ALGEBRAIC AUTOMATA THEORY

Lecture Notes

Jürgen Koslowski

Theoretical Computer Science

Technische Universität Braunschweig

January 2018

Contents

0 Overview: the arena of formal languages

0.0 $\mathbb{R}$AT and $\mathbb{R}$EC : two classes of subsets of monoids via closures

We consider three operations, initially on sets:

- ▷ $(-)^*$, the formation of free monoids as sets of finite lists;
- ▷ $\mathbb{P}$, the formation of power-sets;
- ▷ $\mathbb{F}$, the formation of the set of finite subsets.

Formal languages over an alphabet X are subsets of X^* , *i.e.*, elements of $\mathbb{P}(X^*)$, which of course contains $\mathbb{F}(X^*)$, the set of finite formal languages over X . Instead of free monoids X^* one can form these powersets for arbitrary monoids M . Both $\mathbb{P}(M)$ and $\mathbb{F}(M)$ inherit a monoid structure from M , the latter being a sub-monoid of the former. We think of them as *fibres* over the monoid M .

[Diagram]

To obtain interesting submonoids of $\mathbb{P}(M)$ containing $\mathbb{F}(M)$, one can "vertically" close the latter under certain operations. We will be mainly interested in the subset $\mathbb{R}\text{AT}(M)$ of *rational* subsets of M , which is closed under

- ▷ finite unions (just like $\mathbb{F}(M)$);
- ▷ concatenation (just like $\mathbb{F}(M)$, hence one obtains a submonoid);
- ▷ the Kleene star (a special infinite union).

Hence rational sets may be thought of as describable by *rational expressions* (which coincide with the "regular expressions" known from TI1). The class of rational subsets of monoids will turn out to be automatically closed under direct and inverse homomorphic images.

An á priori unrelated subclass $\mathbb{R}\text{EC}(M) \subseteq \mathbb{P}(M)$ of *recognizable sets* is obtained by unioning the inverse images of $\mathbb{P}(N) = \mathbb{F}(N)$ for all homomorphisms $M \xrightarrow{f} N$, where N is a finite monoid. Compared to the construction of $\mathbb{R}\text{AT}(M)$ by fibre-wise "vertical" closures, $\mathbb{R}\text{EC}(M)$ may be thought of as a "horizontal" closure.

It may not be entirely obvious from the definition that the class of recognizable subsets is closed under finite unions and intersections as well as under direct images, but for *recognizable languages*, *i.e.*, recognizable subsets of finitely generated free monoids, this is an easy consequence of the following observation, which also justifies the term "recognizable".

0.0.00 Proposition. *The class of recognizable languages ist the class of languages that can be recognized by a finite automaton.*

Proof. Suppose $L \subseteq X^*$ is the inverse image of $U \subseteq N$ under $X^* \xrightarrow{h} N$ with N finite. Define a complete deterministic automaton $A = \langle N, e, A, \delta \rangle$, where $e \in N$ is the neutral element, and for each letter $x \in X$, the transition function $N \xrightarrow{\delta(x)} N$ is given by right-multiplication with $h(x) \in N$. It is clear that $h^{-1}[U] = L = \mathcal{L}(A)$.

Conversely, start with a complete deterministic automaton $A = \langle Q, q_o, F, \delta \rangle$, and define N to be the submonoid of the monoid of all endo-functions $Q \longrightarrow Q$ generated by the transition functions $\delta(x)$, $x \in X$. Then all those elements of N that map q_0 into F form the subset U whose inverse image under uniquely determined extension $X^* \xrightarrow{\bar{\delta}} N$ of δ to a monoid homomorphism. □

Hence finite monoids turn out to provide an alternative mechanism to recognizing rational languages.

0.0.01 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. ◁

Notice also that the minimal automaton of a language $L \subseteq X^*$ induces a minimal monoid that recognizes L , the so-called syntactic monoid, which we will study in more detail below.

Clearly, every finite monoids arise as the syntactic monoid of some recognizable language.

A somewhat surprising result connects the classes of rational and of recognizable sets:

0.0.02 Theorem. (Kleene) *Every finite set X satisfies $\mathbb{R}\mathrm{AT}(X^*) = \mathbb{R}\mathrm{EC}(X^*)$.*

0.0.03 Corollary. *For every finite set X , the set $\mathbb{R}\mathrm{AT}(X^*)$ 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 priori it is not clear, if the rational subsets of monoids that are non-finetely generated freely can change under this extension!)

In the 1960's subclasses of rational languages were intensely studied, which resulted from certain constraints on the allowed rational expressions, the so-called "star-free" languages that do not require the Kleene star for their description forming a prime example. Schützenberger managed to characterize this particular class of rational languages in terms of their recognizing syntactic monoids: they are aperiodic. Since then a large number of results in this vein have been obtained.

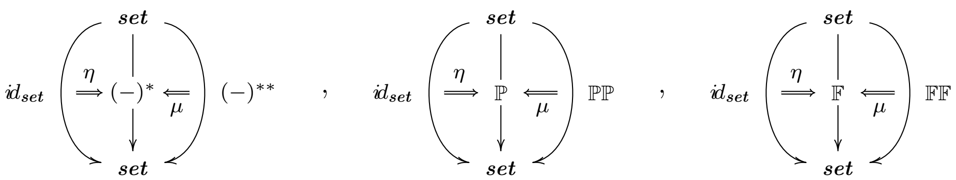
On a more gneral level, Eilenberg managed to relate so-called *varieties* of finite monoids, *i.e.*, classes of finite monoids closed under homomorphic images (H), sub-monoids (S) and under finite products (F), with nice subclasses of recognizable languages, also called "varieties".

However, the efficient specification of varietes of finite monoids remained elusive. While varieties of unconstrained algebras, *i.e.*, classes closed under (H), (S) and not necessarily finite products (P) had already been characterized by means of classes of "equations" by Birkhoff in 1935, in general this fails for varieties of finite monoids (or other algebras). Only in 1980 Reiterman used profinite methods to provide a characterization in this case as well.

0.1 The interplay of $(-)^*$ and $\mathbb{P}$, resp. $\mathbb{F}$

The operators $(-)^*$ as well as $\mathbb{P}$ and $\mathbb{F}$ are easily seen to be functors on ***set***. While for $A \xrightarrow{f} B$ the function $A^* \xrightarrow{f^*} B^*$ simply applies f to every symbol of a word $w \in A^*$, and in particular preserves the empty word, $\mathbb{P}(A) \xrightarrow{\mathbb{P}(f)} \mathbb{P}(B)$ maps $U \subseteq A$ to its f-*image* $\{ b \in B : \exists a \in A. f(a) = b \}$, which we denote by $f_\exists[U]$. The case of $\mathbb{F}$ is analogous.

In addition, all three functors carry rather canonical natural transformations that give rise to monads, see Appendix.



The list monad

Here the unit $A \xrightarrow{\eta_A} A^*$ maps letters in A to singleton words of length 1, while the multiplication $A^{**} \xrightarrow{\mu_A} A^*$ concatenates a list of words to a single word. It is straightforward to check that these families of functions indeed yield natural transformations η and μ respetively, and that μ is associative, while η is a two-sided unit for μ.

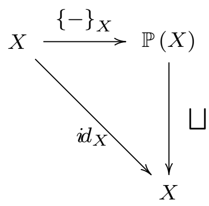
0.1.00 Proposition. *The EM-algebras for the free monoid monad bijectively correspond to classical monoids, while the EM-homomorphisms are precisely the monoid homomorphisms.* □

0.1.01 Corollary. *Every monoid is a quotient of a free monoid.* □

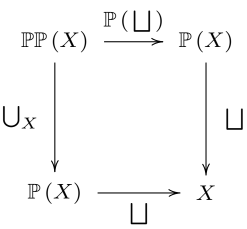
The power-set monad

Here the unit $X \xrightarrow{\eta_X} \mathbb{P}(X)$ maps elements $x \in X$ to singleton subsets $\{x\} \subseteq X$, while the multiplication $\mathbb{PP}(X) \xrightarrow{\mu_X} \mathbb{P}(X)$ forms the union of a set of subsets of X, resulting in a subset of A. Again, the proof that these yield natural transformations is easy, just as the confirmation of the monad axioms.

