}_f\text{-}\boldsymbol{pro}$, which is known as $\boldsymbol{priest}$, the category of *Priestley spaces*; these can be identified, *e.g.*, as ordered compact totally disconnected Hausdorff spaces.

Extensive work on dualities has been done by, *e.g.*, Brian Davey, from the perspective of universal algebra, and by Hans Porst, from a categorical perspective [needs to be expanded].

For the rest of this Chapter, we consider a fixed duality $\langle F, G, \gamma, \delta\rangle$ between $\mathscr{C}$ and $\mathscr{D}$.

**4.0.04 Proposition.** *Up to isomorphism there exists a bijective correspondence between monads* $\boldsymbol{T} = \langle T, \eta, \mu\rangle$ *on* $\mathscr{D}$ *and co-monads* $\boldsymbol{B} = \langle B, \varphi, \nu\rangle$. *Moreover, the duality between* $\mathscr{D}$ *and* $\mathscr{C}$ *lifts to a duality between the categories* $\mathscr{D}^{\boldsymbol{T}}$ *of EM-algebras and* ${}^{\boldsymbol{B}}\mathscr{C}$ *of EM-co-algebras.*

**Proof.** Given a monad $\boldsymbol{T} = \langle T, \eta, \mu\rangle$ on $\mathscr{D}$, combine the co-monad $\boldsymbol{T}^{\mathrm{op}} = \langle T^{\mathrm{op}}, \eta^{\mathrm{op}}, \mu^{\mathrm{op}}\rangle$ with $\gamma^{-1}$ and $\delta^{\mathrm{op}}$ to obtain a co-monad structure on the endo-functor

$$\mathscr{C} \xrightarrow{F} \mathscr{D}^{\mathrm{op}} \xrightarrow{T^{\mathrm{op}}} \mathscr{D}^{\mathrm{op}} \xrightarrow{G^{\mathrm{op}}} \boldsymbol{C}$$

Rest: HW □

## 4.1 The categorical version of Birkhoff's Theorem

In classical universal algebra one considers signatures $\Sigma$ consisting only of function symbols, and for every set $X$ the corresponding term algebras $XT_\Sigma$. Then equations are simply pairs of terms for some $X$, and a $\Sigma$-algebra $A$ "satisfies" such an equation $\langle e_0, e_1\rangle \in (XT_\Sigma)^2$, if any homomorphism $XT_\Sigma \xrightarrow{h} A$ identifies the terms $e_0$ and $e_1$.

Alternatively, one may consider the quotient $XT_\Sigma/\sim$, where $\sim$ is the smallest congruence on $XT_\Sigma$ containing the pair $\langle e_0, e_1\rangle$. Then the satisfaction of of this equation can equivalently be expressed by requiring that any homomorphism $XT_\Sigma \xrightarrow{h} A$ factors through the quotient $XT_\Sigma \xrightarrow{q} XT_\Sigma/\sim$.

Of course, sensible classes of equations ought to be closed under substitution.

The following is based on [BH76]. Let $\mathscr{D}$ be a complete category with a factorization system $\langle \mathcal{E}, \mathcal{M}\rangle$, let $\hat{\mathscr{D}} \subseteq \mathscr{D}_0$ be a class of objects (where the "variables" of the "equations" live), and let $\boldsymbol{T} = \langle T, \eta, \mu\rangle$ be a monad on $\mathscr{D}$, subject to the following assumptions:

**(B0)** $\langle \mathcal{E}, \mathcal{M}\rangle$ is a proper factorizations system on $\mathscr{D}$;

**(B1)** $T$ preserves the class $\mathcal{E}$;

**(B2)** the free $T$-algebras over $\hat{\mathscr{D}}$-objects are $\mathcal{E}$-projective;

**(B3)** every $T$-algebra is an $\mathcal{E}$-quotient of the free $T$-algebra over some $\hat{\mathscr{D}}$-object;

**(B4)** up to isomorphism free $T$-algebras over $\hat{\mathscr{D}}$-objects admit only a set of $\mathcal{E}$-quotients.

**4.1.01 Definition.** A full subcategory $\mathscr{K}$ of $\mathscr{D}^{\boldsymbol{T}}$ is called an $\langle \mathcal{E}, \mathcal{M}\rangle$-*variety*, if it is closed under $\mathcal{E}^{\boldsymbol{T}}$-quotients, $\mathcal{M}^{\boldsymbol{T}}$-sub-objects and products.

To appreciate the following definition, recall that the Kleisli-category $\mathscr{D}_{\boldsymbol{T}}$ can be identified with the full subcategory of $\mathscr{D}^{\boldsymbol{T}}$ spanned by the free algebras. Henceforth we will be interested in the full subcategory $\hat{\mathscr{D}}_{\boldsymbol{T}}$ of $\mathscr{D}^{\boldsymbol{T}}$ spanned by the free algebras over $\hat{\mathscr{D}}$-objects, and the corresponding inclusion $J$.

**4.1.02 Definition.** An *equational theory* over $\hat{\mathscr{D}}$ is a pair $\mathbb{E} = \langle \varepsilon, Q\rangle$ consisting of a functor $\hat{\mathscr{D}}_{\boldsymbol{T}} \xrightarrow{Q} \mathscr{D}^{\boldsymbol{T}}$ and a natural transformation $J \xRightarrow{\varepsilon} Q$ that point-wise belongs to $\mathcal{E}^{\boldsymbol{T}}$.

A $\boldsymbol{T}$-algebra $A$ *satisfies* $\mathbb{E}$, written as $\mathbb{E} \models A$, provided $A$ is $\varepsilon$-injective. Write $\mathbb{E}$-***mod*** for the full subcategory of $\mathscr{D}^{\boldsymbol{T}}$ of those $\boldsymbol{T}$-algebras that satisfy $\mathbb{E}$.

**4.1.03 Theorem.** *There exists an essentially bijective correspondence between varieties in $\mathscr{D}^{\boldsymbol{T}}$ and equational theories over $\hat{\mathscr{D}}$.*

**Proof.** For an equational theory $\mathbb{E}$ it is quite easy to see that the class $\mathbb{E}$-***mod*** $\varepsilon$-injective objects is closed under $\mathcal{E}$-quotients, $\mathcal{M}$-sub-objects and products, and hence forms a variety (HW).

Conversely, given a variety $\mathscr{K}$ in $\mathscr{D}^{\boldsymbol{T}}$, an object $X \in \hat{\mathscr{D}}$, and a $\boldsymbol{T}$-homomorphism $\langle XT, X\mu\rangle \xrightarrow{h} h\partial_1 \in \mathscr{K}$, *i.e.*, an object of the comma-category $\langle XT, X\mu\rangle \downarrow \mathscr{K}$, consider its $\langle \mathcal{E}^{\boldsymbol{T}}, \mathcal{M}^{\boldsymbol{T}}\rangle$-factorization $h = XT \xrightarrow{e_h} X_h \xrightarrow{m_h} h\partial_1$. By (B4), there exists a small set $F_X \subseteq \langle XT, X\mu\rangle \downarrow \mathscr{K}$ such that any $e_h$ is isomorphic to $e_f$ for some $f \in F_X$.

Define $XR$ to be the product $\prod_{f\in F_X} X_f$ in $\hat{\mathscr{D}}^{\boldsymbol{T}}$, and consider the unique $\boldsymbol{T}$-homomorphism $\langle XT, X\mu\rangle \xrightarrow{X\rho} XR$ that satisfies $X\rho \cdot f\pi = e_f$ for each $f \in F$. Its $\langle \mathcal{E}^{\boldsymbol{T}}, \mathcal{M}^{\boldsymbol{T}}\rangle$-factorization

$$\langle XT, X\mu\rangle \xrightarrow{X\varepsilon} XQ \xrightarrow{X\omega} XR$$

yields a candidate for the desired $\mathcal{E}^{\boldsymbol{T}}$-quotient $X\varepsilon \in \mathcal{E}$. We now claim that $R$ and $Q$ are functors $\hat{\mathscr{D}}_{\boldsymbol{T}} \longrightarrow \mathscr{D}^{\boldsymbol{T}}$, and that $\rho$ and $\varepsilon$ are natural transformations from the inclusion $J$ to $R$, resp., $Q$. (HW) □

## 4.2 Birkhoff + Duality in case of the list monad

In view of the duality between $\mathscr{D}^{\boldsymbol{T}}$ and ${}^{\boldsymbol{B}}\mathscr{C}$ we now obtain a co-equational theory $\mathbb{W} = \langle L, \lambda\rangle$ over ${}_{\boldsymbol{B}}\check{\mathscr{C}}$, the dual of $\hat{\mathscr{D}}_{\boldsymbol{T}}$. Here the components of $\lambda$ belong to the image of $\mathcal{E}^{\boldsymbol{T}}$ under the duality, and therefore are monos.

It will be useful to address the underlying set and the structure morphism of $XQ$ explicitly, hence we set $XQ = \langle X\Omega, X\omega\rangle$.

**4.2.01 Example.** Let $\boldsymbol{T}$ be the list monad on $\mathscr{D} = \boldsymbol{set} = \hat{\mathscr{D}}$ with the factorization system consisting of the class $\mathcal{E}$ of surjections (= regular or strong or extremal epis) and the class $\mathcal{M}$ of all monos.

An equational theory assigns to each set $X$ an $\mathcal{E}^{\boldsymbol{T}}$-quotient $\langle X^*, X\mu\rangle \xrightarrow{X\varepsilon} XQ$ in ***mon***, and the corresponding surjection $X^* \xrightarrow{X\varepsilon} XQ$, and this assignment is natural in $\boldsymbol{set}_{\boldsymbol{T}} \subseteq \boldsymbol{set}^{\boldsymbol{T}}$.

Dually, in the category ***caba***, the objects are essentially power-sets, and the dual of a free monoid $\langle X^*, X\mu\rangle$ is a co-free co-algebra $\langle X^*P, X\mu^{\leftarrow}\rangle$ in ***caba***. (Do not get confused by the fact that the objects of ***caba*** are certain Boolean *algebras*!) .

The category ${}_{\boldsymbol{B}}\check{\mathscr{C}}$ has precisely these free quantales over sets as objects, while the morphisms are the inverse image functions of monoid homomorphisms $X^* \xrightarrow{h} Y^*$. These, in turn, are the unique extensions $h = X\eta \cdot fT \cdot Y\mu$ of *substitutions* $X \xrightarrow{f} Y^*$ in ***set***.

Now the $\mathcal{E}^{\boldsymbol{T}}$-quotient $\langle X^*, X\mu\rangle \xrightarrow{X\varepsilon} XQ$ has an underlying surjection $X^* \xrightarrow{X\varepsilon} X\Omega$ in $\mathcal{E}$, whose inverse image function, *i.e.*, ***caba***-morphism, $X\Omega P \xrightarrow{X\varepsilon^{\leftarrow}} X^*P$ is an injective $\mathcal{B}$-co-algebra homomorphism.

Hence the co-equational theory amounts to a family of languages $XL \xrightarrow{X\iota} X^*P$ that are

- ▷ complete Boolean sub-algebras (due to injective ***caba*** - morphisms);
- ▷ sub - $\boldsymbol{B}$ - co-algebras;
- ▷ and are closed under inverse images along extensions of substitutions, *i.e.*, homomorphisms between free monoids (due to naturality). ◁

Of these conditions it is not immediately clear how to check the second one. In view of the duality it suffices to understand, under which conditions a surjective function $X^* \xrightarrow{e} Z$ induces a monoid structure $\zeta$ on $Z$ such that $e$ becomes a monoid homomorphism from $\langle X^*, X\mu \rangle$ to $\langle Z, \zeta \rangle$ .

## 4.3 From varieties to pseudo-varieties

# 5 Appendix: categorical foundations

It is useful to consider "all sets and functions" as well as "all monoids and homomorphisms", "all ordered sets and order-preserving functions", "all topological spaces and continuous functions" and other collections of "structures sets" and "structure-preserving" functions of this type as single mathematical entities, called "categories". In contrast to familiar algebraic or order-theoretic structures, these will be 2-sorted: we distinguish "objects", like sets or monoids, and "arrows", like functions or homomorphisms. The the latter point from some object to another, hence can be assigned a "domain" or "source" and a "codomain" or "target". These data specify a directed graph, possibly large, *i.e.*, with proper classes of nodes and edges.

Abstracting from these concrete examples, categories are directed graphs with a sensible composition operation on the edges, or arrows. Among these there are two extreme but very familiar cases: monoids, corresponding to single-node graphs, and pre-ordered sets (not necessarily anti-symmetric), with at most one arrow between any two nodes. Hence categories can be thought of as the common generalization of these two concepts.

**Terminology.** Unfortunately, in mathematics as well as in computer science the term "graph" is overloaded with a number of different concepts. Especially in view of the mathematical discipline of "graph theory", which only deals with a rather restricted notion of graph compared to the one of interest in category theory, in 1972 Peter Gabriel [**?**] (not from Genesis) suggested the name "quiver" (German: "Köcher") instead. This is increasingly used in category theory (*e.g.*, the nLab), but has not spread to other areas. For this script we will stick to the term "graph" and explicitly point out the differences with graph theory.

## 5.0 Comprehension and graphs

It is well-known that for any set $X$ the the *subsets* $U \subseteq X$ bijectively correspond to *characteristic functions* $X \xrightarrow{U\chi} 2 = \{0,1\}$; this can be seen as a simple form of *comprehension*. Let us identify the subset $U \subseteq X$ with the injective inclusion function $U \xrightarrow{i} X$. What about general injective functions into $X$? We think of any two such as *essentially the same*, if their *images* in $X$ coincide; this obviously specifies an equivalence relation on all injective functions into $X$. From this point of view the subsets of $X$ are just canonical representatives of the equivalence classes.

But while an injection $U \xrightarrow{i} X$ is essentially determined by which pre-images of $X$-elements are singletons, respectively, empty, a general function $U \xrightarrow{f} X$ also is precisely determined by the family of pre-images $af^{-1}$, $a \in X$; these are pairwise disjoint sets, possibly empty, partitioning $U$. Conversely, given any $X$-indexed family of sets, which we think of as a function $X \xrightarrow{\varphi} \boldsymbol{set}$ into the proper class of all sets, their disjoint union $\sum_{a \in X} a\varphi$ admits a canonical function into $X$. (To obtain a precise match with the situation for injective functions one would have to consider families of *cardinal numbers* as canonical representatives of sets of a given size, but we will refrain here from this extra complication.)

The first part of this observation allows us to view relations $R \subseteq A \times B$ alternatively as binary $A \times B$-matrices, or $A \times B$-indexed families of truth-values. Similarly, instead of general functions $R \xrightarrow{\rho} A \times B$ between sets we can work with set-valued $A \times B$-matrices, or functions $A \times B \xrightarrow{\bar{\rho}} \boldsymbol{set}$.

Using the obvious projections from $A \times B$ to $A$, resp., $B$, a function $R \xrightarrow{\rho} A \times B$ equivalently can be expressed as a *span* of functions $A \xleftarrow{\partial_0} R \xrightarrow{\partial_1} B$. Here we think of $R$ as a global set of *arrows*. These originate in points of $A$, their *domains* determined by $\partial_0$, and terminate in points of $B$, their *codomain* determined by $\partial_1$. (For disjoint $A$ and $B$ this corresponds to a *bipartite graph*.) On the other hand, a function $A \times B \xrightarrow{\bar{\rho}} \boldsymbol{set}$ encodes a local view: each pair of objects $\langle a, b\rangle \in A \times B$ is equipped with a set of arrows from $a$ to $b$. These so-called “hom-sets” need not be disjoint as they just parameterize the various arrow sets.

Directed graphs correspond to the case $A = B$. Relations on $A$ my be seen as arrows linking the elements of $A$ with at most one such link; loops are allowed. In contrast, graph theory mostly deals with reflexive symmetric relations, *i.e.*, undirected graphs without loops.

In order to allow parallel arrows one has to move to spans on $A$. (Graph-theorists speak of “multi-graphs” in case of loops and parallel arrows; unfortunately this term as well has a different meaning in category theory.)

It turns out that the hom-set-view of graphs lends itself better to generalizations of the concept of category that arise later.

**Size considerations**: Notice, however, that the approach to graphs and categories outlined above is limited to “small” graphs, resp., categories with a set of objects. In contrast, already the category $\boldsymbol{set}$ is large, *i.e.*, has a proper class of objects. But at least the collection of functions between any two fixed sets is small. In this case we speak of “locally small” graphs, resp., categories. Occasionally, even the arrows between two objects may fail to form a set. Consider, for example, the spans $R \xrightarrow{\rho} A \times B$ between two sets $A$ and $B$. Since $R$ is not limited in size, there is a proper class of such spans. In general, category theory has to be careful to avoid Russel-type paradoxes. One possible solution is the use of several so-called “Grothendieck universes”, classes of sets whose cardinality belong to some fixed *strongly inaccessible cardinal* and which therefore have nice closure properties. However, in these notes we will not discuss this issue further and limit ourselves to “sets” (also called “small sets”) and “classes” (also called “large sets”).

## 5.1 Categories

**5.1.00 Definition.** A *locally small span*

$$\boldsymbol{A}_0 \times \boldsymbol{B}_0 \xrightarrow{\boldsymbol{S}} \boldsymbol{set}$$

from a set $\boldsymbol{A}_0$ to a set $\boldsymbol{B}_0$ (both potentially large), written as $\boldsymbol{A}_0 \xrightarrow{\boldsymbol{S}} \boldsymbol{B}_0$, assigns to every pair $\langle A, B\rangle \in \boldsymbol{A}_0 \times \boldsymbol{B}_0$ of *objects* or *0-cells* a local *hom-set* of *arrows* or *morphisms* or *1-cells* from $A$

to $B$. We often write $A \xrightarrow{f} B$ instead of $f \in \langle A, B\rangle \boldsymbol{S}$. Call such a span *small*, if both $\boldsymbol{A}_0$ and $\boldsymbol{B}_0$ are small. The *global set of arrows* refers to the disjoint union of all hom-sets.

If source and target coincide, we speak of a *locally small*, resp., *small graph* $\boldsymbol{C}_0 \times \boldsymbol{C}_0 \xrightarrow{\mathscr{C}} \boldsymbol{set}$ (now denoted by a script letter).

Moreover, $\mathscr{C}$ is a *locally small*, resp., *small category*, if in addition there are families

- ▷ $1 \xrightarrow{A\mathscr{C}} \langle A, A\rangle\mathscr{C}$, $A \in \boldsymbol{C}_0$ of distinguished *identity morphisms* (also denoted as $id_A$), and
- ▷ $\langle A, B\rangle\mathscr{C} \times \langle B, C\rangle\mathscr{C} \xrightarrow{\langle A,B,C\rangle\mathscr{C}} \langle A, C\rangle\mathscr{C}$, $A, B, C \in \boldsymbol{C}_0$ of *local composition operations*, usually shortened to "$\cdot$",

subject to the following axioms:

**(CT-0)** identities are neutral w.r.t. local composition, *i.e.*, $f \cdot B\mathscr{C} = f = A\mathscr{C} \cdot f$ for $A \xrightarrow{f} B$;

**(CT-1)** local composition is associative, *i.e.*, $(f \cdot g) \cdot h = f \cdot (g \cdot h)$ for $A \xrightarrow{f} B \xrightarrow{g} C \xrightarrow{h} D$.

Dropping the requirement for identity morphism leads to the notion of *semi-category*, the many-object generalization of a semi-group.

**5.1.01 Examples.**

(0) The paradigmatic category is $\boldsymbol{set}$ with sets as objects and functions as morphisms. $\boldsymbol{set}$ is a locally small category with a proper class of objects. Function composition is associative and has identity functions as neutral elements.

(1) There are many categories of structured sets with structure-preserving functions as morphisms. This notion will be made precise in Definition **??** below. Often these can be classified as *algebraic* in nature like the categories

- ▷ $\boldsymbol{mon}$ of monoids and monoid homomorphism;
- ▷ $\boldsymbol{grp}$ of groups and group homomorphism;
- ▷ $\boldsymbol{lat}$ of lattices and lattice homomorphisms;
- ▷ $\boldsymbol{bool}$ of Boolean algebras and Boolean homomorphisms;
- ▷ $\boldsymbol{ring}$ of rings and ring homomorphisms;

or topological in nature, like the categories

- ▷ $\boldsymbol{top}$ of topological spaces and continuous maps;
- ▷ $\boldsymbol{met}$ of metric spaces and non-expanding maps;
- ▷ $\boldsymbol{pre}$ of pre-ordered sets and order-preserving functions;

- ▷ $\boldsymbol{vec}(K)$ of vector spaces over a field $K$ and linear functions;
- ▷ $\boldsymbol{cpo}$ of complete partial orders and continuous functions.

(2) Categories of structured sets cannot always be named after their objects, since the same class of objects may admit various sensible choices of morphisms:

- ▷ $\boldsymbol{rel}$ has sets as objects, but binary relations as morphisms; as we will need to distinguish the composition of relations from the generic composition $\cdot$ in a category $\mathscr{C}$, we write $\star$ for the composition of relations, *i.e.*, for $A \xrightarrow{R} B \xrightarrow{S} C$

$$R \star S = \{ \langle a, c \rangle \; : \; \exists b \in B. \; aRbSc \}$$

- ▷ $\boldsymbol{prt}$ has sets as objects, but partial functions as morphisms; clearly $\boldsymbol{set}$ is contained in $\boldsymbol{prt}$, which is contained in $\boldsymbol{rel}$, and the composition operation is the same in all cases.
- ▷ $\boldsymbol{spn}$ has sets as objects, but spans as morphisms (the composition can be realized as the matrix-product of the corresponding set-valued matrices). In this case all hom-sets fail to be small, as the cardinality of the sets in such matrices is unbounded. And while $\boldsymbol{rel}$ sits inside of $\boldsymbol{spn}$, the composition of relations in $\boldsymbol{spn}$ may not result in a relation anymore: the multiplication of set-valued matrices in $\boldsymbol{spn}$ uses disjoint union and Cartesian product, whereas the corresponding operation on binary matrices in $\boldsymbol{rel}$ uses Boolean join $\vee$ and meet $\wedge$.
- ▷ *Complete lattices*, *i.e.*, (small) lattices where every subset has a *supremum* (or *least upper bound*) can alternatively be characterized by every subset having an *infimum* (or *greatest lower bound*). For morphisms, one can require the preservation of suprema, or of infima (notice that these requirements are not equivalent!), or of both suprema and infima. The corresponding categories may be called $\bigsqcup$-$\boldsymbol{slat}$, $\bigsqcap$-$\boldsymbol{slat}$ and $\boldsymbol{clat}$, respectively. (Here $\boldsymbol{s}$ indicates "semi", while $\boldsymbol{c}$ indicates "complete". Note that both $\bigsqcup$-semi-lattices and $\bigsqcap$-semi-lattices are in fact complete lattices, but considered in contexts where the morphisms are only guaranteed to preserve suprema, resp., infima.)

(3) As mentioned before, the objects of a category need not be structured sets:

- ▷ Every monoid $\mathscr{M} = \langle M, \cdot, e \rangle$ is a category with a single object $*$, the morphism-set $M$, and identity morphism $* \xrightarrow{e} *$. In particular, this includes the so-called *terminal category* $\mathbf{1}$ with one object and one morphism (and one $n$-cell for each $n > 1$).
- ▷ Every pre-ordered set $\langle P, \leq \rangle$, where $\leq \; \subseteq P \times P$ is a reflexive and transitive relation, is a small category. Objects are the elements of $P$, while there is at most one arrow between any two objects: $p \longrightarrow q$ iff $p \leq q$. This includes the category $\mathbf{2}$ with $0 < 1$ the only non-identity arrow. Special cases include the *discrete category* and the *indiscrete category* on a set $X$ of objects. In the locally small case, proper classes of objects are allowed.