However, for the power-set monad it may harder to identify the EM-algebras as familiar algebras. The stucture map $\mathbb{P}(X) \xrightarrow{\bigsqcup} X$ satisfies



and



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

$$x \sqsubseteq y \qquad \text{iff} \qquad \bigsqcup\{x, y\} = y \tag{0.1-00}$$

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

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

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

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

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

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

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

Hence the EM-algebras for the power-set monad are those posets where every subset has a supremum, and hence also an infimum, *i.e.*, complete lattices. However, the corresponding morphisms are functions that only have to preserve suprema, hence one calls these EM-algebras $\bigsqcup$-*semi-lattices*, while the corresponding category is denoted by $\bigsqcup$-***slat***. Preservation of suprema automatically makes the morphisms automatically monotone.

The power-sets now can be identified as the "free $\bigsqcup$-semi-lattices", and every $\bigsqcup$-semi-lattice is a quotient of a free one.

The monad of finite subsets

After replacing $\bigsqcup$ by $\sqcup$ to indicate finiteness of the argument, one can argue essentially as above. The devinition of the partial order is exactly the same.

The only change concerns the characterization of the EM-algebras. They are no longer complete lattices, not even lattices, but just $\sqcup$-semilattices with bottom element $\bot = \sqcup\emptyset$. When every finite subset $A \in X$ has a least upper bound, it suffices to consider subsets of size 2 and of size 0. The EM-homomorphisms have to preserve finite suprema, in particular the order and $\bot$.

The set $\mathbb{F}(X)$ of finite subsets now is the free $\sqcup$-semilattice on the set X. In computer science the corresponding monad $\boldsymbol{F} = \langle \mathbb{F}, \{-\}, \cup\rangle$ is used for modelling non-determinism.

The distributive laws linking $(-)^*$ with $\mathbb{P}$ resp. $\mathbb{F}$

While, in general, monads on the same category need not give rise to a monad on the composite functor, in our special case the composition of $(-)^*$ with $\mathbb{P}$, resp. $\mathbb{F}$ can be extended to a monad. The appropriate distributive law

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

(and similarly for $\mathbb{F}$) maps a string of n (finite) subsets $A_i \subseteq X$ to their cartesian product, *i.e.*, a (finite) subset of $X^n \subseteq X^*$. The axioms of a distributive law are readily established HW.

Therefore in both cases, $\mathbb{P}$ and $\mathbb{F}$ 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?

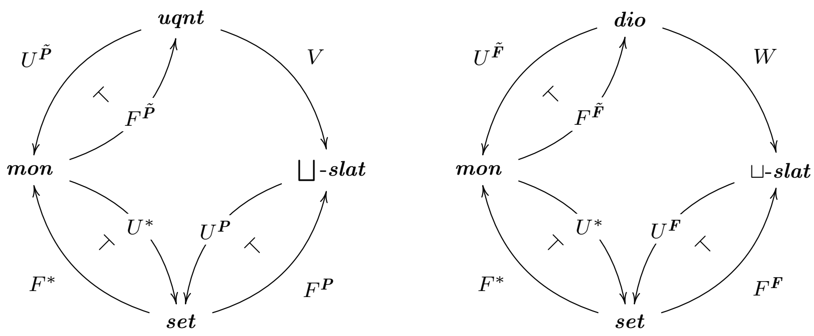
Fortunately, in both cases the categories of EM-algebras for the composite monad on ***set*** and the lifted monad on ***mon*** agree.

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

$\mathbb{F}$: we obtain the category ***dio*** of *unital dioids*, or *idempotent semirings*. Recall that a *ring* is (classically) a set R equipped with the structure of an abelian group $\langle R, +, 0\rangle$ and a monoid $\langle R, \cdot, 1\rangle$ such that multiplication $\cdot$ distributes over addition. (There are various fancyer characterizations of rings, *e.g.*, as the EM-algebras for the composition of the free abelian group monad with the free monoid monad over ***set***, or the EM-algebras of the lifting of the free abelian group monad to the category ***mon*** of monoids, or as monoids *internal* to the category ***ab*** of abelian groups with the tensor product.) One can now weaken the requirement on $\langle R, +, 0\rangle$ to be just a commutative monoid, in which case one obtains *unital semirings*, or only ask for $\langle R, +\rangle$ to be a commutative semigroup, which yields *semirings*. Adding the requirement that $+$ be idempotent 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.)

Further functors linking our categories of algebras

We have the following diagrams in ***Cat*** :



0.2 First closure properties of rational sets

By definition, rational sets are closed under

▷ finite union,

▷ multiplication,

▷ Kleene star.

The latter combines multiplication with countable disjoint union.

As direct images preserve arbitrary unions and multiplication, we immediately get

0.2.01 Proposition. *Rational sets are closed under direct images.* □

By Kleene's Theorem, the rational sets on finitely generated free monoids are closed under complement, hence by De Morgan's rules also under intersection.

For general monoids this is not the case:

0.2.02 Example. ([**?**, Example IV.1.3]) Consider the monoid $M = \{a\}^* \times \{b,c\}^*$ and the rational sets

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

with intersection

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

The projection into $\{b,c\}^*$ maps $R \cap S$ to the context-free language $\{\, b^n c^n \;:\; n \in I\!N \,\} \subseteq \{b,c\}^*$ that is well-known not to be rational (or "regular" in that context). ◁

Hence in view of De Morgan's rules rational sets in general cannot be closed under complementation.

0.2.03 Theorem. *([?, Theorem IV.1.2]) Rational subsets are closed under formation of cartesian products.*

Proof. Note that for monoids M_i, $i < 2$, the projections $M_0 \times M_1 \xrightarrow{\pi_i} M_i$ are split epi in $\boldsymbol{mon}$ with the canonical left inverses $M_i \xrightarrow{\sigma_i} M_0 \times M_1$ that combine the identity on M_i with the unit $1 \xrightarrow{e_{1-i}} M_{1-i}$.

Now the cartesian product of rational subsets $R_i \subseteq M_i$, $i < 2$, can be expressed as a multiplication of two subsets

$$R_0 \times R_1 = ((\sigma_0)_\exists[R_0]) \cdot ((\sigma_1)_\exists[R_1])$$

both of which are rational, see Proposition 0.2.06. □

0.2.04 Proposition. *([?, Proposition IV.1.3]) For every monoid M,*

$$\mathbb{R}\mathrm{AT}\,(M) = \bigcup \{\, \mathbb{R}\mathrm{AT}\,(M') \ :\ M' \text{ is a finitely generated submonoid of } M \,\}$$

Proof. It suffices to show the inclusion " $\subseteq$ ". But the right hand side contains all finite subsets of M, and hence their closure under finite unions, multiplication and Kleene star. □

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

0.2.06 Proposition. *([?, Proposition IV.1.1]) The power-monoid functor $\boldsymbol{mon} \xrightarrow{\mathbb{P}} \boldsymbol{mon}$ admits a pointwise restriction $\boldsymbol{mon} \xrightarrow{\mathbb{R}\mathrm{AT}} \boldsymbol{mon}$ that maps a monoid M to its set of rational subsets. This inherits a monad-structure $\boldsymbol{R} = \langle \mathbb{R}\mathrm{AT}, \eta_{\boldsymbol{R}}, \mu_{\boldsymbol{R}} \rangle$ from $\boldsymbol{P}$.*

Moreover, any regular epi $M \xrightarrow{\varphi} N$ induces a surjection $\mathbb{R}\mathrm{AT}\,(M) \xrightarrow{\mathbb{R}\mathrm{AT}\,(\varphi)} \mathbb{R}\mathrm{AT}\,(N)$.