- ▷ Every directed graph $\boldsymbol{C}_0 \times \boldsymbol{C}_0 \xrightarrow{\mathscr{C}} \boldsymbol{set}$ gives rise to the so-called *free category* category $\mathscr{C}^*$ with the same objects and the *directed paths* in $\mathscr{C}$ as morphisms. This generalizes the construction of free monoids over sets (viewed as hom-sets of directed graphs with a single node) to general directed graphs.

**5.1.02 Remarks.**

- ▷ For every category $\mathscr{C}$ and each $\mathscr{C}$-object $A$ the hom-set $\langle A, A \rangle \mathscr{C}$ is automatically a monoid under composition, with $A\mathscr{C} = id_A$ as neutral element.

- ▷ Replacing for a given category the non-empty hom-sets by $1 \in \mathbf{2}$ results in a (possibly large) pre-ordered set, the so-called “posettal collapse” of $\mathscr{C}$.

The fact that certain composites of 1-cells coincide in a category is often expressed in terms of so-called “commutative diagrams”:

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ {\scriptstyle h}\downarrow & & \downarrow{\scriptstyle g} \\ C & \xrightarrow[k]{} & D \end{array} \tag{5.1-00}$$

expresses the fact that $f \cdot g = h \cdot k$.

Invertible elements of monoid play an important role. The same is true for invertible arrows in categories:

**5.1.03 Definition.**

- ▷ An arrow $A \xrightarrow{f} B$ in a category $\mathscr{C}$ is called *isomorphism*, or *iso* for short, if there exists some $B \xrightarrow{g} A$ with $f \cdot g = A$ and $g \cdot f = B$.

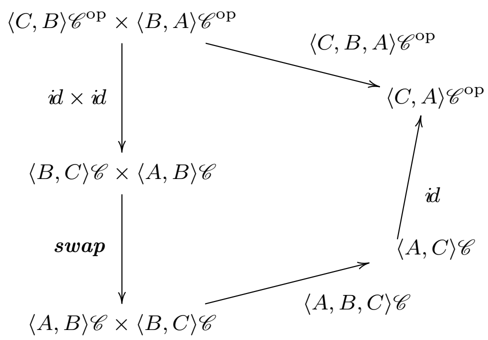
- ▷ Two $\mathscr{C}$-objects $A$ and $B$ are called *isomorphic*, if there exists an iso $A \xrightarrow{f} B$.

- ▷ $\mathscr{C}$ is called *essentially small*, if the collection of isomorphism-classes is in bijective correspondence with a set.

Isomorphisms are a weakening of the notion of equality. It will turn out that all relevant concepts inside a category will only be defined “up to isomorphism”. Further weakenings become available as more levels of structure are considered.

In directed graphs reversing all arrows is an elementary operation.

**5.1.04 Definition.** The *opposite* span $\boldsymbol{B}_0 \times \boldsymbol{A}_0 \xrightarrow{\boldsymbol{S}^{\mathrm{op}}} \boldsymbol{set}$ of a span $\boldsymbol{A}_0 \times \boldsymbol{B}_0 \xrightarrow{\boldsymbol{S}} \boldsymbol{set}$ is obtained by pre-composing $\boldsymbol{S}$ with the swap-function $\boldsymbol{B}_0 \times \boldsymbol{A}_0 \xrightarrow{\boldsymbol{swap}} \boldsymbol{A}_0 \times \boldsymbol{B}_0$. *Opposite graphs* and *opposite categories* are defined accordingly. For the opposite category $\mathscr{C}^{\mathrm{op}}$ the composition also involves suitable swap functions:




(5.1-01)

◁

**5.1.05 Definition.**

- ▷ $\boldsymbol{C}_0 \times \boldsymbol{C}_0 \xrightarrow{\mathscr{C}} \boldsymbol{set}$ is a *sub-graph* of $\boldsymbol{D}_0 \times \boldsymbol{D}_0 \xrightarrow{\mathscr{D}} \boldsymbol{set}$, if $\boldsymbol{C}_0 \subseteq \boldsymbol{D}_0$ and $\langle A, B\rangle\mathscr{C} \subseteq \langle A, B\rangle\mathscr{D}$, for all $\langle A, B\rangle \in \boldsymbol{C}_0 \times \boldsymbol{C}_0$.
- ▷ In case that $\langle A, B\rangle\mathscr{C} = \langle A, B\rangle\mathscr{D}$, for all $\langle A, B\rangle \in \boldsymbol{C}_0 \times \boldsymbol{C}_0$, we speak of a *full sub-graph*.
- ▷ For $\mathscr{C}$ to be a *(full) sub-category* of $\mathscr{D}$, in addition we need that the $\mathscr{C}$-composition is a restriction of the $\mathscr{D}$-composition, and that the identity-morphisms coincide. ◁

**5.1.06 Definition.** The *product* $\mathscr{C} \times \mathscr{D}$ of two graphs $\boldsymbol{C}_0 \times \boldsymbol{C}_0 \xrightarrow{\mathscr{C}} \boldsymbol{set}$ and $\boldsymbol{D}_0 \times \boldsymbol{D}_0 \xrightarrow{\mathscr{D}} \boldsymbol{set}$ is defined component-wise on the object-set $\boldsymbol{C}_0 \times \boldsymbol{D}_0$, *i.e.*,

$$\langle\langle A, B\rangle, \langle A', B'\rangle\rangle(\mathscr{C} \times \mathscr{D}) := \langle A, A'\rangle\mathscr{C} \times \langle B, B'\rangle\mathscr{D}$$

In case of a *product category* the composition has to operate component-wise as well. ◁

## 5.2 Functors and profunctors

In the spirit of category theory, besides categories, we need morphisms between categories. Actually, there are various candidates, but the most prominent ones are called "functors" as they are generalizing structure-preserving functions. They specialize to monoid homomorphisms when both categories have just one object, and to order-preserving functions, if both categories are pre-orders. These two conditions uniquely determine the notion of functor. Notice, however, that order-ideals between pre-ordered sets also can be generalized to categories, which provides a more general notion of morphism than functors that can also be useful.

**5.2.00 Definition.**

(0) A *span-morphism* $\boldsymbol{S} \overset{\varphi}{\Longrightarrow} \boldsymbol{T}$ from $\boldsymbol{A}_0 \times \boldsymbol{B}_0 \xrightarrow{\boldsymbol{S}} \boldsymbol{set}$ to $\boldsymbol{A}_0 \times \boldsymbol{B}_0 \xrightarrow{\boldsymbol{T}} \boldsymbol{set}$ is just a family of functions $\langle A,B\rangle\boldsymbol{S} \xrightarrow{\langle A,B\rangle\varphi} \langle A,B\rangle\boldsymbol{T}$, $\langle A,B\rangle \in \boldsymbol{A}_0 \times \boldsymbol{B}_0$. It can also be viewed as a function-valued $(\boldsymbol{A}_0 \times \boldsymbol{B}_0)$-matrix.

(1) In contrast, a *graph morphism* $\mathscr{C} \xrightarrow{\boldsymbol{F}} \mathscr{D}$ from

$$\boldsymbol{C}_0 \times \boldsymbol{C}_0 \xrightarrow{\mathscr{C}} \boldsymbol{set} \qquad \text{to} \qquad \boldsymbol{D}_0 \times \boldsymbol{D}_0 \xrightarrow{\mathscr{D}} \boldsymbol{set}$$

consists of an object-function $\boldsymbol{C}_0 \xrightarrow{F} \boldsymbol{D}_0$, and a family $\langle A,B\rangle\mathscr{C} \xrightarrow{\langle A,B\rangle\boldsymbol{F}} \langle AF,BF\rangle\mathscr{D}$ of functions. It is called *faithful*, resp., *full*, if all functions $\langle A,B\rangle\boldsymbol{F}$ are injective, resp., surjective.

(2) A graph morphisms between categories is called a *functor*, if it preserves the identity morphisms and the composition, *i.e.*,

$$\begin{array}{ccc} & 1 & \\ {}^{B\mathscr{C}}\swarrow & & \searrow^{BF\mathscr{D}} \\ \langle B,B\rangle\mathscr{C} & \xrightarrow[\langle B,B\rangle\boldsymbol{F}]{} & \langle BF,BF\rangle\mathscr{D} \end{array} \tag{5.2-00}$$

respectively,

$$\begin{array}{ccc} \langle A,B\rangle\mathscr{C} \times \langle B,C\rangle\mathscr{C} & \xrightarrow{\langle A,B\rangle\boldsymbol{F} \times \langle B,C\rangle\boldsymbol{F}} & \langle AF,BF\rangle\mathscr{C} \times \langle BF,CF\rangle\mathscr{D} \\ {\scriptstyle\langle A,B,C\rangle\mathscr{C}}\downarrow & & \downarrow{\scriptstyle\langle AF,BF,CF\rangle\mathscr{D}} \\ \langle A,C\rangle\mathscr{C} & \xrightarrow[\langle A,C\rangle\boldsymbol{F}]{} & \langle AF,CF\rangle\mathscr{C} \end{array} \tag{5.2-01}$$

(3) A *concrete category* is a pair $\langle\mathscr{C},U\rangle$ consisting of a category $\mathscr{C}$ and a faithful functor $\mathscr{C} \xrightarrow{|-|} \boldsymbol{set}$. ◁

The notion of concrete category is intended to make precise the somewhat informal idea of a category of structured sets and structure preserving functions as morphisms.

**5.2.01 Examples.**

(0) For every category $\mathscr{C}$ there is an *identity functor* $\mathscr{C} \xrightarrow{\mathrm{id}_{\mathscr{C}}} \mathscr{C}$, for convenience often also denoted by $\mathscr{C}$, that leaves objects and morphisms invariant.

(1) While in concrete categories $\langle\mathscr{C},U\rangle$ the functor $U$ "forgets" all structure, in the case of several interacting structures it is, of course, possible, to forget only part of the structure. If, say, $\boldsymbol{pomon}$ is the category of partially ordered monoids with order-preserving monoid

homomorphisms, there are obvious functors ***pos*** $\longleftarrow$ ***pomon*** $\longrightarrow$ ***mon*** that forget the monoid structure, resp., the order structure. The same phenomenon also occurs, *e.g.*, for the category ***topgrp*** of topological groups with continuous group homomorphisms.

**5.2.02 Theorem.**

(0) *For fixed sets* $\boldsymbol{A}_0$ *and* $\boldsymbol{B}_0$ *the spans* $\boldsymbol{A}_0 \xrightarrow{\boldsymbol{S}} \boldsymbol{B}_0$ *and the span-morphisms* $\boldsymbol{S} \xRightarrow{\varphi} \boldsymbol{T}$ *form a locally small category.*

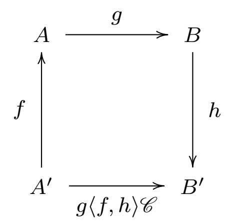
(1) *Small graphs and graph morphisms form a locally small category* ***grph*** .

(2) *Small categories and functors form a locally small category* ***cat*** .

**Proof.** HW. □

By abuse of notation we will often denote the identity functor $\mathscr{C} \xrightarrow{id_{\mathscr{C}}} \mathscr{C}$ just by $\mathscr{C}$ .

**5.2.03 Example.** The assignment of hom-sets $\boldsymbol{C}_0 \times \boldsymbol{C}_0 \xrightarrow{\mathscr{C}}$ ***set*** of a category is the object function of the so-called *hom-functor* $\mathscr{C}^{\mathrm{op}} \times \mathscr{C} \xrightarrow{\mathscr{C}}$ ***set***, also called $\mathscr{C}$ by abuse of notation. It maps a pair of $\boldsymbol{C}$-morphisms $A' \xrightarrow{f} A$ and $B \xrightarrow{g} B'$ to the function $\langle A, B\rangle\mathscr{C} \xrightarrow{\langle f,h\rangle\mathscr{C}} \langle A', B'\rangle\mathscr{C}$ that operates by pre-composing and post-composing $A \xrightarrow{g} B$ with $f$ and $h$, respectively.




(5.2-02)

The functors $\mathscr{C}^{\mathrm{op}} \xrightarrow{\langle -,B\rangle\mathscr{C}}$ ***set*** and $\mathscr{C} \xrightarrow{\langle A,-\rangle\mathscr{C}}$ ***set*** resulting from fixing the second, respectively, the first argument of the hom-functor are called *representable functors*. ◁

**5.2.04 Examples.**

(0) Monoid homomorphisms and order-preserving functions are precisely the functors between monoids, resp., pre-ordered sets. In particular, the components $\langle A, A\rangle F$ with $A \in \mathscr{C}$ are always monoid homomorphisms.

(1) Functors $\mathbf{1} \xrightarrow{C} \mathscr{C}$ can be identified with objects of $\mathscr{C}$ , while functors $\mathbf{2} \xrightarrow{C} \mathscr{C}$ correspond to arrows.

(2) The operation $(-)^*$ on sets of forming the "set of words" over $X$ extends to a functor on ***set*** . There also is a functor from ***set*** to ***mon*** that maps $X$ to the free monoid $\langle X^*, \diamond, \varepsilon_X\rangle$ , and each function $X \xrightarrow{f} Y$ to the induces monoid homomorphism $\langle X^*, \diamond, \varepsilon_X\rangle \xrightarrow{f^*} \langle Y^*, \diamond, \varepsilon_Y\rangle$.

Conversely, there is a "forgetful functor" from ***mon*** to ***set*** that maps a monoid $\mathcal{M} = \langle M, \bullet, e\rangle$ to the set $M$ and a monoid homomorphism to its underlying function.

Similar constructions work for other kinds of "algebras", like magmas, semi-groups, groups, lattices, Boolean algebras *etc.*. Traditionally, this has been studied in the mathematical field of "universal algebra". We will see below how the categorical notion of "monad" captures most of these examples as well.

(3) Adding a new distinguished element to a set amounts to a functor ***set*** $\xrightarrow{+1}$ ***set*** that extends functions $A \xrightarrow{f} B$ by mapping distinguished elements to distinguished elements.

(4) Adding a distinguished element can also be useful for ordered sets. However, there are several possible ways of doing that. The following are most important for our purposes:

- ▷ ***pos*** $\xrightarrow{+\bot}$ ***pos*** adds a new bottom element to every poset, and order-preserving functions are extended so that they preserve the new bottom.
- ▷ ***pos*** $\xrightarrow{+1}$ ***pos*** adds a new element to a every that is incomparable with the original elements. Again, order preserving functions are extended to preserve the new element.

Both functors can be extended to monads on ***ord*** in trivial fashion, and both of the resulting monads can be iterated, as there are trivial distributive laws that interchange the two added objects. However, when we try to combine both monads, there is only a distributive law $(+\bot)(+1) \xRightarrow{\delta} (+1)(+\bot)$, but none in the other direction.

(5) The power-set operation $\mathcal{P}$ induces functors

- ▷ from ***set*** to itself;
- ▷ from ***set*** to the category $\bigvee$***-slat*** of $\bigvee$-semi-lattices (with complete lattices as objects and suprema-preserving functions as morphisms);
- ▷ from ***mon*** to itself
- ▷ from ***mon*** to the category ***uqnt*** of unital quantales (objects are complete lattices with a monoid structure such that the composition preserves suprema, and morphisms are monoid-homomorphisms that in addition preserve suprema);
- ▷ from ***cat*** to itself; here the objects are kept invariant, but the hom-sets are replaced by their power sets.

### 5.2.1 Profunctors

Functors are the standard morphisms between categories are functors, see Definition 5.2.00, a notion certainly biased by the preference of functions over relations in general mathematics. After all, naive category is based on ***set*** which has functions as morphisms, rather than ***rel*** (see Example 5.10.05).

It seems reasonable to generalize functors between categories to more general morphisms in much the same way as (binary) relations generalize functions, or order-ideals generalize monotone functions.

A functor $\mathscr{C} \xrightarrow{F} \mathscr{D}$ maps every $\mathscr{C}$-arrow $A \xrightarrow{f} B$ to a $\mathscr{D}$-arrows $AF \xrightarrow{fF} BF$ in such a way that identity arrows and composition are preserved.

We may indicate the functional assignment $A \mapsto AF$ by introducing new arrows from the $\mathscr{C}$-object $A$ to the $\mathscr{D}$-object $AF$. This effectively "glues" the categories $\mathscr{C}$ and $\mathscr{D}$ together into a graph with the disjoint union of the objects of $\mathscr{C}$ and $\mathscr{D}$ as nodes, and with the disjoint union of the arrows of $\mathscr{C}$ and $\mathscr{D}$ and the new arrows $A \longrightarrow AF$, $A$ a $\mathscr{C}$-object, as arrows. This graph is not yet a category, as no composition between the $\mathscr{C}$-, respectively $\mathscr{D}$-arrows and the new arrows has been defined. Indeed, every new arrow $A \longrightarrow AF$ has to be composed with all $\mathscr{C}$-arrows with codomain $A$, and with all $\mathscr{D}$-arrows with domain $AF$ in suitably associative fashion. Furthermore, one then would expect to obtain commutative squares of the form

$$\begin{array}{ccc} A & \longrightarrow & AF \\ {\scriptstyle f}\big\downarrow & & \big\downarrow{\scriptstyle fF} \\ B & \longrightarrow & BF \end{array}$$

Dropping the requirement that a $\mathscr{C}$-object $A$ can only be linked to a unique $\mathscr{D}$-object, in other words, abstracting from the object-function $\mathscr{C}_0 \xrightarrow{F} \mathscr{D}_0$, one arrives at

**5.2.05 Definition.** By a *profunctor* $\mathscr{C} \xrightarrow{R} \mathscr{D}$ we mean an ordinary functor $\mathscr{C}^{\mathrm{op}} \times \mathscr{D} \xrightarrow{R} \boldsymbol{set}$.

**Interpretation:** Consider the elements of $\langle C, D\rangle R$ as new arrows from the $\mathscr{C}$-object $C$ to the $\mathscr{D}$-object $D$. Functoriality now yields a composition of $\mathscr{C}$-arrows with the new arrows, and of the new arrows with $\mathscr{D}$-arrows, both resulting in new arrows and satisfying suitable associative laws to the effect that the order of composition for

$$C'' \xrightarrow{f} C' \xrightarrow{g} C \xrightarrow{h} D \xrightarrow{h} D' \xrightarrow{k} D''$$

does not matter. In effect, the profunctor $\mathcal{R}$ amounts to a super-category of the disjoint union $\mathscr{C} + \mathscr{D}$ with potentially extra arrows joining $\mathscr{C}$-objects with $\mathscr{D}$-objects, a so-called "gluing" of $\mathscr{C}$ with $\mathscr{D}$.

**5.2.06 Remark.** The composition of profunctors is somewhat problematic: given $\mathscr{A} \xrightarrow{R} \mathscr{B} \xrightarrow{S} \mathscr{C}$, the set $\langle A, C\rangle(R \cdot S)$ ought to be the disjoint union of the sets $\langle A, B\rangle R \times \langle B, C\rangle S$ modulo equalities resulting from $\mathscr{B}$-morphisms. This type of co-limit is also known as a *co-end*.

However, if $\mathscr{B}$ has a proper class of objects, this construction may result in a proper class, *e.g.*, if $\mathscr{B}$ is discrete. This problem disappears, if we restrict ourselves to small categories.

**5.2.07 Examples.**

- ▷ If $\mathscr{C}$ and $\mathscr{D}$ are ordered sets, *i.e.*, categories enriched over $\mathbf{2}$, then a profunctor $\mathscr{C} \times \mathscr{D} \xrightarrow{R} \mathbf{2}$ amounts to combining $\mathscr{C}$ and $\mathscr{D}$ into a new ordered set with the disjoint union of objects, where certain $\mathscr{C}$-objects may be smaller than certain $\mathscr{D}$-objects, but not vice versa. Such $\mathbf{2}$-enriched profunctors are also known as *order ideals*. This name is slightly misleading as "ideals" as defined below are special endo-profunctors, or even sub-1-cells of endo-1-cells carrying a monad structure.

- ▷ If $\mathscr{C}$ and $\mathscr{D}$ are sets $C$ and $D$, respectively, we have $\mathscr{C}^{\mathrm{op}} = \mathscr{C}$, hence a profunctor is just a function $C \times D \xrightarrow{R} \mathbf{2}$, which is essentially a subset of $C \times D$, *i.e.*, an ordinary relation.

- ▷ If $\mathscr{C}$ and $\mathscr{D}$ are monoids $M$ and $N$, a profunctor $M \xrightarrow{R} N$ amounts to a single set $\mathcal{R}$ of new morphisms from the single object of $M$ to the single object of $N$ together with suitable composition functions $M \times R \xrightarrow{\rho_L} R \xleftarrow{\rho_R} R \times N$.

- ▷ For every locally small category $\mathscr{C}$, the hom-functor $\mathscr{C}^{\mathrm{op}} \times \mathscr{C} \xrightarrow{hom} \boldsymbol{set}$ is a profunctor.

- ▷ For every ordinary functor $\mathscr{C} \xrightarrow{F} \mathscr{D}$ we obtain two profunctors $\mathscr{C} \xrightarrow{\hat{F}} \mathscr{D}$ specified by $\mathscr{C}^{\mathrm{op}} \times \mathscr{D} \xrightarrow{F^{\mathrm{op}} \times \mathscr{D}} \mathscr{D}^{\mathrm{op}} \times \mathscr{D} \xrightarrow{hom} \boldsymbol{set}$ and $\mathscr{D} \xrightarrow{\check{F}} \mathscr{C}$ specified by $\mathscr{D}^{\mathrm{op}} \times \mathscr{C} \xrightarrow{\mathscr{D}^{\mathrm{op}} \times F} \mathscr{D}^{\mathrm{op}} \times \mathscr{D} \xrightarrow{hom} \boldsymbol{set}$. With the correct (obvious?) notion of 2-cell between profunctors, these turn out to be adjoint, *i.e.*, $\hat{F} \vdash \check{F}$.

If the new sets of arrows $\langle C, D\rangle R$ are independent from $\mathscr{C}$ and $\mathscr{D}$, the compositions $\rho_L$ and $\rho_R$ with the original arrows have to be defined explicitly. However, in certain cases we can encode the new sets of arrows by actual arrows of some category $\mathscr{A}$ and re-use its composition for $\rho_L$ and/or $\rho_R$.

**5.2.08 Example.** Consider a co-span of functors $\mathscr{C} \xrightarrow{F} \mathscr{A} \xleftarrow{G} \mathscr{D}$ between locally small categories. Any sub-functor $\mathcal{R}$ of $\mathscr{C}^{\mathrm{op}} \times \mathscr{D} \xrightarrow{F^{\mathrm{op}} \times G} \mathscr{A}^{\mathrm{op}} \times \mathscr{A} \xrightarrow{\mathrm{hom}_{\mathscr{A}}} \boldsymbol{set}$ yields a profunctor $\mathscr{C} \xrightarrow{R} \mathscr{D}$. Here by "sub-functor" we mean a natural transformation $R \Longrightarrow (F^{\mathrm{op}} \times G)\mathrm{hom}_{\mathscr{A}}$ with inclusions $\langle C, D\rangle R \subseteq \langle CF, DG\rangle\mathscr{A}$ as components. ◁

## 5.3 Transformations and natural transformations

Just as the order-preserving functions from a pre-ordered set $\langle P, \leq\rangle$ into another pre-ordered set $\langle Q, \sqsubseteq\rangle$ can be ordered point-wise, we expect the functors from some category $\mathscr{C}$ into another category $\mathscr{D}$ to be the objects of a category as well, a so-called "functor-category".

**5.3.00 Definition.** Given graph morphisms $\mathscr{C} \underset{\boldsymbol{G}}{\overset{\boldsymbol{F}}{\rightrightarrows}} \mathscr{D}$, a *transformation* $\boldsymbol{F} \xRightarrow{\tau} \boldsymbol{G}$ is just a family of $\mathscr{D}^*$-arrows $AF \xrightarrow{A\tau} AG$, $A \in \boldsymbol{C}_0$.

If $\mathscr{D}$ is a category, then a transformation $\tau$ is already determined by specifying a family of $\mathscr{D}$-arrows $AF \xrightarrow{A\tau} AG,\ A \in \boldsymbol{C}_0$. A transformation is called *natural*, if for any $\mathscr{C}$-arrow $A \xrightarrow{f} B$ the following diagram commutes:

$$\begin{array}{ccc} AF & \xrightarrow{fF} & BF \\ {\scriptstyle A\tau}\big\downarrow & & \big\downarrow{\scriptstyle B\tau} \\ AG & \xrightarrow[fG]{} & BG \end{array} \tag{5.3-00}$$

◁

The use of $\mathscr{D}^*$-arrows as components of transformations between grph morphisms is necessary to be able to compose such transformations and to have identities for this composition (see below). Considering Diagram (5.3-00) in this situation, although one obtains two finite paths of $\boldsymbol{D}$-morphisms from $AF$ to $BG$, in general these will be different.

**5.3.01 Proposition.** *For two small categories $\mathscr{C}$ and $\mathscr{D}$, the functors $\mathscr{C} \xrightarrow{\boldsymbol{F}} \mathscr{D}$ as objects and the natural transformations as arrows form a locally small so-called* functor category $[\mathscr{C}, \mathscr{D}]$. □

By abuse of notation we often denote the identity natural transformation $F \xRightarrow{id_F} F$ by $F$.

In case $\mathscr{C}$ is not small, the functor category may fail to be locally small (but we will still use it informally).

**5.3.02 Examples.**

(0) Any transformation between order-preserving functions is automatically natural.

(1) A natural transformation between monoid homomorphisms $M \overset{f}{\underset{g}{\rightrightarrows}} N$ is an element $\tau \in N$ such that $mf \cdot \tau = \tau \cdot mg$ for all $m \in M$. ◁

**5.3.03 Examples.**

(0) The family of inclusions $X \xrightarrow{X\eta} X^*$ that map elements $a \in X$ to singleton words over $X$ constitute a natural transformation from (the identity functor on) $\boldsymbol{set}$ to $(-)^*$; recall that by construction $X = X^1$ is a subset of $X^*$.

Furthermore, the family of concatenation maps $X^{**} \xrightarrow{X\mu} X^*$ constitute a natural transformation from $(-)^{**}$ to $(-)^*$.

(1) The family of inclusions $X \xrightarrow{X\,\boldsymbol{inl}} X+1$ constitute a natural transformation from $\boldsymbol{set}$ to $+1$.

Moreover, the family of maps $X+1+1 \longrightarrow X+1$ that collapse the two distinguished points of $X+1+1$ constitute a natural transformation from $(-)+1+1$ to $(-)+1$.

(2) The family of singleton-functions $X \xrightarrow{\{-\}} XP$ that maps elements $a \in X$ to the set $\{a\} \subseteq X$ constitute a natural transformation from ***set*** to the power set functor $P$.

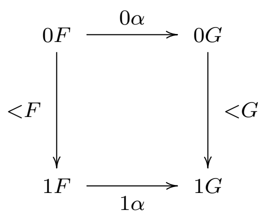
Also the family of union maps $(XP)P \xrightarrow{\bigcup} XP$ that forms the union of a set of subsets of $X$ constitute a natural transformation from $PP$ to $P$. ◁

We will see ins Section 5.6 below how these pairs of natural transformations interact nicely and form so-called "monads" with the underlying functors.

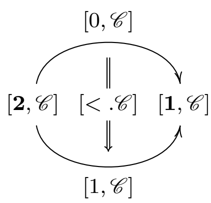
**5.3.04 Examples.**

(0) For any category $\mathscr{D}$ the functor category $[\mathbf{1}, \mathscr{D}]$ is essentially the same as $\mathscr{D}$. In this context natural transformations are just arrows in $\mathscr{D}$.

(1) For any category $\mathscr{D}$ the functor category $[\mathbf{2}, \mathscr{D}]$, also often denoted by $\mathscr{D}^{\to}$, has $\mathscr{D}$-arrows as objects and commutative squares as morphisms:



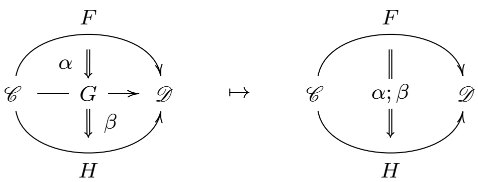

There are obvious domain-and codomain-functors from $[\mathbf{2}, \mathscr{D}]$ to $\mathscr{C}$ that map the square above to $0F \xrightarrow{0\alpha} 0G$, respectively, $1F \xrightarrow{1\alpha} 1G$. More precisely: when we identify $\mathscr{D}$ with $[\mathbf{1}, \mathscr{D}]$ and $\mathbf{2}$ with $[\mathbf{1}, \mathbf{2}]$, the relation $0 < 1$ produces a natural transformation



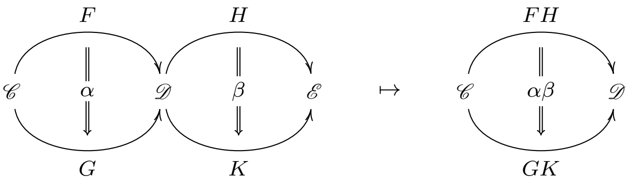

◁

There are two ways how natural transformations can be composed:

▷ *sequentially*:

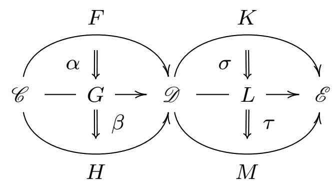


▷ *in parallel*:




which can be expressed as the sequential composition of $FH \xrightarrow{F\beta} FK$ with $FK \xrightarrow{\alpha K} GK$, or equivalently, of $FH \xrightarrow{\alpha H} GK$ with $GH \xrightarrow{G\beta} GK$. These latter obvious parallel composites of (identity natural transformations of) functors with $\alpha$, respectively, $\beta$ are known as *whiskerings*.

Now the obvious question arises, namely whether both possible composites of




(5.3-01)

namely $(\alpha \cdot \beta)(\varphi \cdot \psi)$ and $(\alpha\varphi) \cdot (\beta\psi)$ coincide. If this is the case, we say that the *middle interchange* law is satisfied.

Recall that an ordered set $\boldsymbol{P} = \langle P, \leq \rangle$ can be embedded into the complete lattice of its lower segments under inclusion, and also into the complete lattice of its upper segments under reverse inclusion. In view of ordered sets being $\mathbf{2}$-enriched, this amounts to embeddings

$$[\boldsymbol{P}^{\mathrm{op}}, \mathbf{2}] \xleftarrow{(-)\downarrow} \boldsymbol{P} \xrightarrow{(-)\uparrow} [\boldsymbol{P}, \mathbf{2}]^{\mathrm{op}}$$

via principal lower, resp., upper segments. For ordinary $\boldsymbol{set}$-enriched categories there ought to be similar embeddings, at least in the small case. This is a consequence of the so-called

**5.3.05 Theorem. (Yoneda Lemma)** *For any small category* $\mathscr{C}$ *the functors*

$$[\mathscr{C}^{\mathrm{op}}, \boldsymbol{set}] \xleftarrow{\boldsymbol{Y}} \mathscr{C} \xrightarrow{\boldsymbol{Z}} [\mathscr{C}, \boldsymbol{set}]^{\mathrm{op}}$$

*are full embeddings. Moreover, for any functors* $\mathscr{C}^{\mathrm{op}} \xrightarrow{U} \boldsymbol{set}$ *and* $\mathscr{C} \xrightarrow{V} \boldsymbol{set}$*, respectively, there are natural isomorphisms*

$$\langle -\boldsymbol{Y}, U \rangle [\mathscr{C}^{\mathrm{op}}, \boldsymbol{set}] \cong U \qquad \text{and} \qquad \langle V, -\boldsymbol{Z} \rangle [\mathscr{C}, \boldsymbol{set}] \cong V$$

The functors $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are called the *Yoneda embeddings*, while their values are referred to as *representable functors*. Moreover, $[\mathscr{C}^{\mathrm{op}}, \boldsymbol{set}]$ is often called the category of *pre-sheaves* of $\mathscr{C}$.

**Proof.**

Once the natural isomorphisms for $\boldsymbol{Y}$ and $U$ is established, the one for $\boldsymbol{Z}$ and $V$ follows by duality, while the claim about the fullness of the embeddings is a direct consequence of using representable functors for $U$, respectively $V$.

For every $C \in \mathscr{C}$, we map a natural transformation $C\boldsymbol{Y} \xRightarrow{\varphi} U$ to the value of $id_C$ under its $C$-component $\langle C, C\rangle\mathscr{C} \xrightarrow{C\varphi} CU$. We claim that this is a bijection from $\langle C\boldsymbol{Y}, U\rangle[\mathscr{C}^{\mathrm{op}}, \boldsymbol{set}]$ to $CU$.

Injectivity: If $C\boldsymbol{Y} \xRightarrow{\varphi} U$ differs from $\varphi$, there exists some $A \in \mathscr{C}$ with $A\varphi \neq A\psi$, which in turn implies the existence of some $\mathscr{C}$-morphisms $A \xrightarrow{f} C$ with $fA\varphi \neq fA\psi$. But in view of the naturality of $\varphi$ and $\psi$, the following diagrams in $\boldsymbol{set}$ commute:

$$\begin{array}{ccc}
\langle C, C\rangle\mathscr{C} & \underset{C\psi}{\overset{C\varphi}{\rightrightarrows}} & CU \\
{\scriptstyle \langle f, C\rangle\mathscr{C}}\downarrow & & \downarrow{\scriptstyle fU} \\
\langle A, C\rangle\mathscr{C} & \underset{A\psi}{\overset{A\varphi}{\rightrightarrows}} & AU
\end{array}$$

and hence $fA\varphi \neq fA\psi$ implies $(id_C)C\varphi \neq (id_C)C\psi$, as required.

Surjectivity: Given $x \in CU$, define a natural transformation $C\boldsymbol{Y} \xRightarrow{\bar{x}} U$ by setting

$$\langle A, C\rangle\mathscr{C} \xrightarrow{A\bar{x}} AU \quad , \quad f \mapsto xfU$$

This automatically makes the desired diagrams commute. □

A very useful construction on co-spans of functors $\mathscr{C} \xrightarrow{F} \mathscr{A} \xleftarrow{G} \mathscr{D}$ is the so-called comma-category $F \downarrow G$.

**5.3.06 Definition.** The *comma category* $F \downarrow G$ has

▷ as objects triples $\langle C, f, D\rangle$ with $C \in \mathscr{C}$, $D \in \mathscr{D}$ and $f \in \langle CF, DG\rangle\mathscr{A}$;

▷ as morphisms from $\langle C, f, D\rangle$ to $\langle C', f', D'\rangle$ pairs of morphisms $\langle u, v\rangle \in \langle C, C'\rangle\mathscr{C} \times \langle D, D'\rangle\mathscr{D}$ subject to the condition

$$\begin{array}{ccc}
CF & \xrightarrow{u} & C'F \\
{\scriptstyle f}\downarrow & & \downarrow{\scriptstyle f'} \\
DG & \xrightarrow[v]{} & D'G
\end{array}$$

**5.3.07 Example.**

- ▷ Given a category $\mathscr{C}$ and an object $C$, the comma category $C \downarrow \mathscr{C}$ has as objects all arrows with domain $C$, and as morphisms commutative triangles with domain $C$.

  Dually, $\mathscr{C} \downarrow C$ has as objects all arrows with codomain $C$ and as morphisms commutative triangles with codomain $C$.

- ▷ Given $\mathscr{C}$-objects $C$ and $C'$, the comma category $C \downarrow C'$ coincides with the set $\langle C, C' \rangle \mathscr{C}$. ◁

Associated with a comma category $F \downarrow G$ are domain and codomain functors $\mathscr{A} \xleftarrow{\partial_0} F \downarrow G \xrightarrow{\partial_1}$ $F \downarrow G \xleftarrow{\partial_1} \mathscr{B}$, as well as an obvious natural transformation $\partial_0 F \xrightarrow{F \triangleright G} \partial_1 G$ satisfying the following universal property (HW):

Then for any span of functors $\mathscr{A} \xleftarrow{U_0} \mathscr{X} \xrightarrow{U_1} \mathscr{B}$ and any natural transformation $U_0 F \xrightarrow{\varphi}$ $U_1 G$, there exists a unique functor $\mathscr{X} \xrightarrow{V} F \downarrow G$ such that $\varphi$ factors as $id_V (F \triangleright G)$.

## 5.4 Monoidal categories and enriched categories

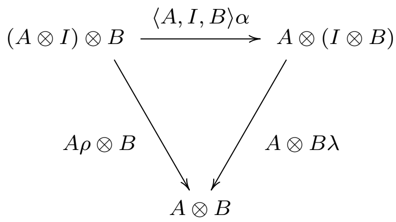
The advantage of considering small graphs as families $\boldsymbol{C}_0 \times \boldsymbol{C}_0 \xrightarrow{\mathscr{C}} \boldsymbol{set}$ of sets is the possibility to replace ***set*** as the category, where the hom-sets live, by some other suitable category $\mathscr{V}$. This must be equipped with a so-called *monoidal structure*, *i.e.*, a counterpart $\otimes$ for $\times$ that is associative and has a unit $I$. Allowing hom-objects to live in $\mathscr{V}$ yields so-called $\mathscr{V}$-*enriched* categories. The first very simple example for such a $\mathscr{V}$ is the ordered set $\mathbf{2}$ with objects $0$ and $1$ and one non-trivial morphism $0 \longrightarrow 1$. Ordered sets then turn out to be enriched over $\mathbf{2}$.

**5.4.00 Definition.** A *monoidal category* $\langle \mathcal{V}, \otimes, I \rangle$ is a category $\mathcal{V}$ equipped with

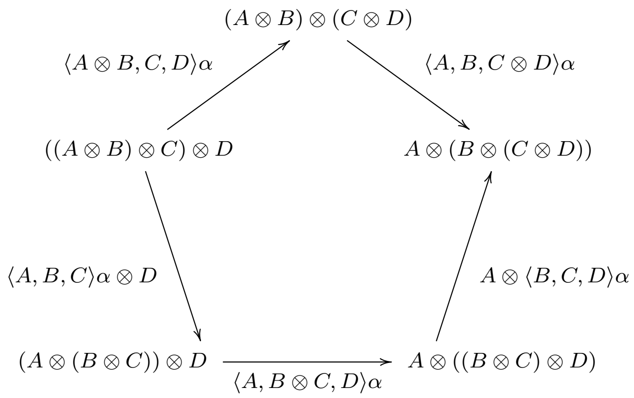
- ▷ a functor $\mathcal{V} \times \mathcal{V} \xrightarrow{\otimes} \mathcal{V}$, called *tensor product*;
- ▷ a so-called *unit object* $I \in \mathcal{V}$;
- ▷ a natural isomorphism $\alpha$ from $\mathcal{V}^3 \xrightarrow{\otimes \times \mathcal{V}} \mathcal{V}^2 \xrightarrow{\otimes} \mathcal{V}$ to $\mathcal{V}^3 \xrightarrow{\mathcal{V} \times \otimes} \mathcal{V}^2 \xrightarrow{\otimes} \mathcal{V}$, as of recently named the *associator*;
- ▷ natural isomorphisms $I \otimes \mathcal{V} \xRightarrow{\lambda} \mathcal{V}$ and $\mathcal{V} \otimes I \xRightarrow{\rho} \mathcal{V}$, nowadays called the *left* and *right unitor*, respectively;

subject to two *coherence conditions*:

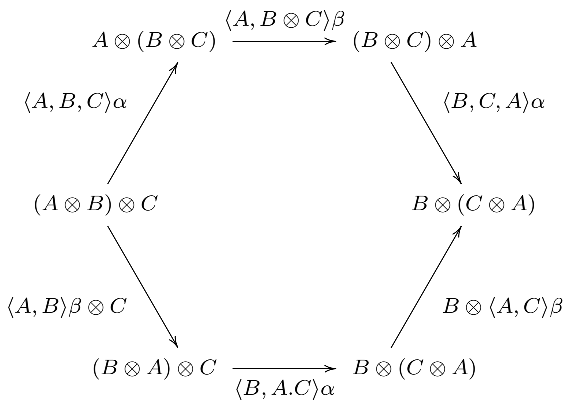
- the *triangle identity*

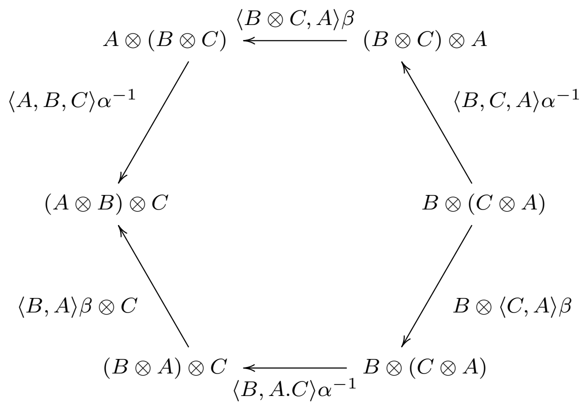


- the *the pentagon identity*




A monoidal category $\langle \mathcal{V}, \otimes, I\rangle$ is called *braided*, if in addition there is a natural isomorphism $A \otimes B \xrightarrow{\langle A,B\rangle\beta} B \otimes A$, the *braiding*, subject to the following *hexagon identities*



,



Finally, $\langle \mathcal{V}, \otimes, I\rangle$ is called *symmetric*, if in addition $\langle B, A\rangle\beta = \langle A, B\rangle\beta^{-1}$; in that case one of the hexagons suffices.

To save space and enhance readability we will often just write $\alpha$, $\lambda$ or $\rho$ without the arguments, when those can easily be recovered from the context.

**5.4.01 Examples.**

(0) An important example of a monoidal category of structured sets, where the tensor operation $\otimes$ differs from the cartesian product, is the category ***ab*** of abelian groups with group homomorphisms.

Given abelian groups $A$ and $B$, their tensor product $A \otimes B$ is defined (uniquely up to isomorphism) by the following "universal property": any *bi-linear* function $A \times B \xrightarrow{f} C$ (*i.e.*, this is a group-morphism in each component separately) uniquely factors through a group homomorphism $A \otimes B \xrightarrow{\bar{f}} C$ by means of a bilinear map $A \times B \xrightarrow{\langle A, B\rangle\beta} A \otimes B$. The neutral element with respect to this tensor product is $\mathbb{Z}$, the free group on one generator.

(1) The notion of a monoidal structure also makes sense for pre-ordered sets. For the sake of simplicity we restrict attention to posets, where the order is anti-symmetric.

Since the hom-sets are at most singletons, the coherence requirements for the associator and the unitors disappear; the only remaining requirement is that the multiplication preserves the order.

Monoidal posets with certain completeness properties occur in algebraic automata theory: monoidal $\sqcup$-semi-lattices are called *dioids*, while monoidal $\bigsqcup$-semi-lattices are known as *quantales*. As we will see below, the sets $MF$ and $MP$ of finite, resp., all subsets of a monoid $M$ form such structures.

**5.4.02 Remark.** The construction of a tensor product by a universal property wrt. "bilinear" maps works in many categories of algebras. The free algebra on one generator will serve as a neutral element. However, to guarantee the associativity of the resulting tensor, one has to require that every algebra operation $A^n \xrightarrow{o} A$ is in fact an algebra homomorphism. Such categories of algebras are called *entropic* [DD85].

**5.4.03 Definition.** Given a monoidal category $\langle \mathcal{V}, \otimes, I\rangle$, a $\mathcal{V}$-*category* $\mathcal{C}$ consists of

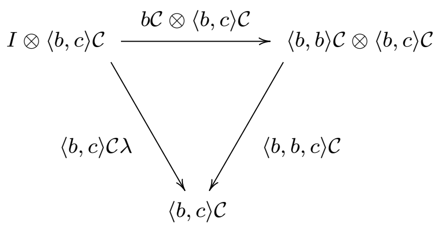
▷ a class $C_0$ of *objects*;

▷ a family $\langle a, b\rangle\mathcal{C}$ of $\mathcal{V}$-objects, called *hom-objects*;

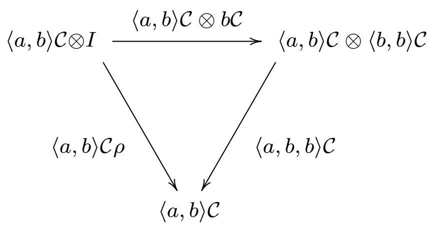
▷ a family of $\mathcal{V}$-morphisms $I \xrightarrow{a\mathcal{C}} \langle a, a\rangle\mathcal{C}$;

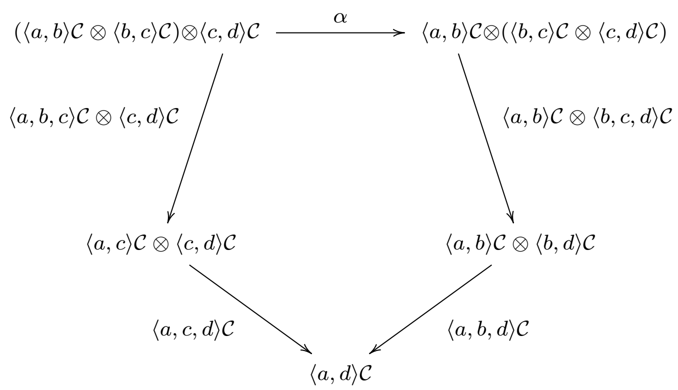
▷ a family of $\mathcal{V}$-morphisms $\langle a, b\rangle\mathcal{C} \otimes \langle b, c\rangle\mathcal{C} \xrightarrow{\langle a,b,c\rangle\mathcal{C}} \langle a, c\rangle\mathcal{C}$

subject to the obvious unit and associativity conditions (note that the number of arguments to $\mathcal{C}$ distinguishes between identities, hom-sets and composition morphisms):




and






**5.4.04 Examples.**

- ▷ ***ab*** -categories are also known as *abelian categories* and play an important role within mathematics.
- ▷ ***pos*** -categories are also known as *order-enriched* categories.
- ▷ Consider the non-negative reals $[0, \infty[$ with the reverse order $\geq$ and the monoidal structure given by addition, and denote this monoidal category by $\mathbb{R}^+$. Lavwere [**?**] identified $\mathbb{R}^+$-categories as *generalized metric spaces* (they do not need to be symmetric). The "hom-number" is just the distance, while the triangle inequality amounts to the composition morphism.

Note that a monoidal category $\langle \mathcal{V}, \otimes, I\rangle$ need not be "self-enriched", or have *internal homs*, *i.e.*, it need not be a $\mathcal{V}$-category itself. Of course, ***set*** and **2** have this property, as does ***cat***, and it seems highly desirable, if $\mathcal{V}$-category theory is supposed to subsume ordinary category theory.

In case that $\mathcal{V}$ does have internal homs, say $A \triangleright B$, we at least expect them to relate to the external homs $\langle A, B\rangle\mathcal{V}$ via

$$\langle I, A \triangleright B\rangle\mathcal{V} \cong \langle A, B\rangle\mathcal{V}$$

Since $I$ is the unit for $\otimes$, one furthermore should expect

$$\langle I, A \triangleright B \triangleright C\rangle\mathcal{V} \cong \langle A, B \triangleright C\rangle\mathcal{V} \cong \langle B \otimes A, C\rangle\mathcal{V}$$

This interplay between the monoidal structure and internal homs is known to computer scientists as "Currying". In case of ***set*** this phenomenon is called "Cartesian closedness", while in **2** it is responsible for the deduction theorems in logic.

As we do not yet have the categorical tools to address this phenomenon, we postpone further discussion until Section **??**. Notice, however, that self-enrichment is possible without a monoidal structure. In computer science the typed $\lambda$-calculus gives an example, which á posteriori can be enhances with pairing, *i.e.*products.