Proof. Given a monoid homomorphism $M \xrightarrow{\varphi} N$, the direct image function $\mathbb{P}\,(M) \xrightarrow{\mathbb{P}\,(\varphi)} \mathbb{P}\,(N)$ is a monoid homomorphism and therefore preserve multiplication. As a left adjoint between complete lattices, it also preserves (arbitrary) unions, and thus the Kleene star. Hence $\mathbb{P}\,(\varphi)$ restricts to a monoid homomorphism $\mathbb{R}\mathrm{AT}\,(M) \xrightarrow{\mathbb{R}\mathrm{AT}\,(\varphi)} \mathbb{R}\mathrm{AT}\,(N)$. The unit of the monad $\boldsymbol{R}$ on $\boldsymbol{mon}$ carries over directly to $\boldsymbol{R}$, while for the multiplication $\mu_{\boldsymbol{R}}$ we observe that rational unions of rational sets are again rational (*cf.*. homework).

If φ is surjective, so are $\mathbb{P}\,(\varphi)$ and $\mathbb{F}\,(\varphi)$, and since $\mathbb{R}\mathrm{AT}\,(N)$ is the closure of $\mathbb{F}\,(N)$ under finite union, multiplication and Kleene star, the claim is obvious. □

Given a homomorphism $M \xrightarrow{\varphi} N$, in contrast to the direct image function $\mathbb{P}(M) \xrightarrow{\varphi_\exists} \mathbb{P}(N)$, the other two interesting functions linking the unital quantales $\mathbb{P}(M)$ and $\mathbb{P}(N)$, namely

$$\mathbb{P}(N) \xrightarrow{\varphi^{-1}} \mathbb{P}(M) \quad \text{and} \quad \mathbb{P}(M) \xrightarrow{\varphi_\forall} \mathbb{P}(N)$$

need not be monoid homomorphisms. This is immediately clear when considering the units $\{e_M\}$ and $\{e_N\}$. Nevertheless, at least φ^{-1} is still reasonably well-behaved, if we considered in a somewhat larger context.

0.2.07 Definition. Consider two ordered monoids $\langle X, \cdot, e_X, \sqsubseteq\rangle$ and $\langle Y, \cdot, e_Y, \sqsubseteq\rangle$, where the multiplication is order-preserving. A function $X \xrightarrow{\lambda} Y$ is a *lax homomorphisms*, if

$$e_M \sqsubseteq \lambda[e_n] \quad \text{and} \quad \forall u, v \in X.\, \lambda(u) \cdot \lambda \sqsubseteq \lambda((u \cdot v))(v)\,.$$

Dual Notion: *oplax homomorphism*, where the order is reversed. ◁

0.2.08 Proposition. *Given a homomorphism $M \xrightarrow{\varphi} N$, both*

$$\mathbb{P}(N) \xrightarrow{\varphi^{-1}} \mathbb{P}(M) \quad \textit{and} \quad \mathbb{P}(M) \xrightarrow{\varphi_\forall} \mathbb{P}(N)$$

are lax homomorphisms.

Proof. φ^{-1}: Clearly, $\{e_M\} \subseteq \varphi^{-1}[e_n]$. Now consider $U, V \subseteq N$, and elements $u \in \varphi^{-1}[U]$ and $v \in \varphi^{-1}[V]$ in the pre-images. Since φ is a monoid homomorphism, we have $\varphi((u \cdot v)) \in U \cdot V$, and therefore $\varphi^{-1}[U] \cdot \varphi^{-1}[V] \subseteq \varphi^{-1}[(U \cdot V)]$, as required. □

How does $\varphi_\forall$ behave with respect to the multiplication?

0.3 Recognition by homomorphism, and recognizability

0.3.01 Definition. A subset L of a monoid M is said to be *recognized by a morphism* $M \xrightarrow{\varphi} N$, if L belongs to the image of $\mathbb{P}(N)$ under the inverse image map φ^{-1}. If φ is surjective (= regular epi), one says that L is *fully* recognized. ◁

0.3.02 Definition. A subset L of a monoid M is called *recognizable*, if it can be recognized by a morphism with finite codomain. The class of all recognizable subses of M is denoted by $\mathbb{R}\text{EC}(M)$. ◁

Notice that every recognizable subeset L of M can be fully recognized by a homomorphism with finite codomain, since ***mon*** admits ⟨regular epi, mono⟩-factorizations.

Since pullback-squares compose, we immediately have

0.3.03 Proposition. *Recognizable sets are closed under inverse images.* □

Recognition by a given morphism allows various equivalent formulations:

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

(a) L *is recognized by* φ .

(b) L *is union of equivalence classes of the concruence* $\sim_\varphi$ *induced by* φ , i.e., *the kernel pair of* φ .

(c) $(\varphi^{-1} \circ \varphi_\exists)[L] = L$, i.e., L *is a fixed point of the closure-operator on* $\mathbb{P}(M)$ *induced by* φ , *or an EM-Algebra for the monad on* $\mathbb{P}(M)$ *induced by* φ .

Proof.

(a) ⇒ (b): Let L be the φ-pre-image of $P \subseteq N$. If $x \sim_\varphi y$, then $\varphi(x) = \varphi(y)$, and hence x and y either both belong to the L , or both lie outside of L . Hence every equivalene 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) ⇒ (c): Consider $y \in (\varphi^{-1} \circ \varphi_\exists)[L]$. There exists $x \in L$ with $\varphi(x) = \varphi(y)$, and therefore $x \sim_\varphi y$. But this implies $y \in L$.

(c) ⇒ (a): Trivial. □

0.3.05 Proposition. *For every monoid* M *the set* $\mathbb{R}\text{EC}(M)$ *forms a Boolean subalgebra of* $\mathbb{P}(M)$. *In particular, recognizable sets are closed under finite unions, finite intersections and complementation.*

Proof. Closedness under complementation is an immediate consequence of Proposition 0.3.04.

The nullary intersection M is the inverse image of $\mathbf{1}$ under the unique monoid homomorphism $M \xrightarrow{!} \mathbf{1}$.

For $n > 0$ consider morphisms $M \xrightarrow{\varphi_i} N_i$ with finite codomains that recognize $L_i \subseteq M$ by means of $P_i \subseteq N_i$, $i < n$. The product $N := \prod_{i<n} N_i$ is still finite, and the pullback of $P := \prod_{i<n} P_i \subseteq N$ along $M \xrightarrow{\langle \varphi_i \,:\, i<n \rangle} N$ results in $L := \bigcap\{ L_i : i < n \}$, which therefore is recognizable. Because of Proposition 0.3.05, unions now can be handled by De Morgans laws. □

The subsets of M recognized by a fixed homomorphism $M \xrightarrow{\varphi} N$ are closed under further operations:

0.3.06 Definition. The *pre-residuation* of $L \subseteq M$ with $J \subseteq M$, is given by

$$J \backslash L := \{ m \in M : J \cdot m \subseteq L \}$$

This has to be distinguished from the *left quotient* of L with respect to K

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

DUAL NOTIONS: the *post-residucation* L/K; the *right quotient* LK^{-1}. ◁

0.3.07 Remark. The notation for residuations and quotients are confusing in the litrature, so one has to be very careful not to confuse these notions.

In fact, for singleton K both concepts agree, for $j \in M$

$$j\backslash L = j^{-1}L$$

where we have dropped the set-brackets from the singletons. This base case yields residuations by intersection, while quotients are obtained by unions:

$$\begin{aligned} J\backslash L &= \bigcap \{\, j\backslash L \;:\; j \in J \,\} \\ J^{-1}L &= \bigcup \{\, j^{-1}L \;:\; j \in J \,\} \end{aligned}$$

Conceptually, however, both notions are very different: in the case of groups quotions can indeed be considered as multiplications with $J^{-1} = \{\, j^{-1} \;:\; j \in J \,\}$ from the left, respectively, with K^{-1} from the right, since every $x \in L$ can be written as $jj^{-1}x = x = xk^{-1}k$, so removing j as a first/ k as a last factor leaves j^{-1} as first/ k^{-1} as a last factor that is multiplied with x.