## 5.5 2-categories and suitable 1- and 2-cells

Categories enriched in ***cat***, with cartesian product and $I = \mathbf{1}$ providing the required monoidal structure, are called *bi-categories*, or *weak 2-categories*. Fortunately, it can be shown that each bi-category is bi-equivalent to a strict one; in this case one speaks of *2-categories*. We will restrict our attention to those.

In more elementary terms, 2-categories can be viewed as 2-graphs $\boldsymbol{C}_0 \times \boldsymbol{C}_0 \xrightarrow{\mathscr{C}}$ grph, with nodes in $\boldsymbol{C}_0$ as 0-cells, nodes of the hom-graphs as 1-cells, and arrows of the hom-graphs as 2-cells. For $0 < k < 3$ the $k$-cells have $(k-1)$-cells as domains and codomains. The 2-cells then have to admit two compositions, "horizontal" and "vertical" subject to the middle interchange law (*cf.* Diagram 5.3-01). The horizontal composition, restricted to the identity-2-cells, turns 0-cells and 1-cells into an ordinary category.

**5.5.00 Proposition.**

- ▷ *Just as every set $X$ can be viewed as a discrete graph by means of the hom-functor $X \times X \xrightarrow{\emptyset}$* ***set****, very graph $\boldsymbol{C}_0 \times \boldsymbol{C}_0 \xrightarrow{\mathscr{C}}$* ***set*** *is a "2-discrete 2-graph" w/o 2-cells, also denoted by $\mathscr{C}$.*
- ▷ *For every graph $\mathscr{C}$ its* suspension *contains every $i$-cell of $\mathscr{C}$ as an $(i+1)$-cell and has a single new 0-cell, which of course is the domain and codomain of every 1-cell.* □

**5.5.01 Remark.** The generalization of the previous result from graphs to categories requires some care:

- ▷ In order to view a set $X$ as a discrete category, one has to use the characteristic function of the diagonal $X \times X \xrightarrow{X\Delta_X}$ ***set*** on $X$ as hom-functor; which also amounts to interpreting $X$ as a discrete reflexive graph.
- ▷ For a mere category $\mathscr{C}$ the suspension construction only yields a 2-graph with vertical composition. In order to obtain a horizontal composition as well, one has to start with a monoidal category $\langle \mathscr{C}, \otimes, I \rangle$. In fact, one can establish a bijective correspondence between monoidal categories and bi-categories with one 0-cell, which restricts to strict monoidal categories and 2-categories with one 0-cell, respectively.

**5.5.02 Example.** The familiar construction of a free monoid $X^*$ over some set $X$ should be seen as involving an additional step:

- ▷ consider $X$ as a discrete category, with the elements of $X$ as objects and only the identity morphisms; this may also be seen as a discrete reflexive graph;
- ▷ form the suspension $X$-$\Sigma$, which is a single-node graph with hom-set $X$;

- ▷ form the free category on $X$-$\Sigma$.

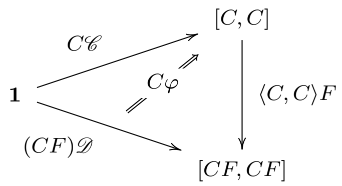
A *2-functor* between 2-categories preserves $i$-cells, $i < 3$, as well as their composition and their units on the nose. Unfortunately, in "nature" a weaker version is much more prevalent, where the functor axioms only hold up to "coherent isomorphism". These are known as *pseudo-functors*.

### 5.5.1 Relaxing the requirements on 2-functors

Besides 2-functors there are (at least) two further reasonable morphisms between 2-categories:

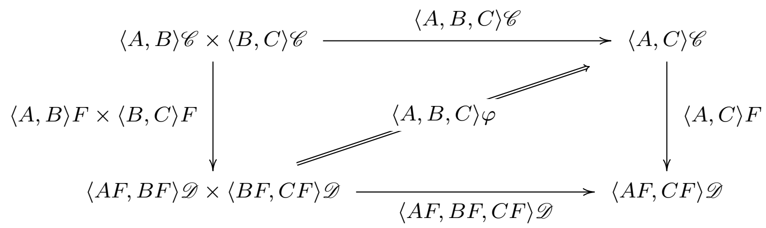
**5.5.03 Definition.** A *lax functor* $\mathscr{C} \xrightarrow{\langle F,\varphi\rangle} \mathscr{D}$ between 2-categories consists of

- ▷ an object-function $\mathscr{C}_0 \xrightarrow{F} \mathscr{D}_0$;
- ▷ a family of functors $\langle A,B\rangle\mathscr{C} \xrightarrow{\langle A,B\rangle F} \langle AF,BF\rangle\mathscr{D}$, for $A,B \in \mathscr{C}_0$;
- ▷ a family of natural transformations




, $C \in \mathscr{C}_0$ ;

- ▷ a family of natural transformations



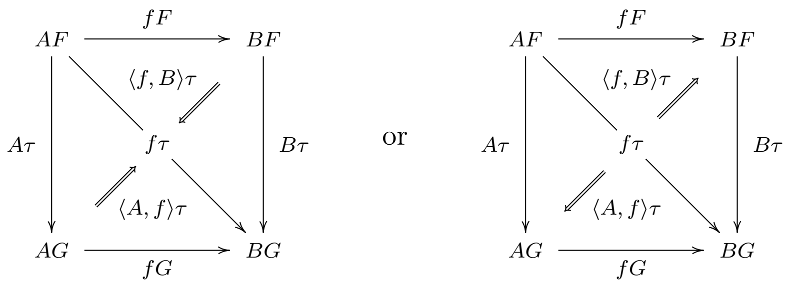

, $A, B, C \in \mathscr{C}_0$ ;

A lax functor is called *normal*, if all natural transformations $C\varphi$, $C \in \mathscr{C}_0$, are identities.
Dual Notion: For an *oplax functor* the natural transformations are reversed.

Concretely, in the lax case, this means that for 1-cells $A \xrightarrow{f} B$ and $B \xrightarrow{g} C$ in $\mathscr{C}$, there are well-behaved 2-cells $fF \cdot gF \xRightarrow{\langle f,g\rangle\langle A,B,C\rangle\varphi} (f \cdot g)\,F$ and $CF\mathscr{D} \xrightarrow{C\varphi} C\mathscr{C}F$. Often we drop the 1-cell indices to simplify the notation.

Since most of the 2-categories needed in this course will be ***ord***-enriched, the naturality conditions above do not need to be mentioned, as they are automatically satisfied. Our main example of lax functors will be lax monoid homomorphisms between ordered monoids.

It is less clear, what "2-natural transformations" ought to be. Diagram 5.3-00 seems inadequate when 2-cells $f \overset{\alpha}{\Longrightarrow} g$ in $\mathscr{C}$ have to be considered. With an extra "dimension" available in the form of 2-cells, it seems reasonable to expect a "2-natural transformations" to assign data not only to the objects, *i.e.*, 0-cells, of the domain, but also to the arrows, *i.e.*, 1-cells. For natural transformations between functors linking categories, in Diagram 5.3-00 one could define $f\tau := fF \cdot B\tau = A\tau \cdot fG$, so here these data can be derived. But for "2-natural transformations" they should be part of the specification. Of course, we then cannot expect the resulting triangles to commute. Instead, one can consider




(5.5-00)

and require suitable compatibility conditions with the 2-cells of $\mathscr{C}$ (HW). We refer to these kind of transformations as *modules*, resp. *opmodules*. Certain special cases are important in applications:

- ▷ modules where $\langle f, B\rangle\tau$ is always an isomorphisms, or opmodules where $\langle A, f\rangle\tau$ is always an isomorphisms, are generally known as *lax natural transformations*;
- ▷ opmodules where $\langle f, B\rangle\tau$ is always an isomorphisms, or modules where $\langle A, f\rangle\tau$ is always an isomorphisms, are generally known as *oplax natural transformations*.

**Warning**: The terminology for these two concepts is not consistent in the literature, not even within the papers of some authors (*cf.*, nlab). Moreover it clashes with the notion of lax and oplax functors (see below), in particular, as both types of transformations can be defined between these types of functors.

## 5.6 Monads

The key observation in Subsection **??** was that the free monoid functor and the power-set functor on $\boldsymbol{set}$, carry extra structure that much resembles a monoid. Here come the official definitions.

**5.6.00 Definition.** A *monad* $\boldsymbol{T} = \langle T, \eta, mu\rangle$ on a category $\mathscr{C}$ consists of an endo-functor $\mathscr{C} \xrightarrow{T} \mathscr{C}$ equipped with natural transformations $TT \xrightarrow{\mu} T$ and $id_{\mathscr{C}} \xrightarrow{\eta} T$ subject to

- ▷ $\mu$ is associative, *i.e.*,

$$\begin{array}{ccc} TTT & \overset{T\mu}{\Longrightarrow} & TT \\ {\scriptstyle \mu T}\Big\Downarrow & & \Big\Downarrow{\scriptstyle \mu} \\ TT & \underset{\mu}{\Longrightarrow} & T \end{array}$$

- ▷ $\eta$ is left- and right-neutral with respect to $\mu$, *i.e.*

$$\begin{array}{ccccc} T & \overset{T\eta}{\Longrightarrow} & TT & \overset{\eta T}{\Longleftarrow} & T \\ & {\scriptstyle id_T}\searrow & \Big\Downarrow{\scriptstyle \mu} & \swarrow{\scriptstyle id_T} & \\ & & T & & \end{array}$$

DUAL NOTIONS: *co-monad* $\boldsymbol{Q} = \langle Q, Q \xrightarrow{\varphi_{\boldsymbol{Q}}} id_{\mathscr{C}}, Q \xrightarrow{\nu_{\boldsymbol{Q}}} QQ \rangle$ with a *co-unit* $\varphi_{\boldsymbol{Q}}$ and a *co-multiplication* $\nu_{\boldsymbol{Q}}$.

**5.6.01 Proposition.** *Monads bijectively correspond to lax functors* $\mathbf{1} \longrightarrow \boldsymbol{Cat}$.

**Proof.** HW! □

Several important monads are based on the various ways to form collections of "elements":

- ▷ as *sets*, where neither the order nor the (positive) multiplicity matters;
- ▷ as *bags* or *multi-sets*, where the order does not matter, but the multiplicity does;
- ▷ as *tuples*, where the order matters (which leaves no room to ignore the multiplicity).

In all three cases, one can form "power-collections", either unconstrained, or finite, or at most singleton collections of elements of a given collection. In CS, mostly the finite collections will be of interest. Moreover, one can distinguish, if the empty collection is allowed or not. This should give rise to a number of monads.

**Power-sets:** The (unconstrained) power-set monad $\boldsymbol{P} = \langle P, \{-\}, \bigcup \rangle$ may be restricted to finite sub-sets, resulting in a monad $\boldsymbol{F}$, and to at most singleton subsets, resulting in the exception monad $\boldsymbol{E}$. The EM-algebras in case of $\boldsymbol{P}$ are $\bigsqcup$-semi-lattices, which are in fact complete, but the morphisms only have to preserve suprema, IN case of $\boldsymbol{F}$ the algebras are $\sqcup$-semi-lattices, and the morphisms preserve joins $\sqcup$. Alternatively, one can think of the objects as idempotent commutative monoids (with the prefix-order). For $\boldsymbol{E}$ the algebras are flat domains with a bottom element; the latter need to be preserved by the morphisms.

Requiring the sets to be non-empty does not cause any problems in the unit or multiplication. In terms of the EM-algebras, the requirement for a least element $\bot$ has to be dropped, which in case of the exception monad results in the identity monad.

**Power-bags:** The singleton- and union-operations carry over from sets to multi-sets. But rather than a supremum-operation, where multiple occurrences of elements do not matter, we now obtain a commutative "addition" that need not be idempotent. Hence in the finite case we obtain as EM-algebras commutative monoids, or commutative semi-groups, if non-empty sets are required. In the infinite case infinite sums are available as well.

**Power-tuples:** In the finite case we recover the list monad $(-)^*$, resp., the non-empty list monad $(-)^+$, with monoids, resp. semi-groups as EM-algebras. The infinite case is considerably more complicated, one has to consider functions from ordinal numbers into the alphabet $X$ in question, and the multiplication of a potential monad would seem to require ordinal addition of ordinal numbers. [Presently I have no idea what the EM-algebras might be.]

However, the countably infinite case is of interest, as this produces streams (= countably infinite words), while the countable case in addition considers finite words. In case of streams the usual unit via singletons is no longer available, and it is also not clear how to handle the multiplication of a countably infinite sequence of streams. It seems to be the case that streams are better handles by so-called co-algebraic methods.

Several of these monads turn out to be connected by so-called "distributive laws" (see Section 5.7 below), which allow the formation of interesting "composite monads".

**5.6.02 Examples.**

(0) For every *finitary algebraic theory* defined by a finite signature $\Sigma$ of function symbols with finite arity the category of $\Sigma$-algebras arises as the the category of EM-algebras for the following monad $\boldsymbol{T}_\Sigma$: the functor $T_\Sigma$ maps a set $X$ to its term algebra, which in particular contains $X$ as the set of "variables". This induces the unit $\eta$, while the multiplication $\mu$ governs the substitution and flattening of terms.

(1) Algebraic theories unbounded signatures can fail to induce to monads: *e.g.*, the theory of complete lattices requires supremum operator ***clat*** of complete lattices does not arise in this fashion (the "free complete lattice" on three generators has a proper class of elements and hence does not exist over ***set***).

(2) On the other hand, *compact Hausdorff spaces* arise as EM-algebras for the ultra-filter monad, an example for a signature with a function-symbol of infinite arity.

**5.6.03 Examples.**

(0) For any monad $\boldsymbol{T} = \langle T, \eta^{\boldsymbol{T}}, \mu^{\boldsymbol{T}} \rangle$ on a category $\mathscr{C}$ the composition $\mathscr{C}^{\boldsymbol{T}} \xrightarrow{U^{\boldsymbol{T}}} \mathscr{C} \xrightarrow{F^{\boldsymbol{T}}} \mathscr{C}^{\boldsymbol{T}}$ automatically yields a co-monad on $\mathscr{C}^{\boldsymbol{T}}$.

**5.6.04 Definition.**

(0) Given an endo-functor $\mathscr{C} \xrightarrow{T} \mathscr{C}$, an *algebra* for $T$ is a pair $\langle X, XT \xrightarrow{\xi} X \rangle$ consisting of a $\mathscr{C}$-object $X$ and a so-called *structure morphism* $\xi$. Hence $T$-algebras are objects of the arrow category $\mathscr{C}^{\rightarrow}$, which coincides with the functor category $[\mathbf{2}, \mathscr{C}]$.

If $\langle Y, \zeta \rangle$ also is a $T$-algebra, a $\mathscr{C}$-morphism $X \xrightarrow{f} Y$ is called an *algebra-homomorphism*, if it preserves the structure maps, *i.e.*,

$$\begin{array}{ccc} XT & \xrightarrow{fT} & YT \\ {\scriptstyle \xi}\downarrow & & \downarrow{\scriptstyle \zeta} \\ X & \xrightarrow[f]{} & Y \end{array} \qquad (5.6\text{-}00)$$

The corresponding category $T$-***alg*** usually will be a *non-full* subcategory of $\mathscr{C}^{\rightarrow} = [\mathbf{2}, \mathscr{C}]$, since not necessarily all commutative squares involving two $T$-algebras arise as algebra homomorphisms.

(1) In case of a monad $\boldsymbol{T} = \langle T, \eta, \mu \rangle$ on $\mathscr{C}$, an *Eilenberg-Moore algebra*, or *EM-algebra* for short, is an algebra $\langle X, XT \xrightarrow{\xi} X \rangle$ for $T$, subject to two compatibility conditions with $\eta$ and $\mu$, respectively:

$$\begin{array}{ccc} X & \xrightarrow{X\eta} & XT \\ & {\scriptstyle id_X}\searrow & \downarrow{\scriptstyle \xi} \\ & & X \end{array} \quad \text{and} \quad \begin{array}{ccc} XTT & \xrightarrow{\xi T} & XT \\ {\scriptstyle X\mu}\downarrow & & \downarrow{\scriptstyle \xi} \\ XT & \xrightarrow[\xi]{} & X \end{array} \qquad (5.6\text{-}01)$$

Notice that $\langle XT, XTT \xrightarrow{X\mu} XT \rangle$ is always an EM-algebra, the so-called *free* EM-algebra over $X$. Furthermore, the structure map $\xi$ of an EM-algebra $\langle X, \xi \rangle$ by definition is always a $\mathscr{C}$-retraction and an algebra homomorphism from the free EM-algebra over $X$ into $\langle X, \xi \rangle$.

Denote by $\mathscr{C}^{\boldsymbol{T}}$ the full subcategory of $T$-***alg*** $<$ spanned by the EM-algebras.

DUAL NOTIONS: *co-algebra* $\langle X, X \xrightarrow{\zeta} XT \rangle$; *co-algebra homomorphism*; *Eilenberg-Moore co-algebra* or *EM-co-algebra*; ${}^{\boldsymbol{T}}\mathscr{C}$.

**5.6.05 Proposition.** *The endo-functor* $\mathscr{C} \xrightarrow{T} \mathscr{C}$ *factors through* $\mathscr{C}^{\boldsymbol{T}}$ *by via*

$$\begin{array}{ccc} & \mathscr{C}^{\boldsymbol{T}} & \\ {\scriptstyle F^{\boldsymbol{T}}}\nearrow & & \searrow{\scriptstyle U^{\boldsymbol{T}}} \\ \mathscr{C} & \xrightarrow[T]{} & \mathscr{C} \end{array}$$

*where* $F^{\boldsymbol{T}}$ *maps* $X \xrightarrow{f} Y$ *in* $\mathscr{C}$ *to the algebra-homomorphism*

$$\begin{array}{ccc} XTT & \xrightarrow{fTT} & YTT \\ {\scriptstyle X\mu}\downarrow & & \downarrow{\scriptstyle Y\mu} \\ XT & \xrightarrow[fT]{} & YT \end{array}$$

*between the free algebras over* $X$, *and* $Y$, *respectively. On the other hand,* $U^{\boldsymbol{T}}$ *maps an algebra-homomorphism (5.6-00) to the* underlying $\mathscr{C}$*-morphism* $X \xrightarrow{f} Y$. □

### 5.6.1 Monads vs. extension systems

There exists an alternative description of monads that doe not use a multiplication $\mu$ and hence avoids repeated application of the functor $T$. Instead it requires morphisms of the form $A \xrightarrow{f} BT$ to be extendable to morphisms $AT \longrightarrow \bar{f}BT$ subject to certain compatibility conditions. [see semantics notes]

This style of working with monads is rather popular in computer science, in particular in functional programming [Haskell].

### 5.6.2 Monads abstractly

The notion of monad can be formulated in any 2-category $\mathscr{B}$: it just amounts to an endo-1-cell $C \xrightarrow{t} C$ equipped with 2-cells $C \xRightarrow{\eta} t \xLeftarrow{\mu} t \cdot t$ subject to the axioms of Definition 5.6.00. In fact, this is nothing but a lax functor $\mathbf{1} \xrightarrow{T} \mathscr{B}$. This allows us to subsume many known phenomena under the monad label:

**5.6.06 Examples.**

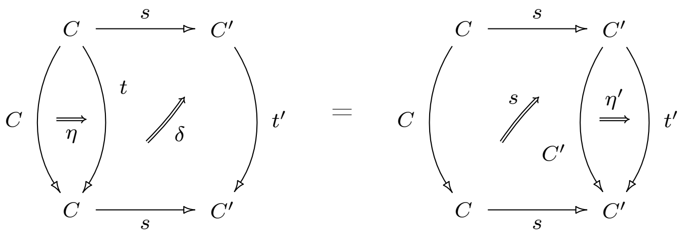
- ▷ Monads in ***rel*** are just pre-orders, *i.e.*, reflexive and transitive relations on a set.
- ▷ Monads in ***pos*** are precisely the closure operators, while co-monads are interior operators.
- ▷ Monads in the suspension of ***set*** are precisely the monoids.

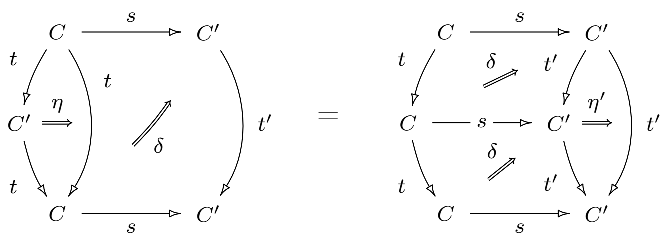
- ▷ Monads in the suspension of ***ab*** (with the tensor product of abelian groups as 1-cell composition) are precisely the rings.
- ▷ Monads in ***spn*** are precisely the small categories.

From this point of view it becomes possible to consider the monads in a 2-category $\mathcal{B}$ as objects of another 2-category. However, there are various sensible notions of morphisms between monads that can be of interest in concrete cases.

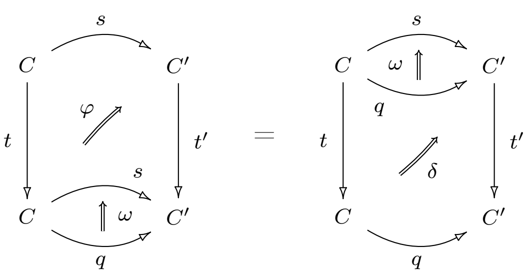
**5.6.07 Definition.** The 2-category $\mathscr{B}$-$\boldsymbol{mnd}_\ell$ of monads in a 2-category $\mathscr{B}$ is specified as follows:

- ▷ Objects are monads $T = \langle C, t, \eta, \mu \rangle$, where $C \xrightarrow{t} C$ is an endo-1-cell of $\mathscr{B}$;
- ▷ lax 1-cells from $T = \langle C, t, \eta, \mu \rangle$ to $T' = \langle C', t', \eta', \mu' \rangle$ are pairs $\langle s, \delta \rangle$ with $C \xrightarrow{s} C'$ and $ts \overset{\delta}{\Longrightarrow} st'$ subject to two compatibility requirements concerning the monads' units and multiplications:





- ▷ 2-cells from $\langle s, \delta \rangle$ to $\langle q, \varphi \rangle$ are 2-cells $q \overset{\omega}{\Longrightarrow} s$ (note the reversal of the direction!) that are compatible with $\delta$ and $\varphi$:



**5.6.08 Remarks.**

(0) Lax 1-cells between monads can always be composed.

(1) Another candidate for 1-cells between monads are the modular ones *cf.*, Diagram 5.5-00. In order to compose these, $\mathscr{B}$ locally, *i.e.*, in the hom-categories, has to admit co-equalizers, and these have to be preserved by left and right composition with further 1-cells. The second condition is guaranteed in closed bi-categories, *cf.* Definition 5.10.02 below. We denote the resulting 2-category by $\mathscr{B}$-$\boldsymbol{mnd}_{\mathrm{m}}$ .

(2) Further interesting 2-categories arise by restricting the 1-cells to maps, *i.e.*, left adjoint 1-cells.

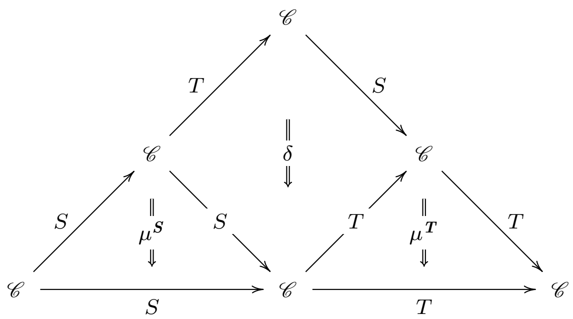
**5.6.09 Examples.** *Cf.*, Examples 5.6.06:

(0) $\boldsymbol{idl} = \boldsymbol{rel}$-$\boldsymbol{mnd}_{\mathrm{m}}$ has pre-ordered sets as objects, order-ideals as 1-cells and inclusions as 2-cells. The corresponding 2-category of maps coincides with $\boldsymbol{pre}$ , the 2-category of pre-ordered sets, order-preserving functions and point-wise comparisons of those.

(1) Since left-adjoint spans are functions as well, $\boldsymbol{cat}$ can be identified with the map-2-category of $\boldsymbol{spn}$-$\boldsymbol{mnd}_{\mathrm{m}}$ .

## 5.7 Distributive laws

Given two monads $\boldsymbol{S} = \langle S, \eta^{\boldsymbol{S}}, \mu^{\boldsymbol{S}} \rangle$ and $\boldsymbol{T} = \langle T, \eta^{\boldsymbol{T}}, \mu^{\boldsymbol{T}} \rangle$ on the same category $\mathscr{C}$ , the question arises, whether the composite functor $\mathscr{C} \xrightarrow{S} \mathscr{C} \xrightarrow{T} \mathscr{C}$ carries a monad structure as well. The problem is to define a meaningful multiplication $STST \Longrightarrow ST$. This can fail in general, but in the presence of a natural transformation $TS \stackrel{\delta}{\Longrightarrow} ST$ satisfying suitable axioms, the original multiplications $SS \stackrel{\mu^{\boldsymbol{S}}}{\Longrightarrow} S$ and $TT \stackrel{\mu^{\boldsymbol{T}}}{\Longrightarrow} T$ can be brought to bear, as indicated in the following diagram:



The axioms for $\delta$ are chosen such that the composite 2-cell above yields a monad together with the obvious unit $id_{\mathscr{C}} = id_{\mathscr{C}} id_{\mathscr{C}} \stackrel{\eta^{\boldsymbol{S}}\eta^{\boldsymbol{T}}}{\Longrightarrow} ST$, namely:

$$\begin{array}{ccccc} TSS & \stackrel{\delta S}{\Longrightarrow} & STS & \stackrel{S\delta}{\Longrightarrow} & SST \\ {\scriptstyle T\mu^{\boldsymbol{S}}}\Downarrow & & & & \Downarrow{\scriptstyle \mu^{\boldsymbol{S}}T} \\ TS & & \stackrel{\delta}{\Longrightarrow} & & ST \end{array} \quad \text{and} \quad \begin{array}{ccccc} TTS & \stackrel{T\delta}{\Longrightarrow} & TST & \stackrel{\delta T}{\Longrightarrow} & STT \\ {\scriptstyle \mu^{\boldsymbol{T}}S}\Downarrow & & & & \Downarrow{\scriptstyle S\mu^{\boldsymbol{T}}} \\ TS & & \stackrel{\delta}{\Longrightarrow} & & ST \end{array}$$

as well as

$$\begin{array}{ccc} & T & \\ {\scriptstyle T\eta^{\boldsymbol{S}}}\swarrow & & \searrow{\scriptstyle \eta^{\boldsymbol{S}}T} \\ TS & \xRightarrow[\delta]{} & ST \end{array} \qquad \text{and} \qquad \begin{array}{ccc} & S & \\ {\scriptstyle \eta^{\boldsymbol{T}}S}\swarrow & & \searrow{\scriptstyle S\eta^{\boldsymbol{T}}} \\ TS & \xRightarrow[\delta]{} & ST \end{array}$$

the reader is encouraged to check that the unit and multiplication above indeed provide a monad structure on $ST$ .

**5.7.00 Proposition.** *For a bi-category $\mathscr{B}$ , there is a bijective correspondence between*

- *distributive laws $TS \xRightarrow{\delta} ST$ between monads $\boldsymbol{T} = \langle T, \eta^{\boldsymbol{T}}, \mu^{\boldsymbol{T}} \rangle$ and $\boldsymbol{S} = \langle S, \eta^{\boldsymbol{S}}, \mu^{\boldsymbol{S}} \rangle$ on the same $\mathscr{B}$-object $C$ ;*
- *monads in $\mathscr{B}$-$\boldsymbol{mnd}_\ell$ .*

*In case of $\mathscr{B} = \boldsymbol{Cat}$ these bijectively correspond to*

- *liftings of the monad $\boldsymbol{T}$ on $\mathscr{C}$ to a monad $\boldsymbol{T_S}$ on the category $\mathscr{C}^{\boldsymbol{S}}$ of EM-algebras*

**5.7.01 Examples.**

(0) As seen in the first sections, the list monad is connected with the finite and the infinite power-set monad by rather simple distributive laws. These are easily seen to be unaffected by the presence or absence of the empty word/subset.

(1) These distributive laws carry over to the finite and infinite power-bag monads, where they take the form of distributivity of multiplication over addition.

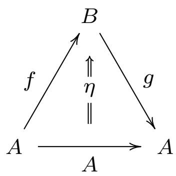
**5.7.02 Conjecture.** In order to obtain the category of commutative semi-rings as a category of EM-algebras for a monad, one would expect to find a non-trivial distributive law on the finite bag monad.

## 5.8 Adjunctions

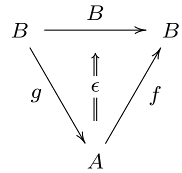
One of the fundamental notions of category theory is the notion of adjunction, due to Daniel Kan in 1958. It may be seen as a weakening of the notion of inverse functions in $\boldsymbol{set}$ . The only sensible way that a function $X \xrightarrow{f} Y$ can have an inverse in $\boldsymbol{set}$ is to have a function $Y \xrightarrow{g} X$ such that $f \cdot g = id_X$ and $g \cdot f = id_Y$ . In a 2-category, where the hom-sets are themselves categories, these equality requirements may be weakened to the existence of 2-cells mediating between $f \cdot g$ and $id_X$ , respectively, $g \cdot f$ and $id_Y$ . Also recall the notion of a generalized inverse matrix $M^{\mathrm{g}}$ in linear algebra that satisfies $MM^{\mathrm{g}}M = M$ .

We start with the abstract concept available in any 2-category, which takes a diagrammatic form that is easy to memorize (we hope).

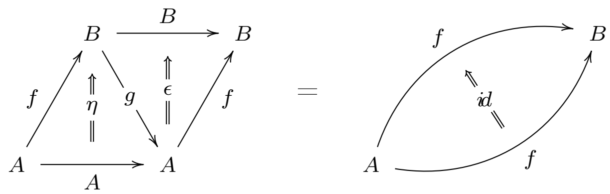
**5.8.00 Definition.** Two 1-cells $A \xrightarrow{f} B$ and $B \xrightarrow{g} A$ are called *adjoint*, if there exist 2-cells



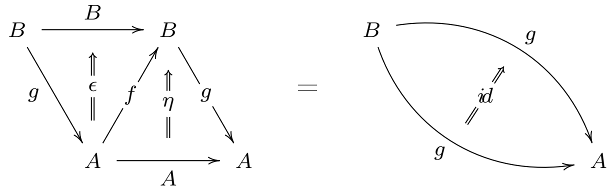
and



subject to the following axioms:



(5.8-00)



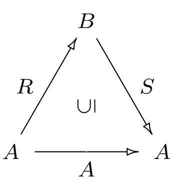
(5.8-01)

Provided both $\eta$ and $\varepsilon$ are isomorphisms, $f$ and $g$ establish an *equivalence* between $A$ and $B$, which is an appropriate weakening of the notion of isomorphism in the context of 2-categories.

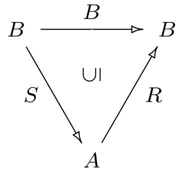
Notation: $f \dashv g$; $f$ is called *left adjoint* and $g$ *right adjoint*, while $\eta$ and $\epsilon$ are called the *unit*, respectively *co-unit* of the adjunction.

**5.8.01 Examples.**

(0) Since the 2-category ***rel*** of sets, (binary) relations and inclusions is is locally partially ordered, the very existence of unit and co-unit inclusions characterizes adjunctions; the compositions then automatically yield the required equalities: hence relations $A \xrightarrow{R} B$ and $B \xrightarrow{S} A$ are adjoint, if

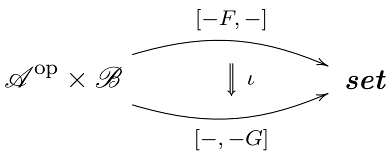


and



The first inclusion says that $R$ is *total*, while the second inclusion forces $\mathcal{R}$ to be *single-valued*. This characterizes the left adjoint relations as precisely the functions, while the right adjoint relations are precisely the duals of functions, *i.e.*, $S = R^{\mathrm{op}}$ above.

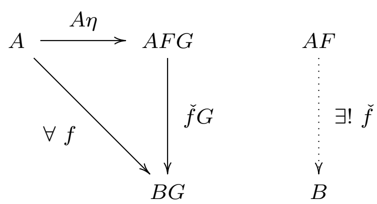
(1) Adjunctions in $\boldsymbol{Cat}$ can also be described in somewhat different terms: given functors $\mathscr{C} \xrightarrow{F} \mathscr{D}$ and $\mathscr{D} \xrightarrow{G} \mathscr{C}$, we have $F \dashv G$ provided there is a family of bijections between the hom-sets $\langle A, BG\rangle\mathscr{C}$ and $\langle AF, B\rangle\mathscr{D}$ that is natural in $A \in \mathscr{C}$ and $B \in \mathscr{D}$. In other words, there is a *natural isomorphism*




(The position of $F$ and $G$ in these hom-sets/functors indicates which of them is left-, resp. right-adjoint.)

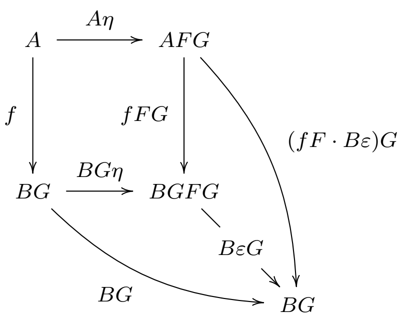
The images of the identities on $AF$ in $\boldsymbol{D}$, resp. the pre-images of the identities on $BG$ in $\mathscr{C}$ yield the components $A \xrightarrow{A\eta} AFG$ of the unit and $BGF \xrightarrow{B\epsilon} B$ of the co-unit.

Conversely, given $\mathscr{C} \xrightarrow{\eta} FG$ and $A \in \mathscr{C}$ as well as $B \in \mathscr{D}$, we wish to show that every $\mathscr{C}$-morphism $A \xrightarrow{f} BG$ admits a unique $\mathscr{D}$-morphism $AF \xrightarrow{\tilde{f}} B$ such that $\tilde{f}G$ extends $f$ along $A\eta$:




(5.8-02)

An obvious candidate for $\check{f}$ is $fF \cdot B\epsilon$:

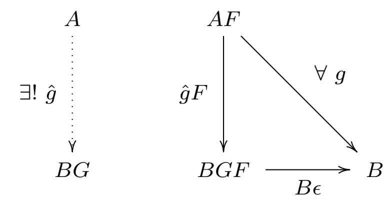



(5.8-03)

The square commutes by naturality of $\eta$, the right triangle by functoriality of $G$, and the lower triangle by one of the adjointness conditions.

Alternatively, starting with the co-unit $GF \xrightarrow{\epsilon} \mathscr{D}$ and objects $A$ and $B$ as above, every $\mathscr{D}$-morphism $AF \xrightarrow{g} B$ is supposed to induce a unique $\mathscr{C}$-morphism $A \xrightarrow{\hat{g}} BG$ such that

$\hat{g}F$ lifts $g$ along $B\epsilon$ :




(5.8-04)

A dual argument to the one above shows that $A\eta \cdot gG$ is a candidate for $\hat{g}$ .

Clearly, the assignments $f \mapsto fF \cdot B\varepsilon$ and $g \mapsto A\eta \cdot gG$ are mutual inverses, hence bijective. Their naturality follows easily from the naturality of $\eta$ and of $\varepsilon$ .

(2) Every monad $\boldsymbol{T} = \langle T, \eta^{\boldsymbol{T}}, \mu^{\boldsymbol{T}} \rangle$ on $\mathscr{C}$ induces an adjunction between the free functor $\mathscr{C} \xrightarrow{F^{\boldsymbol{T}}} \mathscr{C}^{\boldsymbol{T}}$ into the category of EM-algebras, and the forgetful functor $\mathscr{C}^{\boldsymbol{T}} \xrightarrow{U^{\boldsymbol{T}}} \mathscr{C}$.

A particular instance occurs for the lifting of the exception monad to the category ***sgr*** of semi-groups, whose EM-category is ***mon*** . This allows an obvious generalization to the many-object-case: the counterpart of the exception monad is defined on the category ***grph*** of graphs and adds a distinguished loop to every node; this has to be preserved by the morphisms in the EM-category. On the other hand, extending a graph by all non-empty paths leads to a semi-category, and there is an obvious distributive law linking these two monads and generalizing $\delta_1$ of Lemma 1.1.02(1). (Of course, $\delta_0$ can also be generalized in a similar fashion.)

Hence ***cat*** , or ***Cat*** for that manner, arises as the category of EM-algebras of the generalized exception monad lifted to the category ***scat*** (or ***sCat*** ) of semi-categories. Hence the functor ***Cat*** $\xrightarrow{U}$ ***sCat*** that undistinguishes the identities has a left adjoint that freely adds identities to the endo-hom-sets. ◁

The functor $U$ of the last example also has a right adjoint that appears to be useful in the context of algebraic automata theory.

**5.8.02 Definition.** An endomorphism $A \xrightarrow{i} A$ of a semi-category $\mathscr{D}$ is called *idempotent*, if $i \cdot i = i$ . An idempotent $i$ is said to *split*, if there exists $A \xrightarrow{e} B$ and $B \xrightarrow{m} A$ with $e \cdot m = i$ and $m \cdot e = id_B$ .

Having all idempotents split turns out to be a useful property, but it is not always satisfied.

**5.8.03 Definition.** The *Karoubi envelope* $\tilde{\mathscr{D}}$ (also known as *Cauchy-completion* in the enriched context) of a semi-category $\mathscr{D}$ is defined as the following non-full sub-semi-category of the arrow-category $\mathscr{D}^{\to}$ :

▷ objects are the idempotent 1-cells of $\mathscr{D}$ ;

▷ morphisms from $i$ to $j$ are $\mathscr{D}$-morphisms $i\partial_0 \xrightarrow{f} j\partial_0$ such that $i \cdot f = f = f \cdot j$, or equivalently, $i \cdot f \cdot j = f$.

Clearly, $\tilde{\mathscr{D}}$ is always a category, since an idempotent $i$ of $\mathscr{D}$ is the identity on the object $i$ of $\tilde{\mathscr{D}}$. Since idempotents are preserved by any functor, it is also easy to see that for any category $\mathscr{C}$ the semi-functors $\mathscr{C}U \longrightarrow \boldsymbol{D}$ bijectively correspond to the functors $\mathscr{C} \longrightarrow \tilde{\mathscr{D}}$. [What happens at the level of 2-cells?]

## 5.9 Special morphisms

As the prototypical category is ***set***, with functions as morphisms, it is no surprise that besides the notion of bijection also injections and surjections have been generalized to arbitrary categories. As it turns out, the aspects of cancellability and invertibility lead to different generalizations.

**5.9.00 Definition.** A morphism $B \xrightarrow{g} C$ in a category $\mathscr{C}$ is called a

▷ *monomorphism*, or *mono* for sort, if the function

$$[A, B] \xrightarrow{[A,g]} [A, C]$$

defined by post-composition with $g$ is injective for every $\mathscr{C}$-object $A$. In other words, $g$ is *post-cancellable* in the sense that $r \cdot g = s \cdot g$ implies $r = s$ for any $\mathscr{C}$-morphisms $A \underset{s}{\overset{r}{\rightrightarrows}} B$;

▷ *section*, also known as *split mono*, if $g$ has a *right inverse* $C \xrightarrow{h} B$ satisfying $g \cdot h = id_B$.

Dual notions: *epimorphism*, *pre-cancellable*; *retraction* or *split epi*.

It is easy to see that for an iso $B \xrightarrow{g} C$ a left inverse $C \xrightarrow{f} B$ and a right inverse $C \xrightarrow{h} B$ have to agree. Moreover, any isomorphism is of course both epi and mono, as the pre- and post-composition functions are bijective. However, the converse need not be true:

**5.9.01 Examples.**

(a) The inclusion of the natural numbers $\mathbb{N}$ into the integers $\mathbb{Z}$ is an injective epi, but not an iso in the category ***mon*** of monoids. Notice that $\langle \mathbb{N}, +, 0 \rangle$ is a free monoid, while $\langle \mathbb{Z}, +, 0 \rangle$ is a free group on a singleton set. Consider different monoid-homomorphisms $\mathbb{Z} \underset{g}{\overset{f}{\rightrightarrows}} M$. The images of $\mathbb{Z}$ under $f$ and $g$ automatically are groups, and hence $f$ and $g$ are uniquely determined by the values of $1 \cdot f$ and $1 \cdot g$, respectively, which by hypothesis are different elements of $M$. But then also $1 \cdot i \cdot f \neq 1 \cdot i \cdot g$, which implies $i \cdot f \neq i \cdot g$.

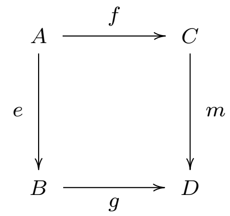
(b) The inclusion of the integers $\mathbb{Z}$ into the rationals $\mathbb{Q}$ is an injective epi, but not an iso in the category ***ring*** of rings. ◁

Since morphisms that are both mono and epi can fail to be iso, one can ask for the least strengthening of the notions of mono and epi that will result in isos in the presence of the unaltered other property:

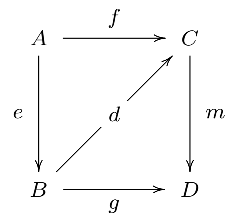
**5.9.02 Definition.** An epimorphism $A \xrightarrow{e} B$ is called

(0) *extremal epi*, if $e$ does not factor through a proper mono, *i.e.*, whenever $e = f \cdot m$ with $m$ mono, then $m$ is already iso;

(1) *strong epi*, if for every diagram of the form




with $m$ mono there exists a unique diagonal $B \xrightarrow{d} C$ making both triangles commute:




(5.9-00)

Dual Notions: *extremal mono*; *strong mono*.

Diagram 5.9-00 is a first instance of a so-called “diagonalization relation” that will be useful in various other contexts, see Subsection 5.14 below.

**5.9.03 Proposition.**

(0) *Every split epi is strong.*

(1) *A morphism is iso iff it is mono and strong epi.*

(2) *Every strong epi is extremal.*

**Proof.**

(0) Suppose $e$ in Diagram (5.9-00) is split with left-inverse $B \xrightarrow{h} A$. We claim that $d := h \cdot f$ is the desired diagonal. The lower triangle commutes because of

$$h \cdot f \cdot m = h \cdot e \cdot g = g$$

while the upper triangle commutes since $m$ is mono and

$$f \cdot m = e \cdot g = e \cdot h \cdot f \cdot m = e \cdot d \cdot m$$

(1) An iso is both split mono and split epi, hence mono and split epi, hence by (0) mono and strong epi. Conversely, set $e = m$ in Diagram (5.9-00), and use identities for the horizontal morphisms. Then $d$ is left- and right-inverse to $e$.

(2) Suppose in Diagram (5.9-00) $D = B$ and $g = id_B$. Then the diagonal $d$ is a left-inverse for $m$. Therefore $m$ is mono and split epi, thus by (0) mono and strong epi, and by (1) iso. □

**5.9.04 Example.** In ***set*** every surjective function is both an epi and a retraction. On the other hand, while all injective functions are monos, only those with non-empty domain are split mono: the inclusion of $\emptyset$ into an nonempty set cannot have a right-inverse. ◁

**5.9.05 Proposition.** *If* $B \xrightarrow{g} C \xrightarrow{h} D$ *is mono, so is* $g$.

**Proof.** If $A \overset{r}{\underset{s}{\rightrightarrows}} B$ are distinct, so are $r \cdot g \cdot h$ and $s \cdot g \cdot h$, which forces $r \cdot g \neq s \cdot g$. □

In most categories of structured sets the monos turn out to have underlying functions that are injective, however the underlying functions of epis need not be surjective, as Example 5.9.01 shows. Furthermore, isos in such categories have to be bijective, however bijective homomorphisms may not be isos: just consider the identity function from a discretely ordered set into an indiscretely ordered one; it only preserves order in one direction.

**5.9.06 Definition.** Monos into an object $C$ are called *sub-objects* of $C$. If every object up to isomorphism has only a set of sub-objects, $\mathscr{C}$ is called *well-powered*.

Dual notion: *super-object*, *cowell-powered*.

**5.9.07 Remarks.**

(a) Some authors reserve the term "sub-object" for equivalence classes of monos into $C$, where two monos into $C$ are equivalent, if they differ by an isomorphism in the domain. This point of view is important when size questions arise, *e.g.*, whether all objects have a set of sub-objects, or if proper classes of sub-objects can occur. To avoid speaking explicitly about equivalence classes, phrases like “up to isomorphism” or “essentially” are being employed.

(b) The terminology of the dual of a sub-objects is not uniform in the literature. While some authors use the term “quotient” for this purpose, otherwise reserve this term for a more specialized notion in connection with a categorical version of equivalence relation (called “congruence”, see below). In that case I’ve seen the term “co-sub-object” being employed, which strikes me as rather awkward. So for these notes we will use “super-object” instead.

(c) For many applications the notions of monomorphism and of epimorphism are too weak, while the notions of split mono/epi are too strong. Hence one finds a number of intermediate notions that are of importance (eg, "strong", "regular", "extremal", "effective",...). In Section 5.13, we will need to consider regular epis.

**5.9.08 Examples.**

▷ In ***set*** sub-objects may be identified with subsets. In fact, every function $A \xrightarrow{f} C$ factors through a smallest subset of $C$, namely the image $f[A] \subseteq C$.

On the other hand, super-objects or quotients in ***set*** correspond to set-indexed families of sets: if $B \xrightarrow{g} C$ is surjective, we have a $C$-indexed family of sets, namely the pre-images of the elements of $C$.

▷ In ***Cat*** the notion of sub-category needs further qualification. The *full subcategory* generated by a sub-set (or sub-class) of $\mathcal{A} \subseteq \mathscr{C}$-***Ob***, has the elements of $\mathcal{A}$ as objects and the corresponding hom-sets of $\mathscr{C}$ as hom-sets. *E.g.*, the category ***ab*** of abelian groups is a full subcategory of the category ***grp*** of groups. However, the category ***sgr*** of semi-groups yields an example of a *non-full subcategory*, as every monoid is a semi-group, but not every semi-group-homomorphism between monoids is a monoid homomorphism; it may fail to preserve the neutral element.

Another example is given by $\bigsqcup$-semi-lattices, which are, of course, $\sqcup$-semi-lattices, but not all $\sqcup$-***slat***-morphisms preserve arbitrary suprema. Moreover, $\bigsqcup$-***slat*** is a non-full sub-category of ***clat***, the category of complete lattices: the objects coincide, but in the first case morphisms are only required to preserve suprema, while in the second case suprema and infima are to be preserved.

The notions of monomorphism and epimorphism admit generalizations to families of morphisms with common domain, resp., codomain. Their relevance will become clear in Section 5.11.

**5.9.09 Definition.** A family of morphism with common domain is called a *source*. If for any different parallel morphisms into the source's domain the composite sources differ, we have a *mono-source*.

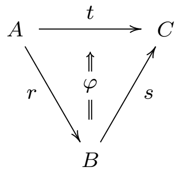
DUAL NOTION: *sink*, *epi-sink*.

## 5.10 Extensions and liftings

We have seen that many important notions of category theory, like monads and adjunctions, can in fact be formalized in any 2-category. Another such concept, *Kan-extensions* (due to Daniel M. Kan, 1960) can be abstracted in a similar fashion and provides a very economical, if slightly abstract approach, that subsumes many other notions. MacLane [Mac71] has a section entitled "All Concepts Are Kan Extensions", where he claims

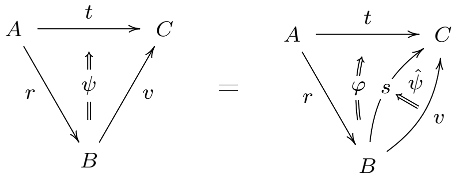
The notion of Kan extensions subsumes all the other fundamental concepts of category theory.