On the other hand, pre-residuation with J is right-adjoint to pre-multiplication with J in the categorical sense, and similarly with post-residuation with K; this is part of the definition of a unital quantale of which power-monoids are prime examples. For subsets J, K, and L of M the following are equivalent:

$$K \subseteq J\backslash L \quad \text{and} \quad J \cdot K \subseteq L \quad \text{and} \quad J \subseteq L/K \tag{0.3-01}$$

In particular, fixing the set L gives an adjunction reminiscent of a *polarity*, see Diagram (4.5-05)

$$\mathbb{P}(M) \underset{-\backslash L}{\overset{L/-}{\rightleftarrows}} \mathbb{P}(M)^{\mathrm{op}} \qquad (\top) \tag{0.3-02}$$

Therefore both $-\backslash L$ and $L/-$ map unions (= colimits in $\mathbb{P}(M)$ = limits in $\mathbb{P}(M)^{\mathrm{op}}$) to intersections (= colimits in $\mathbb{P}(M)^{\mathrm{op}}$ = limits in $\mathbb{P}(M)$).

In the case of quotients, both $-^{-1}L$ and $L-^{-1}$ are order-preseving functions on $\mathbb{P}(M)$. ◁

0.3.08 Proposition. *If $L \subseteq M$ is recognized by $M \xrightarrow{\varphi} N$, so are all residuations of L and all quotients of L.*

Proof. Consider $J \subseteq M$. We first show $J \backslash L = \varphi^{-1}((\varphi_\exists[J]) \backslash (\varphi_\exists[L]))$.

$$\begin{aligned} m \in \varphi^{-1}((\varphi_\exists[J]) \backslash (\varphi_\exists[L])) &\iff \varphi(m) \in (\varphi_\exists[J]) \backslash (\varphi_\exists[L]) \\ &\iff \varphi_\exists[J] \cdot (\varphi(m)) \subseteq \varphi_\exists[L] \\ &\iff \varphi_\exists[(J \cdot m)] \subseteq \varphi_\exists[L] \\ &\iff J \cdot m \subseteq \varphi^{-1}\varphi_\exists[(J \cdot m)] \subseteq L \\ &\iff m \in J \backslash L \end{aligned}$$

(Theorem 4.5.08 only seems to imply " $\supseteq$ ".)

Next we claim $J^{-1}L = \varphi^{-1}\big((\varphi_\exists[J])^{-1}(\varphi_\exists[L])\big)$.

$$\begin{aligned} m \in \varphi^{-1}\big((\varphi_\exists[J])^{-1}(\varphi_\exists[L])\big) &\iff \varphi(m) \in (\varphi_\exists[J])^{-1}(\varphi_\exists[L]) \\ &\iff \varphi_\exists[J] \cdot \{\varphi(m)\} \cap \varphi_\exists[L] \neq \emptyset \\ &\iff \varphi_\exists[(J \cdot m \cap L)] \neq \emptyset \\ &\iff J \cdot m \cap L \neq \emptyset \\ &\iff m \in J^{-1}L \end{aligned}$$

The proofs for the post-residuation, respectively, the right quotient are analogous. □

0.3.09 Corollary. *The Boolean algebras $\mathbb{R}\text{EC}(M)$ are closed under arbitrary residuations and quotients.*

The following rules may enable us to simplify later calculations.

0.3.10 Proposition. *([?, Proposition IV.2.6]) For $M \xrightarrow{\varphi} N$ the direct image function $\mathbb{P}(M) \xrightarrow{\mathbb{P}(h)} \mathbb{P}(N)$ preserves intersections with and relative complements of subsets recognized by φ.*

Proof. Consider $L, R \subseteq M$ with $L = \big(\varphi^{-1} \circ \varphi_\exists\big)[L]$. Clearly $\varphi_\exists[R \cap L] \subseteq \varphi_\exists[R] \cap \varphi_\exists[L]$. Given $y \in \varphi_\exists[R] \cap \varphi_\exists[L]$, there exists $x \in R$ with $\varphi(x) = y$. But $y \in \varphi_\exists[L]$ implies $x \in R \cap L$, hence $y = \varphi(x) \in \varphi_\exists[(R \cap L)]$.

On the other hand, $R - L = R \cap (M - L)$, hence by Proposition 0.3.05 and the first part we get $\varphi_\exists[(R - L)] = \varphi_\exists[R] \cap \varphi_\exists[(M - L)] = \varphi_\exists[R] \cap (N - \varphi_\exists[L]) = \varphi_\exists[R] - \varphi_\exists[L]$. □

0.3.11 Corollary. *([?, Corollary IV.2.7]) If $M \xrightarrow{\varphi} N$ recognizes a relative complement $L = X_0 - X_1 \subseteq M$ with $X_1 \subseteq X_0$, then $\varphi_\exists[(X_0 - X_1)] = \varphi_\exists[X_o] - \varphi_\exists[X_2]$.*

Proof. Since $X_1 \subseteq X_0$ the condition $L = X_0 - X_1$ is equivalent to $X_1 = X_0 - L$, Proposition 0.3.10 yields $\varphi_\exists[X_1] = \varphi_\exists[(X_0 - L)] = \varphi_\exists[X_0] - \varphi_\exists[L]$, and consequently, $\varphi_\exists[L] = \varphi_\exists[X_0] - \varphi_\exists[X_1]$. □

Counter-examples

The following notion seems to be useful when constructing counter-examples.

0.3.12 Definition. Call a subset G of a monoid $\langle M, \cdot, e\rangle$ a *local group* of M, if the binary multiplication $\cdot$ induces a group structure on G, with a *local neutral element* $e_G \in G$.

Notice that e_G can differ from e, *i.e.*, G does not have to be a submonoid. However, G is of course closed under composition. The following result is immediate:

0.3.13 Lemma. *Local groups are preserved by monoid homomorphisms, and these induce group-homomorphisms from the local group to its image.* □

0.3.14 Example. It is well-known that one can add a new neutral element to every semi-group $\boldsymbol{S} = \langle S, \cdot\rangle$, regardless whether S already has a neutral element (this is *not* the construction of s^1 from [?]). Call the new semigroup ES.

On the category ***srg*** of semi-groups this yields a pointed endo-functor $\boldsymbol{sgr} \stackrel{\iota}{\Longrightarrow} E$ via the inclusions of the original semi-groups into the extended ones.

On the category ***mon***, however, this yields a co-pointed endo-functor $E \stackrel{\sigma}{\Longrightarrow} \boldsymbol{mon}$, where σ_M collapses the original and the new neutral element to the original one. For any group G then G becomes a local group in EG.

0.3.15 Example. Finite non-empty subsets of monoids need not be recognizable: Consider an arbitray monoid homomorphism $\mathbb{Z} \xrightarrow{\varphi} N$ with finite codomain. Since $\mathbb{Z}$ is a local group the image $N' := \varphi_{\exists}[\mathbb{Z}]$ is a local group in N, in this case even a sub-monoid, and the co-restriction of φ to N' is a group homomorphism. The cosets of the induced congruence on $\mathbb{Z}$ are all isomorphic to the kernel $U = \varphi^{-1}[e]$, which is a *normal subgroup*: they all have the form $n + \mathbb{Z}$ for some $n \in \mathbb{Z}$. As $\mathbb{Z}$ is partitioned by finitely many isomorphic co-sets, they all have to be infinite. Therefore no nonempty finite subset of $\mathbb{Z}$ is recognizable.

0.3.16 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 $f(a) = 1$ and $f(b) = -1$. By Kleene's Theorem every finite subset of $\{a, b\}^*$ is recognizable, but by Example 0.3.15 no φ-image of a non-empty finite subset has this property.

0.3.17 Example. (Shmuel Winograd) Recognizable sets need not be closed under the binary operation and under Kleene star. Extend the addition of $\mathbb{Z}$ to $\mathbb{Z} + \{e, a\}$ as follows: a behaves almost like a second copy of $0 \in \mathbb{Z}$:

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

while e becomes a new neutral element (observe that 0 and a cannot be neutral anymore). Now $\mathbb{Z} \subseteq M$ is a local group that fails to be a submonoid.

Claim: every $L \in \mathbb{R}\text{EC}(M)$ satisfies $L \cap \mathbb{Z} \in \mathbb{R}\text{EC}(\mathbb{Z})$.

Let N be a finite monoid and L be the inverse image of $P \subseteq N$ along $M \xrightarrow{\varphi} N$. Then $L \cap \mathbb{Z}$ is the inverse image of $P \cap \varphi_{\exists}[\mathbb{Z}]$ and hence 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.

0.3.18 Open Problem. Given a morphism $M \xrightarrow{\varphi} N$, is there a Boolean algebra morphism induced by φ linking $\mathbb{R}\text{EC}(M)$ and $\mathbb{R}\text{EC}(N)$? The direct image map $\varphi_{\exists}$ clearly does not work.

The interplay of $\mathbb{R}\text{AT}$ and $\mathbb{R}\text{EC}$

0.3.19 Theorem. *The following are equivalent for any monoid M:*

(a) *M is finitely generated.*

(b) $\mathbb{R}\text{EC}(M) \subseteq \mathbb{R}\text{AT}(M)$.

(c) $M \in \mathbb{R}\text{AT}(M)$.

Proof.

(a) ⇒ (b): By hypothesis, there exists a regular epi $A^* \xrightarrow{\pi} M$. If $L \subseteq M$ is recognizable, so is $\pi^{-1}[L]$ (see Theorem 0.3.03), which by Kleene's Theorem is rational as well. But then so is $\pi_{\exists} \circ \pi^{-1}[L]$. Since π is surjective, this set coincides with L.

(b) ⇒ (c): Clear, since $M \in \mathbb{R}\text{EC}(M)$.

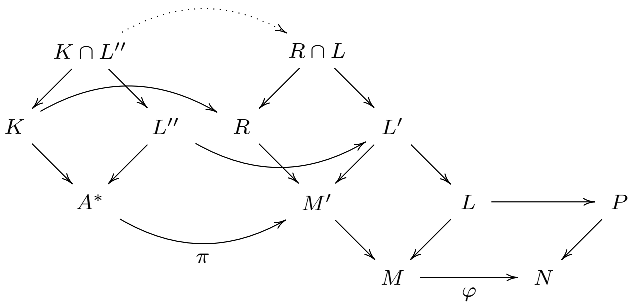
(c) ⇒ (a): Clear because of Proposition 0.2.04. □

0.3.20 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 $A^* \xrightarrow{\pi} M'$ with a finite alphabet A. By Proposition 0.2.06 there exists a rational and hence recognizable set $K \subseteq A^*$ with $\pi_{\exists}(K) = R$ (K need not be the pullback of R along π).

Let $L \subseteq M$ be recognizable, say, the pullback of $P \subseteq N$ along $M \xrightarrow{\varphi} N$ where N is finite. Then $L' := M' \cap L$ is recognizable in M', and $L'' := \pi^{-1}(L')$ is recognizable in A^*. Hence

$K \cap L''$ is recognizable and therfore rational in A^*.



Using Proposition 0.3.10 and the surjectivity of π we obtain

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

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

This result together with Example 0.2.02 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.

0.3.21 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 $\mathbb{R}\text{EC}((M_0 \times M_1))$ is closed under finite unions.

Conversely, suppose $M_0 \times M_1 \xrightarrow{\varphi} N$ recognizes $L \subseteq M_0 \times M_1$. Compose φ with the canonical left inverses $M_i \xrightarrow{\sigma_i} M_0 \times M_1$ to the projections (introduced in the proof of Theorem 0.2.03) 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 φ:

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

Therefore we can pull $P \subseteq N$ back along φ in two stages, first along $N \times N \xrightarrow{\cdot} N$, which results in $Q \subseteq N \times N$, and then along β, which yileds 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

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

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

1 The syntactic monoid

In the categorical approach to congruences 4.7, specifically Definition 4.7.05 and Theorems 4.7.06 and 4.7.08, we have seen precisely how congruences and quotients come about in categories like ***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 0.3.04, 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.

1.0.00 Lemma. *The set of congruences on a monoid forms a compete lattice.*

Proof. The subalgebra-property is preserved by arbitrary intersections of subalgebras, in particular those on $M \times M$. □

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

Recall that a congruence $\sim$ on M admit two equivalent descriptions (*cf.* HW): it is an equivalence relation on M that in addition

- ▷ is a sub-algebra of $M \times M$:

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

- ▷ satisfies

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

Both conditions may be referred to as *stability* of the equivalence relation.

Without reference to elements these conditions can be reformulated as

$$\sim \cdot \sim \;\subseteq\; \sim \qquad \text{resp.} \qquad \Delta_M \cdot \sim \cdot \, \Delta_M \subseteq \sim$$

In fact, since $\langle e, e\rangle \in \sim$ we even have equality in both cases. The second condition resembles the definition of an order-ideal (see Definition 4.12.00), with the order being discrete (Δ_M) and the relation product replaced by the multiplication of subsets of $M \times M$ (however, order ideals do not have to be equivalence relations). Moreover, it suggestes how to define congruences for ordered monoids below.

While the first version of the stability condition directly generalizes to any type of set-based algebra by extending the given condition of all operations, the second version is rather specific for monoids and heavily relies on the neutral element of a monoid. To give a formally similar definition of congruence for a semi-group S, the second quantification has to extend over S^1, defined as S, if S is a monoid, and $S+\{1\}$, if S lacks a neutral element. It is presently (2017-12-16) not clear, what such a categorcally questionalble definition is good for.

It should be clear that every congruence on a monoid M has the form $\Delta_M \cdot E \cdot \Delta_M$ for some equivaence relation E, since reflexivity, symmetry and transitivy are preserved by this construction, while stabilty is automatic.

The main advantage of the second formulation of stability in case of monoids seems to be that it allows the definition of a congruence based on membership in L.

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

$$u \sim_L v \quad \text{iff} \quad \forall s,t \in M.\, s \cdot u \cdot t \in L \iff s \cdot v \cdot t \in L$$

The factor monoid $\boldsymbol{SM}(L) := M/\sim_L$ is known as the *syntactic monoid* of L, while the canoncal surjection $M \xrightarrow{\eta_L} M/\sim_L$ is called the *syntactic quotient*.

1.0.02 Proposition. $\sim_L$ *is in fact a congruence.*

Proof. Reflexivity, symmetry and transitivity are easy. Hence $\Delta_M \cdot \sim_L \cdot \Delta_M$ is a congruence, which of course contains $\sim_L$. It remains to show the other inlusion, *i.e.*, that $u \sim_L v$ and $x, y \in M$ implies $x \cdot u \cdot y \sim_L x \cdot v \cdot y$. But by construction we have

$$\begin{aligned} \forall s,t \in M.\, s \cdot (x \cdot u \cdot y) \cdot t \in L &\iff (s \cdot x) \cdot u \cdot (y \cdot t) \in L \\ &\iff (s \cdot x) \cdot v \cdot (y \cdot t) \in L \\ &\iff s \cdot (x \cdot v \cdot y) \cdot t \in L \end{aligned}$$

□

[Is there no better way of showing stability?]