Here we follow the introduction to Street and Walters [SW78]. Consider a 2-cell in a 2-category $\mathcal{B}$ .



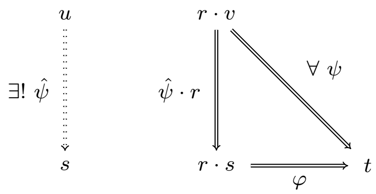
(5.10-00)

**5.10.00 Definition.** Diagram (5.10-00) exhibits the pair $\langle s, \varphi \rangle$ as a *right extension* of $A \xrightarrow{t} C$ *along* $A \xrightarrow{r} B$, if for any 1-cell $B \xrightarrow{x} C$, pasting, *i.e.*. 2-cell composition, with $\varphi$ at $s$ is a bijection between the hom-sets $\langle r \cdot x, t \rangle[A, C]$ and $\langle x, s \rangle[B, C]$ . More precisely, any 2-cell $r \cdot u \xRightarrow{\psi} t$ uniquely factors through $\varphi$



(5.10-01)

Alternatively, this universal property may be depicted as



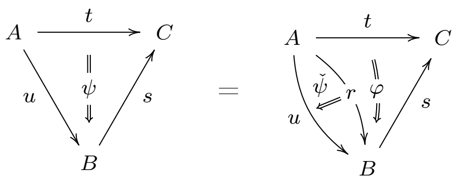
(5.10-02)

with a unique 2-cell $x \xRightarrow{\hat{\psi}} s$.

The right extension $\langle s, \varphi \rangle$ of $t$ along $r$ is called *absolute*, if for every 1-cell $C \xrightarrow{w} D$ the pair $\langle s \cdot w, \varphi \cdot w \rangle$ is a right-extension of $t \cdot w$ along $r$ .

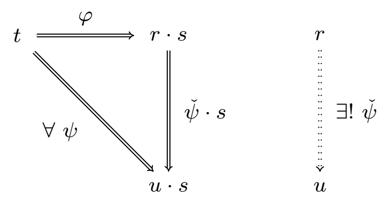
Dual notion: Diagram (5.10-00) exhibits the pair $\langle r, \varphi \rangle$ as a *right lifting* of $A \xrightarrow{t} C$ *through* $B \xrightarrow{s} C$; *absolute right lifting*.

Opposite notion: If the 2-cell $\varphi$ in Diagram (5.10-00) is reversed, it can exhibit the pair $\langle s, \varphi \rangle$ as a *left extension* of $A \xrightarrow{t} C$ *along* $A \xrightarrow{r} B$, resp., the pair $\langle r, \varphi \rangle$ as a *left lifting* of $A \xrightarrow{t} C$ *through* $B \xrightarrow{s} C$. The diagrams for left-liftings look like



(5.10-03)

or, alternatively




(5.10-04)

*Absolute left extensions/liftings* are preserved by post/pre-composition with 1-cells.

**5.10.01 Remarks.**

(0) One can think of the right extensions and liftings as the "best" approximation to a commutative triangle "from below", depending on whether the given two 1-cells have a common domain, or a common codomain. Similarly, left extensions and liftings are "best" approximations "from above". This intuition will be supported by examples below.

(1) In case that $A \xrightarrow{r} B$ exhibits another pair $\langle s', \varphi' \rangle$ as right extension of $t$ along $r$, then $s$ and $s'$ will be isomorphic as objects of the hom-category $[B, C]$: the definition induces mutually inverse 2-cells linking $s$ and $s'$.

(2) Consider $A \xrightarrow{r} B$ and $C$ as fixed, while $A \xrightarrow{t} C$ as well as $B \xrightarrow{u} C$ can vary. Comparing Diagram (5.10-02) with Diagram (5.8-04) of Example 5.8.01(1) shows that the existence of right extensions $\langle s, \varphi \rangle$ along $A \xrightarrow{r} B$ for all 1-cells $A \xrightarrow{t} C$ can be expressed equivalently by saying that the functor

$$[B, C] \xrightarrow{[r, C]} [A, C]$$

that operates by pre-composing with $r$ is left adjoint. The corresponding right adjoint maps $t$ to $s$, the right-extension. Let us denote this functor by

$$[A, C] \xrightarrow{r \triangleright -} [B, C]$$

and refer to it as *pre-residuation* with respect to $r$.

(3) The existence of all right liftings $\langle r, \varphi \rangle$ through $B \xrightarrow{s} C$ for all 1-cells $A \xrightarrow{t} C$ amounts to the functor

$$[A, B] \xrightarrow{[A, s]} [A, C]$$

that operates by post-composition with $s$ being left adjoint. The corresponding right-adjoint

$$[A, C] \xrightarrow{- \triangleleft} [A, B]$$

will be called *post-residuation* with respect to $s$, and can also be written as $-/s$.

(4) the existence of all left-extensions along $r$, resp., left-liftings through $s$ means that the functors $[r, C]$ and $[A, s]$ are right-adjoint, *i.e.*, have left-adjoints. (I'm not aware of terminology corresponding to "residuation" in this case.)

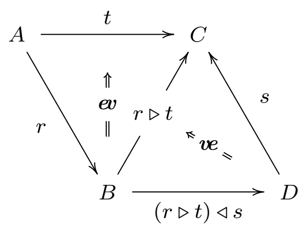
**5.10.02 Definition.** The 2-category $\mathcal{B}$ is called *pre-closed/post-closed*, if pre/post-composition with every 1-cell is left adjoint, *i.e.*, all pre/post-residuations exist. $\mathcal{B}$ is called *closed*, if it is pre- and post-closed.

*Dual Notion*: (*pre-,/post-*) *op-closed*.

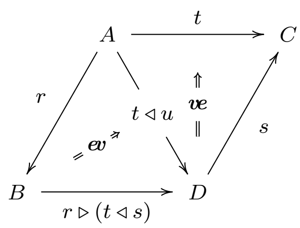
In the special case of a single 0-cell 2-categories can be identified with monoidal categories, see Remark 5.5.01. If such a 2-category is (pre- or post-) closed, one refers to the monoidal category as *monoidal closed* from the appropriate side.

**5.10.03 Lemma.** *For 1-cells* $B \xleftarrow{r} A \xrightarrow{t} C \xleftarrow{s} D$, *if the appropriate right extensions / liftings exist, we have* $(r \triangleright t) \triangleleft s \cong r \triangleright (t \triangleleft s)$ . *Hence we can write* $r \triangleright t \triangleleft s$ .

**Proof.** Just observe that the 2-cells




and

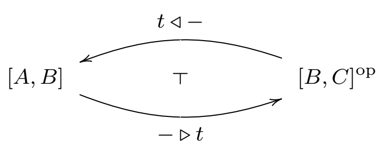



mutually factor though each other. $\square$

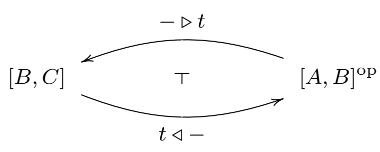
**5.10.04 Remark.** In a closed bicategory $\mathcal{B}$ the bijective correspondences

$$\begin{array}{r}
r \Longrightarrow t \triangleleft s \\ \hline\hline
r \cdot s \Longrightarrow t \\ \hline\hline
s \Longrightarrow r \triangleright t
\end{array}$$

automatically induces further adjunctions




as well as




(5.10-05)

that are known as *polarities*.

**5.10.05 Example.** ***rel*** is closed: given a relation $A \xrightarrow{R} B$, for $A \xrightarrow{T} C$ and $B \xrightarrow{S} C$ the pre- and post-residuations are given by

$$\begin{array}{lll}
\langle b, c\rangle \in R \triangleright T & \text{iff} & \forall a \in A.\ \langle a, b\rangle \in R \Rightarrow \langle a, c\rangle \in T \\
\langle a, b\rangle \in T \triangleleft S & \text{iff} & \forall c \in C.\ \langle a, c\rangle \in T \Leftarrow \langle b, c\rangle \in S
\end{array}$$

In other words, we have adjunctions

$$[B,C] \underset{[R,C]}{\overset{R \rhd -}{\underset{\top}{\leftrightarrows}}} [A,C] \quad \text{as well as} \quad [A,B] \underset{[A,S]}{\overset{- \lhd S}{\underset{\top}{\leftrightarrows}}} [A,C] \tag{5.10-06}$$

and polarities

$$[A,B] \underset{- \rhd T}{\overset{T \lhd -}{\underset{\top}{\leftrightarrows}}} [B,C]^{\mathrm{op}} \quad \text{as well as} \quad [B,C] \underset{T \lhd -}{\overset{- \rhd T}{\underset{\top}{\leftrightarrows}}} [A,B]^{\mathrm{op}} \tag{5.10-07}$$

Interesting things happen when for, say, $B \xrightarrow{S} C$ we chose $A = 1$, as then $[1,B]$ and $[1,C]$ essentially are the power-sets of $B$ and $C$, respectively. Besides the adjunction on the right of (5.10-06), by considering the dual relation $C \xrightarrow{S^{\mathrm{op}}} B$ we obtain a second adjunction in the opposite direction:

$$[1,B] \underset{[1,S]}{\overset{- \lhd S}{\underset{\top}{\leftrightarrows}}} [1,C] \quad \text{as well as} \quad [1,C] \underset{[1,S^{\mathrm{op}}]}{\overset{- \lhd S^{\mathrm{op}}}{\underset{\top}{\leftrightarrows}}} [1,B] \tag{5.10-08}$$

If, moreover, $S$ is left adjoint, *i.e.*, a function $B \xrightarrow{g} C$, then $g \dashv g^{\mathrm{op}}$ and one might suspect that this implies $[1,g] \dashv [1,g^{\mathrm{op}}]$. Indeed, this is the case, hence the two adjunctions above can be combined into

$$[1,B] \;\overset{- \lhd g^{\mathrm{op}}}{\underset{[1,g]}{\overset{\perp}{\underset{\perp}{\xleftarrow{[1,g^{\mathrm{op}}] = - \lhd g}}}}}\; [1,C] \;\cong\; BP \;\overset{g_\forall}{\underset{g_\exists}{\overset{\perp}{\underset{\perp}{\xleftarrow{g^{\leftarrow}}}}}}\; CP \;\cong\; [B,1] \;\overset{g \rhd -}{\underset{[g^{\mathrm{op}},1]}{\overset{\perp}{\underset{\perp}{\xleftarrow{[g,1] = g^{\mathrm{op}} \rhd -}}}}}\; [C,1] \tag{5.10-09}$$

In the middle $g^{-1}$ denotes the inverse image function, which turns out to be left and right adjoint. The other two functions are given by

$$\begin{aligned} V s_\exists &:= \{\, c \in C \,:\, \exists b \in B.\, b \in V \wedge bs = c \,\} = \{\, c \in C \,:\, V \cap cs^{-1} \neq \emptyset \,\} \\ V s_\forall &:= \{\, c \in C \,:\, \forall b \in B.\, bs = c \Rightarrow b \in V \,\} = \{\, c \in C \,:\, cs^{-1} \subseteq V \,\} \end{aligned} \tag{5.10-10}$$

The set $V s_\exists$ is usually referred to as the *direct image* of $V$ under $s$, or the $s$-*image* of $V$.

On the right we are using the dualization $(-)^{\mathrm{op}}$ on ***rel*** that maps $1 \xrightarrow{V} B$ to $B \xrightarrow{V^{\mathrm{op}}} 1$. Hence $1 \xrightarrow{V} B \xrightarrow{g} C$ turns into $C \xrightarrow{s^{\mathrm{op}}} B \xrightarrow{V^{\mathrm{op}}} 1$.

Hence the self-duality of ***rel*** is responsible for the existence of many different descriptions of essentially the same chain of adjunctions.

[HW] What happens, if in the polarities above the set $B$ is chosen to be $1$? ◁

**5.10.06 Examples.**

(0) Notice that ***set*** as a locally discrete 2-category sitting inside ***rel*** is not closed. While the residuations of two functions always exist, they may be proper relations. Even commutative triangles of functions need not correspond to pre- or post-residuations, as the uniqueness requirement could be violated.

(1) Although ***set*** as a locally discrete category is not closed, the *suspension* ***set***-$\Sigma$ of ***set*** is closed: this 2-category has a single object $*$, sets as 1-cells and functions as 2-cells. In particular, ***set***-$\Sigma$ is not locally small. The composition of 1-cells is given by cartesian product $\times$, with neutral 1-cell $1$. Usually the suspension-view is suppressed and ***set*** is directly called *cartesian closed*. Due to the symmetry of $\times$ the pre- and post-residuations agree with the function-space construction.

**Claim.** For every set $X$, the functor ***set*** $\xrightarrow{X \times -}$ ***set*** is left-adjoint, with right adjoint the function-space functor ***set*** $\xrightarrow{[X,-]}$ ***set***.

The $Y$-component of the unit ***set*** $\xRightarrow{\eta}$ $[X, X\times -]$ maps $y \in Y$ to the function $X \longrightarrow X \times Y$ with graph $X \times \{y\}$. On the other hand, the $y$-component of the co-unit $X \times [X, -] \xrightarrow{\epsilon}$ ***set*** is the *evaluation* $X \times [X, Y] \xrightarrow{\boldsymbol{ev}_X} Y$ that maps $\langle x, f\rangle$ to $xf$.

(2) The category ***prt*** of sets and partial functions can be viewed in various different ways: either as Eilenberg-Moore category $\boldsymbol{set}^{\boldsymbol{E}}$ for the exception monad, or as comma category $1 \downarrow$ ***set*** of *pointed sets*, whose objects are sets with distinguished base point and whose morphisms are base-point preserving functions.

**Claim.** ***prt*** is closed with respect to the *smash product* $\wedge$ that first takes the cartesian product and then identifies all pairs with a base point in at least one component.

For a set $X$ we write $X_*$ for the one-point extension $X + \{*\}$, where $*$ is the base point. As the smash product is symmetric, there is no need to distinguish between a left and a right "function space". Define

$$\lfloor X_*, Y_* \rfloor$$

to be the set of partial functions $X \longrightarrow Y$. Then we obtain an "evaluation"

$$X_* \wedge \lfloor X_*, Y_* \rfloor_* \xrightarrow{\boldsymbol{ev}} Y_*$$

specified by

$$\langle x, f\rangle \mapsto xf \quad , \quad \langle *, f\rangle \mapsto * \quad , \quad \langle x, *\rangle \mapsto * \quad \text{and} \quad \langle *, *\rangle \mapsto *$$

whenever $x \in X$ and $f \in \lfloor X_*, Y_* \rfloor$. Except for the case distinctions, the proof of closedness proceeds along similar lines as the proof that $\langle \boldsymbol{set}, \times, 1\rangle$ is cartesian closed.

(3) Much of "categorical topology" was motivated by the fact that ***top*** , the category of topological spaces and continuous functions fails to be cartesian closed, *i.e.*, fails to have good function spaces. Various related categories that do have this property have been constructed.

(4) Algebraic categories like ***grp*** , ***ab*** , or ***ring*** usually are not closed with respect to the cartesian product. Sometimes, there is another product, usually called *tensor product* $\otimes$ with unit $I$ , for which closedness can be established, *e.g.*, for ***ab*** .

(5) As the comparison between ***set*** and ***rel*** shows, enlarging the hom-sets can recover closedness. While rings with ring homomorphisms do not form a closed category, rings with modules do. In general, (small) matrix-categories over a rig (or semi-ring) tend to be closed.

Also ***cat*** (and ***Cat*** ) with functors as 1-cells fails to be closed, but the use of so-called pro-functors $\mathscr{A} \longrightarrow \mathscr{B}$, *i.e.*, functors $\mathscr{A}^{\mathrm{op}} \times \mathscr{B} \longrightarrow$ ***set*** as morphisms instead of functors does yield a closed 2-category. (Notice that every functor $\mathscr{A} \longrightarrow \mathscr{B}$ induces a pair of adjoint profunctors between the same categories, but not all left adjoint profunctors are functors.)

(6) The operation of cartesian product is also available for enriched categories, hence one can consider the notion of cartesian closedness in this setting as well. The category ***ord*** of pre-ordered sets and order-preserving functions is cartesian closed, just as is ***cat*** .

(7) ***2Cat*** , with 2-categories as objects and 2-functors that preserve the composition of 1-cells and both compositions of 2-cells, viewed just as a category (disregarding any the higher-dimensional structure) is cartesian closed. But ***2Cat*** is also closed with respect to the *shuffle-product*, also known as the *funny tensor*. This is the only other monoidal closed structure on ***2Cat*** besides $\times$ and has the same unit $\mathbf{1}$ .

**5.10.07 Example.** As seen in Example 5.4.01, ordered monoids are monoidal categories where the hom-functor factors through $\mathbf{2}$ . An ordered monoid $\langle M, \sqsubseteq \rangle$ that is a quantale, hence in particular a complete lattice, automatically is monoidal closed, *i.e.*, its suspension is closed as a one object bicategory: given $r, t \in M$ , the element

$$r \backslash t := \bigsqcup \{ s \in M \ : \ r \cdot s \sqsubseteq t \}$$

is the largest element $s \in M$ with $r \cdot s \sqsubseteq t$ . Hence the properties

$$s \sqsubseteq r \backslash t \quad \text{and} \quad r \cdot s \sqsubseteq t \quad \text{and} \quad r \sqsubseteq t/s$$

are all equivalent. In particular, the residuation operations $J \backslash -$ and $-/K$ are available in $MP$ and preserve infima for every monoid $M$ and all subsets $J, K \subseteq M$ . On the other hand, the multiplication preserves suprema. Therefore the empty supremum $\perp$ , is an absorber, which in case $|\Omega| > 1$ differs from the neutral element $e$ .

Note, however, that a quantale need not be op-closed: although for fixed $r, t \in M$ the meet $p := \bigsqcap \{ s \in M \ : \ t \sqsubseteq r \cdot t \}$ does exist, we can only conclude that both $t$ and $r \cdot p$ lie below

$\bigsqcap\{ r \cdot s \ : \ t \sqsubseteq r \cdot s \}$ (*cf.* Example 1.2.01), which does not determine the relationship between $t$ and $r \cdot p$. [A concrete example for $t \not\sqsubseteq r \cdot p$ would be nice.] $\lhd$

**5.10.08 Theorem.** *For 1-cells $A \xrightarrow{f} B \xrightarrow{g} A$ and a 2-cell $g \cdot f \xRightarrow{\epsilon} B$ the following are equivalent:*

(a) *$\langle f, \epsilon \rangle$ is an absolute right extension of $B$ along $g$;*

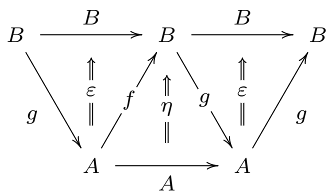
(b) *$\langle f, \epsilon \rangle$ is a right extension of $B$ along $g$ that is preserved by $g$;*

(c) *$\epsilon$ is the co-unit of an adjunction $f \dashv g$.*

**Proof.**

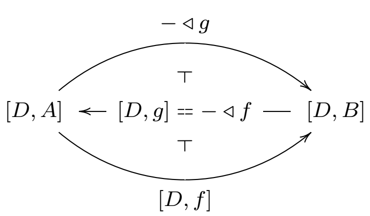
**(a) ⇒ (b):** Trivial.

**(b) ⇒ (c):** The right extension $\langle f \cdot g, \varepsilon \cdot g \rangle$ allows to define a candidate $A \xRightarrow{\eta} f \cdot g$ for the unit that satisfies the second adjointness condition of (5.8-01), while the right extension $\langle f, \varepsilon \rangle$ ensures the first adjointness condition (5.8-00).

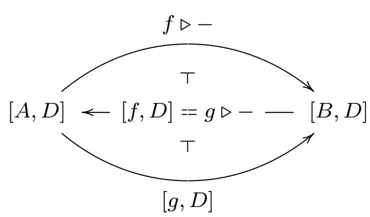



**(c) ⇒ (a):** The desired bijection between $\langle p, f \cdot q \rangle[A, C]$ and $\langle g \cdot p, q \rangle[BC]$ is obtained by pasting with $\varepsilon$ at $f$ and with $\eta$ at $g$, respectively. $\square$

**5.10.09 Corollary.** *The unit of an adjunction $f \dashv g$ is both an absolute left lifting and an absolute left extension, while the co-unit is both an absolute right lifting and an absolute right extension. In other words, $f$ has all left extensions and all right liftings, while $g$ has all left liftings and all right extensions. In particular, this implies that for any $\mathcal{B}$-object $D$ we have the following adjunctions*



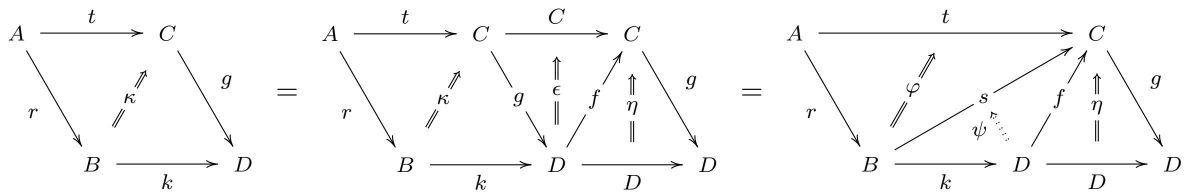

and




$\square$

**5.10.10 Theorem.** *Post-composition with left/right adjoints preserves left/right extensions.*

**Proof.** Suppose Diagram (5.10-00) exhibits $\langle s, \varphi \rangle$ as right extension of $t$ along $r$, and $C \xrightarrow{g} D$ is right adjoint with left adjoint $D \xrightarrow{f} C$. We need to show that pasting with $\varphi \cdot g$ at $s \cdot g$ is a bijection between $\langle k, s \cdot g \rangle [B, D]$ and $\langle r \cdot k, t \cdot g \rangle [A, D]$. Given a 2-cell $r \cdot k \overset{\kappa}{\Longrightarrow} t \cdot g$, we have

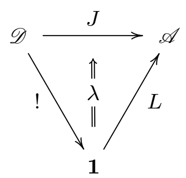



where $\psi$ is the uniquely determined 2-cell by which $\kappa f \cdot t\eta$ factors through $\varphi$. Hence $\hat{\kappa} := k\eta \cdot \psi g$ is a candidate for the image of $\kappa$. Any other such candidate $k \overset{\omega}{\Longrightarrow} sg$ composed with $s \cdot \epsilon$ must yield $\psi$, hence $\omega = \hat{\kappa}$. □

## 5.11 Limits and co-limits

### 5.11.00 Definition.

▷ In ***cat***, as well as ***Cat***, a 2-cell of the form




(5.11-00)

are called a *cone* of the functor $\mathcal{J}$. If it has the universal property of a right extension, it is called a *limit*.

▷ The category $\mathscr{A}$ is called *complete*, if it has all small(!) limits, and *finitely complete*, if it has all finite limits. Alternatively, finitely complete categories are also known as *left exact*[0] categories, or *lex* categories, for short.

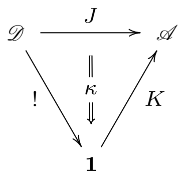
▷ A functor $\mathscr{A} \xrightarrow{F} \mathscr{B}$ *preserves* a limit $\langle L, \lambda \rangle$, if $\langle LF, \lambda F \rangle$ is again a limit.

▷ If $\mathscr{A}$ and $\mathscr{B}$ are finitely complete, or left exact, a functor $\mathscr{A} \xrightarrow{F} \mathscr{B}$ is called *left exact*, or *lex*[1], if it preserves all small limits. The full subcategory of $[\mathscr{A}, \mathscr{B}]$ spanned by all left exact functors is denoted by $\langle \mathscr{A}, \mathscr{B} \rangle$***lex***.

[0] This terminology derives from the study of "exact sequences" in algebraic topology. Unfortunately, here it clashes with the fact that limits are special cases of *right* extensions.