1.0.03 Remark. Observe that $\sim_L$ is also the largest congruence contained in the equivalence relation

$$L \times L + (M - L) \times (M - L)$$

Clearly, for a congruence to be containied in here is equivalent to L (and $M - L$) being exhausted by congruence classes. Unfortunately, this does not seem to give a handle on computing $\sim_L$, since pre- and post-multiplying the equivalence relation above with Δ_M enlarges the relation, but we need to shrink it.

It remains to establish the universal Property of $\sim_L$, or equivalently η_L.

1.0.04 Proposition. *Any quotient of M that recognizes $L \in M$ factors through η_L. In other words, $\sim_L$ is really the largest (w.r.t. $\subseteq$) congruence such that η_L recognizes L.*

Proof. Remark 1.0.03 shows that L (as well as $M - L$) is the union of $\sim_L$-classes, hence by Proposition 0.3.04 η_L recognizes L.

Suppose $M \xrightarrow{\varphi} N$ is surjective and also recognizes L. Its kernel pair is a congruence, hence satisfies

$$\varphi(u) = \varphi(v) \quad \text{implies} \quad \forall x, y \in M. \varphi(x \cdot u \cdot y) = \varphi(x \cdot v \cdot y)$$

which by hypothesis implies

$$\forall x, y \in M.\ x \cdot u \cdot y \in L \Longleftrightarrow x \cdot v \cdot y \in L$$

Hence every φ-class is contained in a $\sim_L$-class. This gives the desired factorization of η_L. By construction, $M/\sim_L$ 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 φ. The e-image of L is essentially the same as the φ-image of L and has the same syntactic monoid as $\varphi_\exists[L]$.

The following Proposition is awkward as it deliberately avoids morphisms in favour 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 restict the concept of a monoid "dividing" another to free dividends. Perhaps a more sugestive name might be *partial quotient*.

1.0.05 Proposition.

(0) *A monoid N recognizes $L \subseteq M$ iff $\boldsymbol{SM}(L)$ is the quotient of some submonoid of N ($\boldsymbol{SM}(L)$ "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 A we have already mentioned that the syntactic monoid $\boldsymbol{SM}(L)$ ought to be the transition monoid of the minimal automaton that recognizes L. Pin provides such a proof.

However, we diverge here from Pin and follow Manfred Kufleitner [**?**] (found too late) who introduces generalized automata over arbitray monoids. Then the result alluded to above will hold in full generality. In addition, some ideas already present in JKs version of "Theoretische Informatik 1" can be developed further.

2 Automata over general monoids

The notion of an automaton over a general monoid M is not too different from the concept of a generalized automaton over A^* for some alphabet A (called $*$NEA in JK's slides for "Theoretische Informatik 1"): transitions can be labeled with arbitrary elements of A^*, as long as there are only finitely many transitions altogether. The latter condition insures that only finitey many elements of A can actually occur in the transitions, a crucial requirement, if one is interested in establishing a correspondence between rational expressions and automata.

If the monoid M is finitely generated, we have a regular epi $A^* \xrightarrow{\pi} M$ for some finite alphabet A and the set $\pi_\exists[A]$ generates M. In case that an M-automaton only has transitions labeled by the generators, *i.e.*, elements of $\pi_\exists[A]$, one arrives at a concept that Kufleitner calls "step-by-step automaton", which is closer to the classical notion of automaton. However, for a general M-automaton, even if a canonical set of generators is known (as for A^*), the possible decompositions of the labels in terms of the generators is irrelevant.

It is important to notice automata over general monoids á priori are non-deterministc in the sense that they employ transition relations rather than transition functions. Only for step-by-step automata the *prefix-property* of the set of generators can be of relevance, *i.e.*, whether or not some generators are prefixes of others.

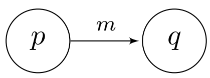
Deterministic automata will be introduced later.

2.0.00 Definition. For any monoid M, a *non-deterministic* M*-automaton* $\mathcal{A} = \langle Q, \delta, I, F\rangle$ consists of

- ▷ a set Q of *states*;
- ▷ a function(!) $M \xrightarrow{\delta} \textbf{\textit{rel}}\langle Q, Q\rangle$ 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 the number of transitions, considered as a subset of $Q \times M \times Q$ is finite.

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



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

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*.

As mentioned before, since $\langle Q,Q\rangle\boldsymbol{rel}$ is an ordered monoid, the canonical notion of morphism is that of a lax homorphism, rather than a monoid homomorphisms. In particular, the function δ ought to introduce a lax homomorphism from M to $\boldsymbol{rel}\langle Q,Q\rangle$. In fact, the *saturation* $\tilde{\delta}$ of the function δ is the lax homomorphism given by

$$\begin{array}{ccccccc}
\mathbb{P}(M) & \xrightarrow{\xi^{-1}} & \mathbb{P}(M^*) & \xrightarrow{\mathbb{P}(\delta^*)} & \mathbb{P}(\langle Q,Q\rangle\boldsymbol{rel}^*) & \xrightarrow{\mathbb{P}(\xi)} & \mathbb{P}(\langle Q,Q\rangle\boldsymbol{rel}) \\
\uparrow{\scriptstyle \eta_M} & & & & & & \downarrow{\scriptstyle \bigsqcup} \\
M & & & \xrightarrow[\tilde{\delta}]{} & & & \langle Q,Q\rangle\boldsymbol{rel}
\end{array} \tag{2.0-00}$$

where ξ denotes the the structure morphisms of the given monoids as EM-algebras and the inverse image map $\mathbb{P}(M) \xrightarrow{xi^{-1}} \mathbb{P}(M^*)$ is the only lax homomorphism (*cf.*, Proposition 0.2.08), while all other components are strict.

Concretely this means: the relation $\tilde{\delta}(w) \subseteq Q\times Q$ is the union of all compostite relations $\delta(w_0);\delta(w_1);\ldots;\delta(w_{n-1})$, where $w_0\cdot w_1\ldots w_{n-1}$ runs through all possible decompositions of $w\in M$.

Now the notion of acceptance becomes easy:

2.0.01 Definition. An M-automaton $\mathcal{A}=\langle Q,\delta,I,F\rangle$ *accepts* $w\in M$ iff $\mathbf{1} \xrightarrow{I} Q \xrightarrow{\tilde{\delta}(w)} Q \xrightarrow{F} 1$ is not empty.

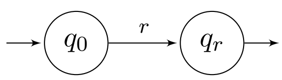
2.0.02 Remark. In view of this somewhat complicated notion of acceptance why not require the function $M \xrightarrow{\delta} \boldsymbol{rel}\langle q,q\rangle$ in Definition 2.0.09 to be a lax homomorphism to start with? In that case we need to introduce the notin of a "finitely generated lax homomorphisms" in order to handle the notion of finiteness, since saturation does not preserve finiteness. Hence one cannot get around introducing at least some notion of saturation.

2.0.1 Rational sets are accepted by finite nondeterministic automata

2.0.03 Proposition. *For any monoid M any finite subset $R\subseteq M$ is accepted by a finite nondeterministic 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 cleat that this automaton is finite and accepts R. □

2.0.04 Proposition. *If two sets* $L_i \subseteq M$, $i < 2$, *are accepted by nondeterministic 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. □

2.0.05 Proposition. *If two sets* $L_i \subseteq M$, $i < 2$, *are accepted by nondeterministic 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,

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

2.0.06 Proposition. *If the set* $L \subseteq M$ *is accepted by a finite nondeterministic 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 2.0.04 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* $\mathcal{A}^+$, where in addition all final states are inked with all initial states by means of e-transitions

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

is finite and accepts L^+ . □

2.0.07 Remark. Kufleitner insists on single initial states even for non-deterministic automata, and requires that initial states have no incoming arrows, while final states have no outgoing arrows. These requirements may be technically convenient, but it is not clear that they are absolutely neccessary.

2.0.08 Theorem. *The language accepted by a finite automaton is rational.*

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

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

- ▷ a set Q of *states*;
- ▷ a monoid homomorphism(!) $M \xrightarrow{\delta} \boldsymbol{set}\langle Q, Q\rangle$ assigning *transition functions* on Q to the elements of M;
- ▷ an *initial state* $q_0 \in Q$ and a set of *final states* $F \subseteq Q$.

$\mathcal{A}$ is called *finite*, if the set Q is finite.

2.0.10 Remark.

- ▷ While in general deterministic M-automate are subsumed by non-deterministic M-automata, this is no longer true in the finite case: if Q is finite, and M is infinite, in the deterministic case the set of transitions as a subset of $Q \times M \times Q$ cannot be finite. On the other hand, the monoid $\boldsymbol{set}\langle Q, Q\rangle$ is finite. In particular, many of the transition functions $\delta(m)$, $m \in M$, have to agree, we just do not know which ones do. A similar argument holds for step-by-step automata, if $A^* \xrightarrow{\pi} M$ is a regular epi. There is no need for A to be finite, even if Q is. Only if both A and Q are finte, finite deterministic step-by-step automata are subsumed by finite non-deterministic step-by-step automata.
- ▷ Using a mere function $M \xrightarrow{\delta} \boldsymbol{set}\langle Q, Q\rangle$ instead of a monoid morphism will always produce a monoid homomorphism from M^* to $\boldsymbol{set}\langle Q, Q\rangle$, but not neccessairy a monoid homomorphism from M to $\boldsymbol{set}\langle Q, Q\rangle$. Since $\boldsymbol{set}\langle Q, Q\rangle$ is only discretely ordered and fails to be a complete lattice, the construction of Diagram (2.0-00) is not available. Whether or not an element $m \in M$ is accepted by an automaton that utilizes just a function δ amounts to the existence of a ξ-preimage $\omega \in M^*$ of m that is accepted by the automaton, not a very deterministic notion.

2.0.11 Theorem. *Every recogngnizable set L of a monoid M is accepted by a deterministic M-automaton, and vice versa.*

Proof. Suppose $L \subseteq M$ is the pre-image of $P \subseteq N$ under the homomorphism $M \xrightarrow{\varphi} N$. Construct an automaton $\mathcal{A}$ by setting

- ▷ $Q := N$;
- ▷ $\delta(m) := (-) \cdot \varphi(m)$, $m \in M$;
- ▷ $q_0 := e_N$ and $F := P$.

Notice that δ is indeed a monoid homomorphism, as required. By construction, $\mathcal{A}$ accepts $m \in M$ iff $q_0 \cdot \varphi(m) \in P$.

Conversely, if $\mathcal{A}$ accepts $L \subseteq M$, consider the δ-image $N := \delta_\exists(M) \subseteq \boldsymbol{set}\langle Q, Q\rangle$ of M and the subset $P := \{ Q \xrightarrow{\psi} Q : \psi \in N \wedge \psi(q_0) \in F \}$. By construction, L is the pre-image of P .

Observe that in both cases $\mathcal{A}$ is finite iff N is finite. □

3 Green's Relations, not just on monoids but on 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 be useful.

In particular, one is interested in the divisibility relations satisfied in N , including the idempotents. An important tool in the analysis are the so-called "Green's relations", a set of five equivalence relations definable in any monoid. Three of these express the fact that two elements are mutually postfixes ($\mathcal{R}$), prefixes ($\mathcal{L}$) or infixes ($\mathcal{J}$) of each other, while the other two relations $\mathcal{H}$ and $\mathcal{D}$ form the meet (or infimum) and join (or supremum) of $\mathcal{R}$ and $\mathcal{L}$ in the complete lattice of equivalence relations.

Traditionally, the equivalence relatins $\mathcal{L}$, $\mathcal{R}$ and $\mathcal{J}$ are derived from pre-orders $\leq_L$, $\leq_R$ and $\leq_J$ that express the fact that the second(!) argumment is a postfix/prefix/infix of the first one. And these pre-orders are expressed in terms of set-inclusions of suitable principal left/right/2-sided ideals. In our view this approach to Green's relations is unnecessarily complicated and obscures the simple fact that these notions make sense in every category, not just in monoids. This is useful when working with *typed alphabets*, or finite graphs, as already indicated in the context of M-automata and *labeled transition systems* after Definition 2.0.09.

For the sake of completness we spell out the ideal-theoretic approach to Green's relations.

3.0.00 Definition. A subset $A \subseteq M$ of a monoid is called a *left ideal*, if $M \cdot A \subseteq A$.

Dual Notion: *right ideal*.

A *2-sided ideal*, or just *ideal*, is a subset $A \subseteq M$ that is both a left and a right (monoid) ideal, and hence satisfiess $M \cdot A \subseteq A \supseteq A \cdot M$.

These notions correspond precisely to the categorical notions of left/right/2-sided ideals mentioned in Examples 4.9.03 and 4.10.01

Since we are interested in monoids rather than semigroups, the inclusions above may be replaced by equalities.

3.0.01 Remark.

- ▷ The notion of ideal makes sense in any category of monoids in a monoidal category (not necessarily symmetric)

  Ideals in monoids have to be distinguished from ideals in rings: the latter in addition have to be subgroups with respect to the addition. This is consistent with the observation that rings are monoids in the category ***ab*** of abelian groups. Notice that $\mathbb{Z}$ as a ring is a principal ideal domain, *i.e.*, all ring ideals are principal ideals of the form $a \cdot \mathbb{Z}$ for some $a \in \mathbb{Z}$. Viewing $\mathbb{Z}$ as just a monoid with respect to multiplications, there exist further ideals, *e.g.*, the sets $I_{n,k} := \{0\} \cup \{ a \in n \cdot \mathbb{Z} : |a| > k \}$. Unless explicitly stated, in these notes "ideals" will refer to "ideals in monoids".

- ▷ Some authors insist on ideals not being empty. But in that case infinite intersections of ideals can fail to be ideals: in $\langle \mathbb{Z}, 0, + \rangle$ consider the ideals $J_k := \{ a \in \mathbb{Z} : |a| > k \}$, $k \in \mathbb{N}$. Their intersection obviously is empty. Moreover, it would exclude the profunctor specified by the empty sub-functor of $M^{\mathrm{op}} \times M \xrightarrow{hom} \boldsymbol{set}$ from being an ideal.

3.0.02 Remark. The left/right/2-sided ideals of a monoid M, respectively, form a subset of $\mathbb{P}(M)$ that is closed under composition, but unless $|M| = 1$ does not contain the unit $\{e\}$ of $\mathbb{P}(M)$. Instead, M is a left/right/2-sided ideal and acts as a neutral element of the corrsponding monoids, which, however, are not sub-monoids of $\mathbb{P}(M)$.

There is no categorical reason that left/right ideals should compose, this opertion does not type-check, but rather is a spurious consequence of the fact that monoids have only one object. On the other hand, the composition of 2-sided ideals does make sense, as does the composition of a left with a right ideal, which results in a 2-sided ideal.

3.0.03 Lemma. *Given a monoid M, the left/right/2-sided ideal* generated *by $B \subseteq M$ is given by $M \cdot B \,/\, B \cdot M \,/\, M \cdot B \cdot M$, respectively. In case of a singleton set B we call such ideals* principal ideals.

For 2-sided ideals we obtain an oplax monoid homomorphism $\mathbb{P}(M) \longrightarrow \boldsymbol{prof}\langle M, M \rangle$

Proof. Clearly, $M \cdot M \cdot B = M \cdot B$, hence $M \cdot B$ is indeed a left ideal. For right idels and 2-sided ideals the argument is similar.