[1] Not to be confused with *lax* functors!

Dual Notions: a *co-cone*, resp., *co-limit*




(5.11-01)

*(finitely) co-complete category*; *right exact* or *rex* category, resp., functor; $\langle \mathscr{A}, \mathscr{B} \rangle \boldsymbol{rex}$ .

**5.11.01 Remarks.**

(0) The composition and units of $\mathscr{D}$ are irrelevant for these notions; it suffices to use graphs $\mathscr{D}$ instead of categories, and graph-morphisms $\mathcal{J}$ instead of functors. This is indicated by calling $\mathscr{D} \xrightarrow{J} \mathscr{A}$ a *diagram of shape* $\mathscr{D}$ in $\mathscr{A}$ . The notion of natural transformation still makes sense for diagrams. Diagrams of shape $\mathscr{D}$ can always be extended uniquely to functors with the free category over $\mathscr{D}$ as domain, if needed.

(1) Think of cones as "lower bounds" and of co-cones as "upper bounds" of the diagram $\mathcal{J}$ .

(2) Limits and co-limits are usually only unique up to isomorphism. Hence it is misleading to talk about "the" limit or co-limit of a diagram. There do not even have to be canonical choices. In ***set*** this is most easily seen with disjoint unions: there is no canonical way to disjointify two sets. But this also applies to the cartesian product, as there are many ways to realize the notion of ordered pair.

(3) Associated with every cone/co-cone is a source/sink (*cf.*, Definition 5.9.09), which results from pre-composing $\mathcal{J}$ with the inclusion of the discrete category $\mathscr{D}$-***Ob*** into $\mathscr{D}$ .

(4) In the $\mathscr{V}$-enriched case the unit $I$ for the tensor product need not be a terminal object of the category $\mathscr{V}$ , *i.e.*, an object that accepts a unique morphism from any object of $\mathscr{V}$ . In that case more general types of limits/co-limits are needed for a useful notion of completeness/co-completeness (weighted limits/co-limits).

The following useful result follows directly from the definition:

**5.11.02 Proposition.** *For every limit the induced source is a mono-source, while for every co-limit the induced sink is an epi-sink.* □

By abuse of notation this will be abbreviated to "every limit is a mono-source", and dually, "every co-limit is an epi-sink".

Size considerations play an important role with regard to (co-)limits:

Requiring (co-)limits of larger, *i.e.*, class-sized diagrams to exist causes the category to collapse:

**5.11.03 Theorem.** *A locally small category with all limits or all co-limits is a complete lattice. The same is true for small categories with all small limits or all small co-limits.* □

Of course, this does not rule out the existence of some large (co-)limits in non-trivial categories. According to the "shape" of the diagram $\mathscr{D}$ we distinguish various special limits and co-limits:

**5.11.04 Examples.**

(0) If $\mathscr{D}$ is empty, we obtain *terminal* / *initial* objects , much like the bottom and top elements in a complete lattice.

(1) If $\mathscr{D}$ itself has an initial / a terminal object $X$ , then $\mathscr{D} \xrightarrow{J} \mathscr{A}$ trivially has a limit / co-limit with vertex $XJ$ .

(2) If $\mathscr{D}$ is discrete, we have *products/co-products*; notice, the qualification "cartesian" works in most categories of structured sets, but may be meaningless in other settings. Notice that the source of projections always is a mono-source.

(3) If $\mathscr{D} = 0 \overset{f}{\underset{g}{\rightrightarrows}} 1$, we obtain *equalizers/co-equalizers*. Truly relevant for equalizers is only the $\mathscr{A}$-morphism into $0J$ , and this is always mono, while for co-equalizers the morphism out of $1J$ is always epi.

(4) If $\mathscr{D}$ is a co-span $1 \xrightarrow{f} 0 \xleftarrow{g} 2$, we obtain *pullbacks* as limits, but no interesting co-limits (why?). The case of $fJ = gJ$ will be of interest in connection with congruences and such pullbacks are also known as *kernel pairs*.

(5) If $\mathscr{D}$ is a span $1 \xleftarrow{f} 0 \xrightarrow{g} 2$, we obtain *push-outs* as co-limits, but no interesting limits. Push-outs in case of $fJ = gJ$ are called *co-kernel pairs*.

(6) If $\mathscr{D}$ has one object $\bullet$ and two morphisms (one being the identity on $\bullet$) and both are idempotent, then a limit as well as a co-limit of $J$ amounts to an idempotent morphism $X \xrightarrow{i} X$ that splits as an $i = e \cdot m$ such that $m \cdot e$ is an identity morphism.

(7) If $\mathscr{A}$ is an ordered set, then limits/co-limits turn into infima/suprema of the image of $\mathscr{D}$, *i.e.*, a certain subset. ◁

Besides these, another type of limit/co-limit deserves a special name as it will be intimately connected with the notion of "finiteness" that is important in automata theory:

**5.11.05 Definition.**

▷ A small category $\mathscr{D}$ is called *filtered*, if every finite diagram in $\mathscr{D}$ admits a co-cone.

▷ By a *filtered co-limit* we mean a co-limit of a functor with filtered domain.

Dual Notions: a *co-filtered* small category, resp., limit. ◁

Note that a filtered category is not empty, as the empty diagram must admit a co-cone. In the ordered or $\mathbf{2}$-enriched setting the notion of "filtered" reduces to "directed".

**5.11.06 Examples.**

- ▷ $\boldsymbol{set}$ is complete and co-complete.
- ▷ $\boldsymbol{rel}$ has all small products, and by self-duality all small co-products, but does not have all (co-)equalizers.
- ▷ In categories of structured sets with structure-preserving functions as morphisms limits can usually be constructed as for the underlying sets, while co-limits often are more complicated. This applies, for instance, to categories of EM-algebras for monads on $\boldsymbol{set}$. In particular, the co-product of two monoids $\mathcal{M} = \langle M, \cdot, e\rangle$ and $\mathcal{N} = \langle N, *, i\rangle$ is *not* their disjoint union, not even with the neutral elements identified, but rather the free monoid on $M + N$ modulo all relations that hold in $\mathcal{M}$ and in $\mathcal{N}$.

**5.11.07 Proposition.** *Left / right adjoint functors in* $\boldsymbol{Cat}$ *preserve co-limits / limits, while left / right adjoint order-preserving functions in* $\boldsymbol{pos}$ *preserve suprema / infima.*

**Proof.** This is an immediate consequence of Theorem 5.10.10. □

**5.11.08 Example.** Recall the laws for exponentiation of numbers:

$$1 = c^0 \quad , \quad c^a \cdot c^b = c^{a+b} \quad \text{and} \quad (c^a)^b = c^{a\cdot b} \quad \text{and} \quad 1^a = 1 \quad , \quad (b\cdot c)^a = b^a \cdot c^a$$

There are similar rules when it comes to function-sets. Writing $C^A$ instead of $[A, C]$ or $\langle A, C\rangle\boldsymbol{set}$, we have

$$1 \cong C^{\emptyset} \quad , \quad C^A \times C^B \cong C^{A+B} \quad \text{and} \quad \left(C^A\right)^B \cong C^{A\times B} \quad \text{and} \quad 1^A \cong 1 \quad , \quad (B\times C)^A \cong B^A \times C^A$$

Where do these rules come from? For any set $C$, the functor $[-, C]$ is self-adjoint, *i.e.*, left adjoint from $\boldsymbol{set}$ to $\boldsymbol{set}^{\mathrm{op}}$, and right adjoint from $\boldsymbol{set}^{\mathrm{op}}$ to $\boldsymbol{set}$. Both properties imply that co-limits in $\boldsymbol{set}$ are mapped to limits in $\boldsymbol{set}$. In particular,

$$[\emptyset, C] \cong 1 \quad \text{and} \quad [A + B, C] \cong [A, C] \times [B, C]$$

which are the first two laws above, albeit in different notation.

The third law arises from the adjunction $A \times - \dashv [A, -]$, since the co-unit $A \times [A, C] \xrightarrow{\boldsymbol{ev}} C$ induces a bijection between $[A \times B, C]$ and $[B, [A, C]]$. In addition, $[A, -]$ being right adjoint implies

$$[A, 1] \cong A \quad \text{and} \quad [A, B \times C] \cong [A, B] \times [A, C]$$

Finally, $A \times -$ being left adjoint implies that $\emptyset$ is absorbing with respect to $\times$ and the distributivity of $\times$ over $+$, *i.e.*,

$$A \times \emptyset = \emptyset \quad \text{and} \quad A \times (B + C) \cong A \times B + A \times C$$

which yields the other laws of arithmetic linking addition and multiplication. $\triangleleft$

### 5.11.1 Special morphisms defined via (co)limits

The fact that co-equalizers are epi (why?) is the basis to define an important class of epimorphisms:

**5.11.09 Definition.** An epimorphism $B \xrightarrow{g} C$ is called *regular*, if it arises as co-equalizer of some parallel pair $A \underset{s}{\overset{r}{\rightrightarrows}} B$ of morphisms.
Dual Notion: *regular mono*.

**5.11.10 Proposition.** *In any category $\mathscr{C}$ we have:*

(0) *Every split epi is regular.*

(1) *Every regular epi is strong.*

(2) *If $\mathscr{C}$ has pullbacks, every extremal epi is strong.*

**Proof.**

(0) If $A \xrightarrow{e} B$ has a left-inverse $B \xrightarrow{s} A$, then $id_A \cdot e = e = (e \cdot s) \cdot e$. We claim that $e$ is a co-equalizer of $A \underset{e \cdot s}{\overset{id_A}{\rightrightarrows}} A$. Consider $A \xrightarrow{f} C$ with $id_A \cdot f = f = (e \cdot s) \cdot f$. By hypothesis $e$ factors through $f$ via $e \cdot s$. Suppose $f = e \cdot g$ for some $B \xrightarrow{g} C$. Now $e \cdot g = f$ implies $g = s \cdot e \cdot g = s \cdot f$, hence $s \cdot f$ is the only possibility for $e$ to factor through $f$.

(1) Suppose $e$ in Diagram (5.9-00) is a co-equalizer of $Z \underset{v}{\overset{u}{\rightrightarrows}} A$. Since $m$ is mono, we have $u \cdot f = v \cdot f$, hence the universal property of the co-equalizer induces a diagonal $d$ that makes the upper triangle commute. By Proposition 5.11.02 $e$ is epi and hence the lower triangle commutes as well.

(2) If $\mathscr{C}$ has pullbacks, form the pullback of Diagram (5.9-00). Then $e$ factors through the pullback of $m$, which is also mono. If $e$ is an extremal epi, the pullback of $m$ has to be iso, which yields the desired diagonal. $\square$

**5.11.11 Proposition.** *In any category $\mathscr{C}$ the following are equivalent:*

(a) $A \xrightarrow{f} B$ *is mono;*

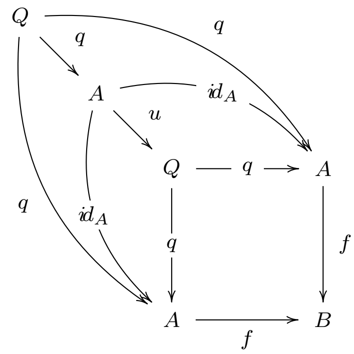
(b) $f$ *has a kernel pair of the form* $\langle q, q\rangle$ *;*

(c) $\langle id_A, id_A\rangle$ *is a kernel-par of* $f$ *.*

**Proof.**

(a) ⇒ (b) Trivial.

(b) ⇒ (c) Consider the following diagram:




where the inner square is a pullback. This induces a unique $A \xrightarrow{u} Q$ that is left-inverse to $Q \xrightarrow{q} A$. Therefore $u$ is split mono and $q$ is split epi. As the outer triangles commute trivially, $q \cdot u$ and $id_Q$ both make the upper and lower composite triangles commute and thus agree. Hence $q$ is split mono while $u$ is split epi. Therefore $q$ and $u$ are mutual inverses, which shows that the middle square is a pullback as well.

(c) ⇒ (a) Trivial. □

## 5.12 Completions and Cocompletions

If a category $\mathscr{A}$ does not have all small limits or co-limits of a certain kind, it can be desirable to suitably enlarge $\mathscr{A}$. The motivation how to do this comes again from the theory of ordered sets.

Recall the Yoneada embeddings of Theorem 5.3.05 and their order-theoretic motivation. Notice that $(-){\downarrow}$ preserves all infima that may exist in a poset $\boldsymbol{P} = \langle P, \leq\rangle$, but fails to preserve existing non-unary suprema. Hence $\langle \boldsymbol{P}^{\mathrm{op}}, \mathbf{2}\rangle\boldsymbol{ord}$ can be thought of as the result of *freely* adding suprema to $\boldsymbol{P}$. Conversely, $(\langle \boldsymbol{P}, \mathbf{2}\rangle\boldsymbol{ord})^{\mathrm{op}}$ arises by *freely* adding infima to $\boldsymbol{P}$. More precisely,

**5.12.00 Proposition.** *For any monotone function* $\langle P. \leq\rangle \xrightarrow{f} \langle Q, \sqsubseteq\rangle$ *into a complete lattice there exists a unique monotone function* $\langle \mathscr{P}^{\mathrm{op}}, \mathbf{2}\rangle\boldsymbol{ord} \xrightarrow{\bar{f}} \langle Q, \sqsubseteq\rangle$ *that preserves suprema and satisfies* $f = \mathcal{D} \cdot \bar{f}$ *.*

**Proof.** HW. □

This simple result about **2**-enriched categories has a counterpart for ordinary ***set***-enriched categories:

**5.12.01 Proposition.** *Consider an essentially small category* $\mathscr{A}$ .

(0) *The functor category* $[\mathscr{A}^{\mathrm{op}}, \boldsymbol{set}]$ *Is complete and co-complete,* i.e., *has all small limits and co-limits.*

(1) *The Yoneda-embedding* $\mathscr{A} \xrightarrow{\boldsymbol{Y}} [\mathscr{A}^{\mathrm{op}}, \boldsymbol{set}]$ *preserves all limits that exist in* $\mathscr{A}$ *, but no non-trivial co-limits.*

(2) *Any functor* $\mathscr{A} \xrightarrow{F} \mathscr{B}$ *into a co-complete category has an essentially unique extension* $[\mathscr{A}^{\mathrm{op}}, \boldsymbol{set}] \xrightarrow{\bar{F}} \mathscr{B}$ *that satisfies* $F = \boldsymbol{Y} \cdot \bar{F}$ *and preserves all co-limits.*

**Proof.** HW. □

In view of this result the functor category $[\mathscr{A}^{\mathrm{op}}, \boldsymbol{set}]$ is sometimes referred to as the *free co-completion* of $\mathscr{A}$ , while $[\mathscr{A}, \boldsymbol{set}]^{\mathrm{op}}$ serves as the *free completion*.

In particular, if one wants to *freely* add a specific type of co-limits, respectively, limits to a given small category $\mathscr{A}$ (see Example 5.11.04, Definition 5.11.05), this can be accomplished by taking appropriate sub-categories of $[\mathscr{A}^{\mathrm{op}}, \boldsymbol{set}]$ , or of $[\mathscr{A}, \boldsymbol{set}]^{\mathrm{op}}$ .

The construction of useful dualities that allow the combination of Birkhoff's and Eilenberg's theorems in the theory of formal languages, utilizes the completion under co-filtered limits, respectively, the co-completion under filtered co-limits of essentially finite categories. Hence these constructions deserve special names.

**5.12.02 Definition.** By $\mathscr{A}$***-ind*** we denote the *free co-completion* of an essentially small category $\mathscr{A}$ *under filtered co-limits*, *i.e.*, the least subcategory of $[\mathscr{A}, \boldsymbol{set}]^{\mathrm{op}}$ that contains $\mathscr{A}$ (qua representables) and is closed under filtered co-limits.
DUAL NOTIONS: the *free completion* $\mathscr{A}$***-pro*** *under co-filtered limits.*

Obviously, one would like to characterize categories that are equivalent to categories of the form $\mathscr{A}$***-ind*** or $\mathscr{A}$***-pro*** for essentially small $\mathscr{A}$ . These form the categorical counterpart of algebraic lattices, which are useful in domain theory.

**5.12.03 Definition.** An object $X$ of a category $\boldsymbol{C}$ is called *finitely presentable*, alternatively *compact*, or even *finite*, if the representable functor $\mathscr{C} \xrightarrow{\langle X, - \rangle \mathscr{C}} \boldsymbol{set}$ preserves finite co-limits. We write $\mathscr{C}_{\mathrm{fp}}$ for the full subcategory spanned by all finitely presentable objects of $\mathscr{C}$ .

**5.12.04 Theorem.** **(*cf.* nlab)** *The following conditions for a category* $\mathscr{C}$ *are equivalent:*

(a) *$\mathscr{C}$ is co-complete* (i.e., *has all small co-limits), $\mathscr{C}_{\mathrm{fp}}$ is locally small, and any $\mathscr{C}$-object $A$ is a filtered co-limit of the canonical diagram induced by the comma category $\mathscr{C} \downarrow A$.*

(b) *There exists a small finitely complete category $\mathscr{K}$ such that $\mathscr{C}$ is equivalent to the full subcategory of $[\mathscr{K}, \boldsymbol{set}]$ spanned by all functors that preserve finite limits,* i.e., *the category of* models *of an* essentially algebraic theory $\mathscr{K}$

(c) *$\mathscr{C}_{\mathrm{fp}}$ has finite co-limits and the image of $\mathscr{C}$ under the restricted Yoneda-embedding*

$$\mathscr{C} \xrightarrow{\ \boldsymbol{Y}\ } [\mathscr{C}^{\mathrm{op}}, \boldsymbol{set}] \xrightarrow{[I^{\mathrm{op}}, \boldsymbol{set}]} [{\mathscr{C}^{\mathrm{op}}}_{\mathrm{fp}}, \boldsymbol{set}]$$

*where $\mathscr{C}_{\mathrm{fp}} \xrightarrow{I} \mathscr{C}$ is the inclusion, essentially coincides with the full subcategory of all functors ${\mathscr{C}^{\mathrm{op}}}_{\mathrm{fp}} \longrightarrow \boldsymbol{set}$ that preserve finite limits.*

*Such categories are called* locally finitely presentable, *or* lfp, *for short.*

**5.12.05 Examples.**

▷ $\boldsymbol{set}$ is lfp: $\boldsymbol{set}$ is co-complete and $\boldsymbol{set}_{\mathrm{fp}}$ consists of the finite sets. Given any set $A$, consider the diagram induced by all functions $X \xrightarrow{f} A$ with $X$ finite. Their image-factorizations determine finite subsets of $A$, which suffice for specifying the co-limit. This is just the directed union of these finite subsets, which, of course, coincides with $A$.

▷ If $\boldsymbol{T}$ is a monad over $\boldsymbol{set}$, the category $\mathscr{C} = \boldsymbol{set}^{\boldsymbol{T}}$ of EM-algebras is lfp: Co-limits in $\mathscr{C}$ are constructed as quotients of free algebras over the underlying sets of the $T$-algebras in the diagram, hence $\mathscr{C}$ is co-complete. For $\mathscr{C}_{\mathrm{fp}}$ we use the finitely generated $T$-algebras, *i.e.*, the quotients of free $T$-algebras over finite sets. Clearly, any such free $T$-algebra up to isomorphism can only have a set of quotients, since any $T$-homomorphism is determined by its values on the generators. Finally, every $T$-algebras is the directed union of its finitely generated sub-algebras.

Notice that the constructions above applied to small categories. If $\mathscr{A}$ is not small, its functor cateogories $[\mathscr{A}^{\mathrm{op}}, \boldsymbol{set}]$ and $[\mathscr{A}, \boldsymbol{set}]^{\mathrm{op}}$ can fail to be locally small. This problem can be overcome in two ways: either one introduces a "hierarchy of smallness" by requiring a sequence of inaccessible cardinals that determine larger and larger "universes" of sets, each "small" with respect to the next one, or one uses the diagrams into $\mathscr{A}$ as objects of the desired completion or co-completion.

## 5.13 Congruences

Any function $B \xrightarrow{g} C$ in $\boldsymbol{set}$ induces an equivalence relation on the domain $B$ by identifying elements with the same $g$-image. In fact, every equivalence relation on sets arises in this fashion.

For a concrete category $\langle \mathscr{C}, |-| \rangle$ consider the equivalence relation $\sim_g$ induced by a homomorphism $B \xrightarrow{g} C$. Under what conditions does the set $|B|_{\sim_g}$ of equivalence classes carry a $\mathscr{C}$-structure that turns the surjective function $|B| \xrightarrow{q} |B|_{\sim_g}$ into a $\mathscr{C}$-morphism with domain $B$? This would warrant calling $\sim_g$ a *congruence*.

Provided the following conditions are satisfied

- ▷ the factorization in ***set*** of $|g|$ into a surjection $e$ followed by an injection $m$ determines a sub-object $g$-***img*** of $C$;
- ▷ any bijection into the set $|g\text{-}\boldsymbol{img}|$ admits a lifting to an isomorphisms with codomain $g$-***img***; this property is sometimes called *transportability*.

the set of equivalence classes is guaranteed to carry a $\mathscr{C}$-structure isomorphic to $g$-***img***. But even without transportability the surjective $\mathscr{C}$-morphism $B \xrightarrow{e} g\text{-}\boldsymbol{img}$ deserves to be called a *quotient* in $\mathscr{C}$.

Of course, not every equivalence relation on $|B|$ ha to be a congruences with respect to $\mathscr{C}$.

If $\mathscr{C}$ is an abstract category without (obvious) faithful functor into ***set***, the notions of congruence and quotient are somewhat more subtle. First we need to ensure that $\mathscr{C}$ admits a reasonable calculus of spans and relations.

**5.13.00 Definition.** Let $\mathscr{C}$ be a category with pullbacks. The bi-category $\mathscr{C}$-***spn*** of $\mathscr{C}$-spans consists of

- ▷ $\mathscr{C}$-objects as objects;
- ▷ 2-sources $\boldsymbol{R} = (A \xleftarrow{r_0} R \xrightarrow{r_1} B)$ of $\mathscr{C}$-morphisms as 1-cells from $A$ to $B$; these are called *spans*, while the mono-sources among the spans are also known as *relations*;
- ▷ $\mathscr{C}$-morphisms $R \xrightarrow{t} S$ that make the obvious triangles commute as 2-cells from $\boldsymbol{R}$ to $\boldsymbol{S} = (A \xleftarrow{s_0} S \xrightarrow{s_1} B)$.

The 1-cell composition of $\boldsymbol{R}$ with $\boldsymbol{S} = (B \xleftarrow{s_0} T \xrightarrow{s_1} C)$ is formed by selecting a specific pullback of the co-span $R \xrightarrow{r_1} B \xleftarrow{s_0} S$, while the identity spans have two $\mathscr{C}$-identities as components. (Notice that the span composition of relations in general fails to be a relation.)
Dual notion: the bicategory $\mathscr{C}$-***csp*** with *co-spans* $\boldsymbol{R} = (A \xrightarrow{r_0} R \xleftarrow{r_1} B)$ as 1-cells.

**5.13.01 Remarks.**

(a) In ***set*** a span from $A$ to $B$ is the obvious generalization of a graph, where all arrows start at some element of $A$ and end at some element of $B$. It may also be seen as an $A \times B$-matrix of (hom-)sets. Span-composition with a span from $B$ to $C$ considers all possible paths of length 2 from elements in $A$ to elements of $C$; it corresponds to a matrix-product where multiplication and addition are replaced by cartesian product and disjoint union, respectively.

Relations from $A$ to $B$ are spans with at most one arrow linking any $a \in A$ with each $b \in B$. Hence the span-composition of relations can fail to be a relation; in general parallel arrows can arise by span composition which need to be identified to arrive at a relation. This amounts to forming the epi–mono-source factorization of the composite span.

(b) Since the composition in ***spn*** is defined by means of pullbacks, which do not need to have canonical representatives and hence involve choices, it is not clear, if these choices can be made in such a way as to make the composition of spans strictly associative. Instead one can accept the mediating isos, in particular since they are well-behaved or "coherent".

(c) ***spn*** is closed in the sense of Definition 5.10.02.

(d) Monads in ***spn*** are small categories: an endo-span on a set $C$ assigns hom-sets of a graph with node-set $C$, while $\eta$ and $\mu$ provide the identities and the composition.

To be able to extract relations from spans, at least in categories with finite products, we need the following notion:

**5.13.02 Definition.** The *image* of a $\mathscr{C}$-morphism $A \xrightarrow{f} B$, if it exists, is the smallest sub-object of $B$ through which $f$ factors. It will be denoted by $f$-***img***.

**5.13.03 Proposition.** *If $A \xrightarrow{f} B$ has an image-factorization $A \xrightarrow{e} f\text{-}\boldsymbol{img} \xrightarrow{m} B$, then $A \xrightarrow{e} f\text{-}\boldsymbol{img}$ is an extremal epi.* $\square$

In the presence of finite products spans $\boldsymbol{R} = (A \xleftarrow{r_0} R \xrightarrow{r_1} B)$ can be identified with morphisms into $R \xrightarrow{\langle r_0, r_1 \rangle} A \times B$, and taking the image of the latter results in sub-objects of $A \times B$, which bijectively correspond to mono-spans from $A$ to $B$. Since we already required $\mathscr{C}$ to have pullbacks to enable the composition of spans, the step to require finite products is a rather small one as it amounts to requiring an initial object in addition to pullbacks.

While a kernel pair, viewed as an "internal relation", due to its universal property as a pullback is trivially reflexive and symmetric, its transitivity initially takes the somewhat strange form that the composite span(!) factors through the kernel pair. If image factorizations exist, the mediating morphism can be factored, which produces the composite relation and shows that it is contained in the original relation.

**5.13.04 Definition.** We will refer to mono-spans $\boldsymbol{R} = (A \xleftarrow{r_o} R \xrightarrow{r_1} B)$ between $\mathscr{C}$-objects $A$ and $B$ as *internal relations* from $A$ to $B$.

An internal relation $\boldsymbol{R} = (A \xleftarrow{r_o} R \xrightarrow{r_1} A)$ on $A$ is called a *congruence* or *internal equivalence relation*, if the composite span exists and the identity span, the opposite span and the composite span all factor through $\boldsymbol{R}$ by means of morphisms $A \xrightarrow{r} R$, $R \xrightarrow{s} R$ and $T \xrightarrow{t} R$. (Note that $t$ need not be mono.)

A congruence is called *effective*, if it is the kernel pair of some morphism.

It is quite elementary to show that:

**5.13.05 Proposition.** *Every kernel pair is a congruence.* □

In general, a good notion of composition for internal relations requires (regular-epi,mono)-factorizations to exist.

The following definition will ensure that image factorizations with a regular epi as first factor exist by requiring that congruences built via kernel pairs do have quotients, analogous to sets of equivalence classes equipped with the relevant structure.

**5.13.06 Definition.** A finitely complete category $\mathscr{C}$ is called *regular*, if

▷ co-equalizers of kernel pairs exist in $\mathscr{C}$ ;

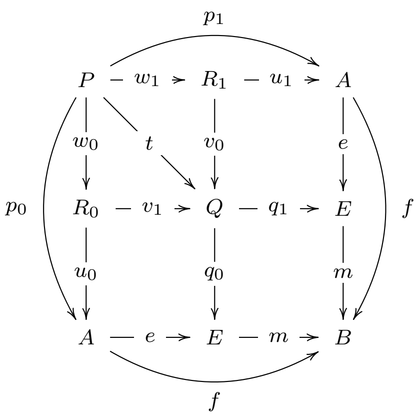
▷ regular epis are stable under pullback;

and *exact* or *Barr-exact*, if in addition

▷ every congruence is a kernel pair.

**5.13.07 Theorem.** *Any morphism of a regular category can be factored as a regular epi followed by a mono, and this factorization is essentially unique.*

**Proof.** Consider the kernel pair $\langle p_0, p_1\rangle$ of $A \xrightarrow{f} B$ and form its co-equalizer $A \xrightarrow{e} E$. Since $p_0 \cdot f = p_1 \cdot f$ there exists a unique morphism $E \xrightarrow{m} B$ satisfying $f = e \cdot m$ . It remains to show that $m$ is mono. For this purpose we sub-divide the pullback diagram for the kernel pair into four smaller pullbacks:




Because of regularity, $v_i$ and $w_i$ , $i < 2$ , are regular epis, which in particular implies that $t$ as the composition of two epis is epi. Now

$$t \cdot q_0 = p_0 \cdot e = p_1 \cdot e = t \cdot q_1$$

shows that $q_0 = q_1$ , hence by Proposition 5.11.11 $m$ is mono. □

**5.13.08 Corollary.** *In regular categories every extremal epi is regular.* □

In exact categories every internal equivalence relation is guaranteed to have a quotient; this may fail in merely regular categories. Fortunately, very many of the relevant categories in practice are exact:

**5.13.09 Theorem. (*cf.* nlab)**

- ▷ *Any category monadic over some power* $\boldsymbol{set}^n$ *is exact.*
- ▷ *Any abelian category (= category enriched in* $\boldsymbol{ab}$ *, the category of abelian groups with the tensor product) is exact.*
- ▷ *Any topos (= replacement for* $\boldsymbol{set}$ *particularly suitable for "geometric" and constructive logic) is exact.*
- ▷ *Any category of models of a Lavwere theory (a particularly nice type of algebraic theory) in an exact category is again exact.* □

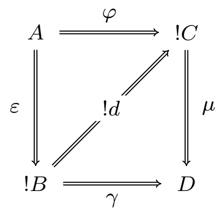
**5.13.10 Example.** In view of the fact that the "forgetful" functor $\boldsymbol{mon} \xrightarrow{U^{\mathbb{L}}} \boldsymbol{set}$ as a right adjoint preserves limits, congruences on a monoid $M$ are simply equivalence relations on $M$ that are sub-monoids of $M \times M$ .

## 5.14 Diagonalization relations

An abstract version of the interplay between quotients and sub-objects will be needed for the categorical version of Birkhoff's Theorem in Section 4.1. But useful relations of this type are not confined to epis and monos.

**5.14.00 Definition.** Consider discrete functors $\mathscr{I} \xrightarrow{A} \mathscr{C} \xleftarrow{D} \mathscr{J}$ into $\mathscr{C}$ .

(0) A sink $A \xRightarrow{\varepsilon} !B$ and a source $!C \xRightarrow{\mu} D$ are called *(weakly) orthogonal*, if whenever a sink $A \xRightarrow{\varphi} !C$ and a source $!B \xRightarrow{\gamma} D$ satisfy $\varphi \cdot \mu = \varepsilon \cdot \gamma$ there exists a (not necessarily) unique $B \xrightarrow{d} C$ with




(5.14-00)

(1) In the special case that $\mathscr{I}$ is a singleton and $\mathscr{J}$ is empty, we say that the object $C$ is *right-orthogonal* to the morphism $A \xrightarrow{e} B$, or *injective* with respect to $e$ in the weak case.

Dual Notions: The object $B$ is *left-orthogonal*, respectively, *projective* with respect to the morphism $C \xrightarrow{m} D$.

**5.14.01 Definition.** A *factorization system* $\langle \mathcal{E}, \mathcal{M} \rangle$ on $\mathscr{D}$ consists of two classes $\mathcal{E}, \mathcal{M} \subseteq \mathscr{D}_1$ of $\mathscr{D}$-morphisms subject to

**(EM-0)** $\mathcal{E}$ and $\mathcal{M}$ are closed and composition with isomorphisms;

**(EM-1)** $\mathscr{D}_1 = \mathcal{E} \cdot \mathcal{M}$ ;

**(EM-2)** the classes $\mathcal{E}$ and $\mathcal{M}$ are *orthogonal* in the following sense: whenever $f \cdot m = e \cdot g$ with $e \in \mathcal{E}$ and $m \in \mathcal{M}$, there exists a unique diagonal $d$ with $e \cdot d = f$ and $d \cdot m = g$, *cf.* condition (1) for strong epimorphisms, Diagram 5.9-00.

A factorization system $\langle \mathcal{E}, \mathcal{M} \rangle$ is called *proper*, if $\mathcal{E}$ consists of epis and $\mathcal{M}$ consists of monos.

**5.14.02 Proposition.** *Let* $\boldsymbol{T} = \langle T, \eta, \mu \rangle$ *be a monad on a category* $\mathscr{D}$ *with a factorization system* $\langle \mathcal{E}, \mathcal{M} \rangle$ *. If* $\mathcal{E}$ *is stable under* $T$ *, then the classes* $\mathcal{E}^{\boldsymbol{T}}$ *and* $\mathcal{M}^{\boldsymbol{T}}$ *of* $T$*-homomorphisms in* $\mathcal{E}$ *, resp.* $\mathcal{M}$ *form a factorization system of* $\mathscr{D}^{\boldsymbol{T}}$ *.*

**Proof.** Easy application of the diagonalization property. □

## 5.15 Ideals (?)

A special case of Example 5.2.08 above deserves a special name:

**5.15.00 Definition.** Profunctors that arise as sub-functors of hom-functors[2] are called *2-sided ideals* or just *i*deals of $\mathscr{C}$ .

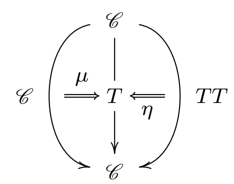
**5.15.01 Example.** If a profunctor $\mathscr{C} \xrightarrow{L} \mathbf{1}$, *i.e.*, $\mathscr{C}^{\mathrm{op}} \times \mathbf{1} \xrightarrow{L} \boldsymbol{set}$ satisfies $\langle C, 1 \rangle L \subseteq C \downarrow \mathscr{C}$ , we call $\mathcal{L}$ a *left ideal*.

Similarly, if a profunctor $\mathbf{1} \xrightarrow{R} \mathscr{D}$, *i.e.*, $\mathbf{1}^{\mathrm{op}} \times \mathscr{D} \xrightarrow{R} \boldsymbol{set}$ satisfies $\langle 1, D \rangle R \subseteq \mathscr{D} \downarrow D$ , we call $\mathcal{R}$ a *right ideal*. ◁

**5.15.02 Remark.** Profunctors $\mathscr{C} \xrightarrow{L} \mathbf{1} \xrightarrow{R} \mathscr{D}$ always compose to a profunctor $\mathscr{C} \xrightarrow{L \cdot R} \mathscr{D}$.

If $\mathscr{C} = \mathscr{D}$ , $\mathcal{L}$ is a left ideal and $R$ is a right ideal, the intersection $\langle C, 1 \rangle L \cap \langle 1, C' \rangle R$ always contains a subset of $\langle C, C' \rangle \mathscr{C}$ . **Conjecture:** these subsets constitute a 2-sided ideal on $\mathscr{C}$ .

There is, however, a second approach to ideals that is more widely applicable and better indicates what ideals are good for. Remark **??** indicated that the monads in a 2-category again form a nice 2-category. Monads $\boldsymbol{T}$ on an endomorphism $\mathscr{C} \xrightarrow{T} \mathscr{C}$ in this setting take the form




[2] There has to be a natural transformation into the hom-functor that point-wise is an inclusion.

subject to the usual identity and associativity laws.

**5.15.03 Definition.** An monomorphism $I \xrightarrow{\iota} T$ is called a *left ideal*, if the composite 2-cell $TI \xrightarrow{T\iota} TT \xrightarrow{\mu} T$ factors through $\iota$ by means of a 2-cell $TI \xrightarrow{\lambda} I$.

DUAL NOTION: *right ideal*, where $IT \xrightarrow{\iota T} TT \xrightarrow{\mu} T$ factors through $\iota$ by means of a 2-cell $IT \xrightarrow{\rho} I$.

A *2-sided ideal*, or just *ideal*, is both a left- and a right-ideal.

**5.15.04 Examples.**

▷ Ordinary monoids are monads in the suspension of ***set***, with a single object, sets as 1-cells that compose with cartesian product, and ordinary functions as 2-cells.

An left ideal for a monoid $\langle M.\cdot, e\rangle$ hence is a subset $A \subseteq M$ such that the restriction of $M \times M \xrightarrow{\cdot} M$ to $M \times A$ factors through $A$, *i.e.*, $M \cdot A \subseteq A$, or even $M \cdot A = A$, since $M$ contains a neutral element.

▷ Recall that rings are monoids in the suspension of the category ***ab*** of abelian groups, with a single object, abelian groups as 1-cells that compose with tensor product, and group homomorphisms as 2-cells. A left ideal for a ring $\langle R, \cdot, 1\rangle$ is a subgroup $A$ of $\mathcal{R}$, such that the restriction of $R \otimes R \xrightarrow{\cdot} R$ factors through $A$, *i.e.*, $R \cdot A \subseteq R$.

▷ Categories themselves are monoids in the 2-category ***spn*** of spans over sets, *cf.* Remark 5.13.01(d). Hence left ideals in some category $\mathscr{C}$ are sets $A$ of $\mathscr{C}$-morphisms such that every pre-composition with some arrow in $A$ again belongs to $A$.

In order to speak about "principal ideals" (of either type) it is necessary that the 1-cells are somehow set-based, so one can distinguish "elements" and therefore has ideals "generated" by such elements. In the suspension of ***set*** or ***ab***, the 1-cells are sets (with an abelian group structure in the second case), so the notion of principal ideal is clear. Spans, on the other hand, assign hom-sets to pairs of elements in the domain, resp. codomain of the span. Hence one can pick an element of one of the hom-sets and generate an ideal from there. For a (small) category $\mathscr{C}$, the arrows factoring through some fixed $X \xrightarrow{a} Y$ form a principal 2-sided ideal, while those with $a$ as last, resp. first factor form a left, resp. right, principal ideal.

One can also interpret 2-sided ideals as generalizations of zero morphisms in the sense outlined below.

If $\mathscr{C}$ is enriched in $1 \downarrow$ ***set***, then every hom-set $[A, B]$ has a distinguished element $0_{A,B}$, a so-called *zero morphism*, such that all compositions with zero morphisms are again zero morphisms. In a sense, the zero morphisms act like a “typed absorber”.

Recall from Remark 5.1.02 that collapsing all hom-sets of a category $\mathscr{C}$ to a singleton yields a potentially large pre-ordered set and a full functor from $\mathscr{C}$ into the latter. This process can be refined using ideals, which themselves may be thought of as absorbing subsets of all $\mathscr{C}$-morphisms.

**5.15.05 Lemma.** *If $\mathscr{C}$ is a category with an ideal $A$, then collapsing all parallel $A$-morphisms into a single one results in a new category $\mathscr{C}(A)$ and a full functor from $\mathscr{C}$ into the latter.*

**5.15.06 Remarks.**

- ▷ Since $\mathscr{C}$ may have several connected components, categories of the form $\mathscr{C}(A)$ need not be enriched in $1 \downarrow \textbf{\textit{set}}$.
- ▷ Even in a connected component there can be morphisms that do not factor through any $A$-morphism, and there can be hom-sets that do not intersect the ideal $A$.
- ▷ For every functor $\mathscr{C} \xrightarrow{F} \mathscr{D}$ the pre-image of all zero morphisms in $\mathscr{D}$ ought to be an ideal in $\mathscr{C}$, and in fact every ideal in $\mathscr{C}$ can be obtained in this fashion, even from a full functor that is surjective on objects.

# 6 Pre-ordered and partially ordered sets

As categories generalize pre-ordered sets, various concepts of order theory can be generalized to the categorical setting. As a reference, we recall some of them here.

**6.0.00 Definition.**

(0) A *pre-ordered set* $\langle P, \sqsubseteq\rangle$ consists of a set $P$ and a reflexive transitive relation $\sqsubseteq$. If the latter is anti-symmetric as well, one has a *partially ordered set*, or *poset*, for short. If in addition $\sqsubseteq \cup \sqsubseteq^{\mathrm{op}} = P \times P$, we speak of a *linearly ordered set*.

(1) A element $x \in P$ is called an *upper bound* of a subset $A \subseteq P$, written $x \sqsubseteq A$, provided $x \sqsubseteq a$ for every $a \in A$. The set of all upper bounds of $A$ is denoted by $A^{\uparrow}$. A *least upper bound*, or *supremum* $A\bigsqcup$ is an element of $A^{\uparrow}$ satisfying $A\bigsqcup \sqsubseteq A^{\uparrow}$.

Dual notions: *lower bound*; $A_{\downarrow}$; *greatest lower bound* or *infimum* $A\bigsqcap$.

(2) Functions $P \xrightarrow{f} Q$ between pre-ordered sets or between posets $\langle P, \sqsubseteq\rangle$ and $\langle Q, \leq\rangle$ are called *order-preserving*, or *monotone*, if $x \sqsubseteq y$ implies $xf \leq zf$ for all $x, y \in P$.

(3) Relations $P \xrightarrow{R} Q$ between pre-ordered sets are called *order-ideals*, if they satisfy $\sqsubseteq; R; \leq\ \subseteq R$ (which by the reflexivity of pre-orders is equivalent to $\sqsubseteq; R; \leq\ = R$). Notice that order-preserving functions usually are not order-ideals, unless both orders are discrete. IN the general case one obtains an order ideal by forming the lower segment of the function graph in $\langle P, \sqsubseteq\rangle^{\mathrm{op}} \times \langle Q \leq\rangle$.

**6.0.01 Remark.** Notice that the supremum, respectively, infimum of $\emptyset \subseteq P$, if it exists, has to be a smallest, resp., largest element element of $P$, usually denoted by $\bot$, resp. $\top$.

**6.0.02 Definition.**

(0) A poset $\langle P, \sqsubseteq \rangle$ where any two elements $x, y \in P$ have a supremum, usually denoted by $x \sqcup y$, is called a $\sqcup$-*semi-lattice*. The slightly stronger notion of $\sqcup$-*semi-lattice with* $\bot$ results from requiring all finite subsets to have a supremum, including the empty set, whose supremum is a least element and thus denoted by $\bot$.

From an algebraic point of view, the $\sqcup$-operator is associative, commutative and idempotent, hence $\langle P, \sqcup \rangle$ is an idempotent commutative semi-group. Since one can recover the order by setting $x \sqsubseteq y$ gdw. $x \sqcup y = y$, and therefore a semi-group morphism automatically preserves the order, the categories $\sqcup$-***slat*** of $\sqcup$-semi-lattices and ***ic-sgr*** of idempotent commutative semi-groups are isomorphic.

Dual notion: $\sqcap$-*semi-lattice*; $\sqcap$-*semi-lattice with* $\top$.

(1) A poset that is both a $\sqcup$-and a $\sqcap$-semi-lattice is called a *lattice*. From the algebraic point of view the operations $\sqcup$ and $\sqcap$ have to be compatible in the sense that both induce the same order relation. This amounts to the *absorption rules* $a \sqcap (a \sqcup b) = a = (b \sqcap a) \sqcup a$.

(2) A lattice $\langle L, \sqsubseteq \rangle$ is called *complete*, if every subset $A \subseteq L$ has a supremum $A\bigsqcap$, or equivalently, every subset $B$ has an infimum $B\bigsqcap$ (namely the supremum of its set $A = B_{\downarrow}$ of lower bounds). A lattice is called *distributive*, provided $\sqcup$ and $\sqcap$, viewed as binary operations, distribute over each other, while a complete lattice is *completely distributive lattice*, if arbitrary suprema distribute over arbitrary infima, and vice versa.

**6.0.03 Examples.**

- Every power-set $XP$ ist partially ordered by set-inclusion $\subseteq$, in fact a completely distributive lattice with suprema given by union $\bigcup$, and infima given by intersection $\bigcap$.

- The natural numbers $\mathbb{N} = \{0, 1, 2, \ldots\}$ are linearly ordered under $\leq$; finite nonempty suprema and infima are given by max and min, respectively. The empty supremum exists and is $0$, but there is no empty infimum, *i.e.*, largest element.

- The division order $\mid$ on the natural numbers is a distributive lattice with $\top = 0$ and $\bot = 1$. Finite infimum and supremum are given by gcd and lcd, respectively, while infinite suprema/infima do not exist.

Alternatively, semi-lattices and lattices can be characterized algebraically as follows:

**6.0.04 Proposition.** *There is a bijective correspondence between* $\sqcap$*-semi-lattices and idempotent commutative semi-groups.*

**Proof.** Clearly, the $\sqcap$-operation of a $\sqcap$-semi-lattice is associative, commutative and idempotent.

Conversely, given an idempotent commutative semi-group $\langle S, \cdot \rangle$, define a relation $\sqsubseteq \ \subseteq S \times S$ by $x \sqsubseteq y$ iff $x \cdot y = x$. Since $\cdot$ is idempotent, $\sqsubseteq$ is reflexive, while associativity of $\cdot$ guarantees transitivity of $\sqsubseteq$, which therefore is a pre-order.

If $x \sqsubseteq y$ and $z \in S$ then $(z \cdot x) \cdot (z \cdot y) = x \cdot x$ and hence $z \cdot x \sqsubseteq z \cdot y$, *i.e.*, the multiplication $\cdot$ preserves this order.

Idempotency of $\cdot$ together with transitivity implies that $x, y \sqsubseteq x \cdot y$, hence $x \cdot y$ is an upper bound of $x$ and $y$. given another upper bound $z \in S$ of $x$ and $y$, observe that

$$(x \cdot y) \cdot z \times \cdot (y \cdot z) = x \cdot y$$

and therefore $x \cdot y$ is the least upper bound of $x$ and $y$. □

**6.0.05 Corollary.** *Reversing the order induces a bijective correspondence between $\sqcap$-semi-lattices and $\sqcup$-semi-lattices, in fact an isomorphism between the categories $\sqcap$-**slat** and $\sqcup$-**slat**.* □

**6.0.06 Remark.** The notions of $\sqcup$-semi-lattice and of (complete) lattice seem to presuppose a partial or at least pre-ordering $\sqsubseteq$. This, however, is only an illusion. An operation @ that is defined at least on non-empty finite sets may be thought of alternatively as an associative idempotent binary operation, possibly with a unit $\bot = \emptyset @$. Then the ordering may be derived via $x \sqsubseteq y$ iff $\{x, y\}@ = y$. In that case @ becomes the supremum-operation for $\sqsubseteq$. But one may equally well define the opposite order $x \sqsupseteq y$ by the same formula $\{x, y\}@ = y$. Of course, then @ has to be interpreted as the infimum operation for $\sqsupseteq$. Using a symbol like $\sqcup$ or $\bigsqcup$ instead of @ for the operation on finite subsets expresses a preference for one of the two possible orders. While the second order is not ruled out by this choice, the reversed notation for suprema resp. infima quickly becomes confusing and error-prone.

# References


[BH76] B. Banaschewski and H. Herrlich. Subcategories defined by implications. *Houston J. Math.*, 2(2):149–171, 1976.

[DA11] MartÃŋn-Vide C. (eds) Dediu AH., Inenaga S., editor. *GreenâĂŹs Relations and Their Use in Automata Theory*, volume 6638 of *LNCS*. Springer, Berlin, Heidelberg, 2011.

[DD85] Brian A. Davey and Gary Davis. Tensor products and entropic varieties. *Algebra Universalis*, 21:68–88, 1985.

[Mac71] Saunders MacLane. *Categories for the Working Mathematician*. Number 5 in Graduate Texts in Mathematics. Springer-Verlag, Berlin, Heidelberg and New York, 1971.

[Pin18] Jean-Eric Pin. Mathematical foundations of automata theory. (preliminary version of a book), June 2018.

[SW78] Ross Street and R. F. C. Walters. Yoneda structures on 2-categories. *Journal of Algebra*, 50:350–379, 1978.

# Index

Jürgen Koslowski (koslowj@tu-bs.de)
Theoretical Computer Science
TU Braunschweig
Mühlenpfordstr. 23
D-38106 Braunschweig
Germany