Composing the 2-sided ideals $M \cdot A \cdot M$ and $M \cdot B \cdot M$ results in the 2-sided ideal generated by $A \cdot M \cdot B$, which in general differs from the 2-sided ideal generated by $A \cdot B$, unless A

or B itself is a 2-sided ideal. Now $A \cdot B \subseteq A \cdot M \cdot B$ implies the corresponding inclusion for the generated 2-sided ideals. Moreover, the unit $\{e\}$ of $\mathbb{P}(M)$ is mapped to the unit M of $\boldsymbol{prof}\langle M, M\rangle$, therefore the assignment $A \mapsto M \cdot A \cdot M$ is oplax (even normalized oplax, as the units are preserved on the nose). □

3.0.04 Remark. Due to the fact that monoids have only one object, the composition of the left ideal $M \xrightarrow{M \cdot A} \mathbf{1}$ with the right ideal $\mathbf{1} \xrightarrow{B \cdot M} M$ results in the 2-sided ideal $M \cdot A \cdot B \cdot M$. In general this differs from the composition of the 2-sided ideals generated by A, respectively, B, which happens to be $M \cdot A \cdot M \cdot B \cdot M$. All of this specializes to the principal case as well.

3.0.05 Definition. (traditional) On a monoid M one defines three pre-order relations (not necessarily anti-symmetric) and three corresponding equivalences, the first three *Green's relations* as follows:

- $a \leq_R c :\iff a \cdot M \subseteq c \cdot M$ and $a\mathfrak{R}c :\iff a \leq_R c \wedge c \leq_R a$;
- $c \leq_L b :\iff M \cdot c \subseteq M \cdot b$ and $c\mathfrak{L}b :\iff c \leq_L b \wedge b \leq_L c$;
- $a \leq_J b :\iff M \cdot a \cdot M \subseteq M \cdot b \cdot M$ and $aJb :\iff a \leq_J b \wedge b \leq_J a$.

As the lattice of equivalence relations on a given set is always complete, we can also form the infimum $H := L \sqcap R = L \cap R$ and the supremum $D := L \sqcup R$; these are the remaining two Green's relations.

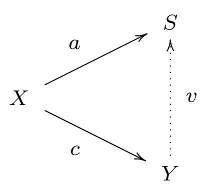
A more direct description of the order relations defined above addresses the notion of divisibility:

$$\begin{array}{rll} a \leq_R c & \text{iff} \quad \exists u.\, a = c \cdot v & \textit{i.e.}, c \text{ is a prefix of } a \\ c \leq_L b & \text{iff} \quad \exists v.\, c = u \cdot b & \textit{i.e.}, b \text{ is a postfix of } c \\ a \leq_J b & \text{iff} \quad \exists u, v.\, a = u \cdot b \cdot v & \textit{i.e.}, b \text{ in an infix of } a \end{array}$$

This suggests the categorical version of these relations below:

3.0.06 Definition. (categorical) Given a category $\mathscr{C}$, we define three pre-orders and three corresponding equivalence relations on the class of $\mathscr{C}$-morphisms:

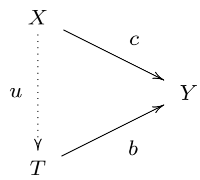
- We set $a \leq_R c$ iff a and c have the same domain X and the hom-set $[a, c]$ in the comma-category $X/\mathscr{C}$ is not empty; *i.e.*, if c is a first factor of a:



(3.0-00)

Hence $a\mathfrak{R}c$ means that a and c have a common domain and are mutually first factors of each other.

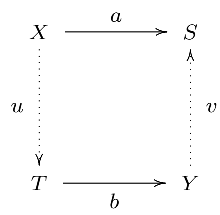
▷ If c and b have the same cocomain Y, we set $b \leq_L c$ iff the the hom-set $[c, b]$ in the comma-category $\mathscr{C}/Y$ is not empty; *i.e.*, if c is a last factor of b:



(3.0-01)

Hence cLB means that c and b have a common codomain and mutually are last factors of each other.

▷ Without constraint on $X \xrightarrow{a} S$ and $T \xrightarrow{b} Y$ we set $a \leq_J b$ if b factors through a; i.e., there exist arrows $X \xrightarrow{u} T$ and $S \xrightarrow{v} Y$ with $a = v \circ c \circ u$:



(3.0-02)

Hence aJb means that a and b mutually factor through each other.

▷ As before, $\mathcal{H}$ and $\mathcal{D}$ are defined as the infimum, resp., supremum of $\mathcal{L}$ and $\mathcal{R}$. Observe that only parallel arrows can be in relation $\mathcal{H}$.

3.0.07 Remark.

▷ In both cases the relations $\leq_R$, $\leq_L$ and $\leq_J$ are clearly pre-orders, hence their symmetrizations $\mathcal{R}$, $\mathcal{L}$ and $\mathcal{J}$, respectively, are equivalence relations.

Alternatively, $\mathcal{R}$, $\mathcal{L}$ and $\mathcal{J}$ also arise als kernel-pairs of the ***set***-functions that map $\mathscr{C}$-arrows to the generated principal ideals in ***prof*** $\langle \mathbf{1}, \mathscr{C} \rangle$ in case of $\mathcal{R}$, in ***prof*** $\langle \mathscr{C}, \mathbf{1} \rangle$ in case of $\mathcal{L}$, and in ***prof*** $\langle \mathscr{C}, \mathscr{C} \rangle$ in case of $\mathcal{J}$. Only the third case the codomain is a monoid, but already in the case that $\mathscr{C} = M$ the assignment $a \mapsto M \cdot a \cdot M$ in view of Lemma 3.0.03 is only a normalized oplax monoid homomorphism from $\mathbb{P}(M)$ to ***prof*** $\langle M, M \rangle$. Whether this suffices to turn $\mathcal{J}$ into a congruence remains to be seen.

▷ In the categorical setting, for $X \xrightarrow{a} S$ the $\mathcal{R}$-class $[a]_{\mathcal{R}}$ and the principal ideal $S/\mathscr{C} \circ a$ determine each other uniquely: if $a\mathcal{R}c$ then $S/\mathscr{C} \circ a = S/\mathscr{C} \circ c$, and if the principal ideals $S/\mathscr{C} \circ a$ and $S/\mathscr{C} \circ c$ coincide, then $a\mathcal{R}c$.

▷ The supremum of two equivalence relations in general is difficult to compute explicitly, similar to the computation of a coequalizer. However, in this special case the relationship of $\mathcal{R}$ and $\mathcal{L}$ to their supremum $\mathcal{D}$ turns out to be rather simple, see Proposition 3.0.11.

3.0.08 Proposition. *Any idempotent arrow $B \xrightarrow{i} B$ of a category $\mathscr{C}$ satisfies: If i is a first factor of $B \xrightarrow{g} C$ (a last factor of $A \xrightarrow{f} B$), then i is a first factor of g (a last factor of f).*

Proof. If $(B \xrightarrow{i} B \xrightarrow{h} C) = (B \xrightarrow{g} C)$, then

$$(B \xrightarrow{i} B \xrightarrow{g} C) = (B \xrightarrow{i} B \xrightarrow{i} B \xrightarrow{h} C) = (B \xrightarrow{i} B \xrightarrow{h} C) = B \xrightarrow{g} C) \qquad \square$$

3.0.09 Proposition. *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 $n > 1$. For every $k < n$ we have $a^{n!} = \left(a^{k+1}\right)^{n!/(k+1)}$. As there are at most n distinct powers of a, there has to be a smallest $k < n$ satisfying $a^{k+1} = \left(a^{k+1}\right)^{n!/(k+1)}$. Since at least one of the exponents $k+1$ and $n!/(k+1)$ is even, we obtain the required idempotent power of a. $\square$

3.0.10 Proposition. *For every category $\leq_J \;=\; \leq_L \,;\, \leq_R \;=\; \leq_R \,;\, \leq_L$.*

Proof. Since combining Diagrams (3.0-00) and (3.0-01) along c yields Diagram (3.0-02), we get $\mathcal{L} \circ \mathcal{R} \subseteq J$.

Conversely, $a \leq_J b$ implies the existence of $\mathscr{C}$-morpgisms u, v such that Diagram (3.0-02) commutes. But then $aR(b \circ u)Lb$ and also $aL(v \circ b)Rb$. $\square$

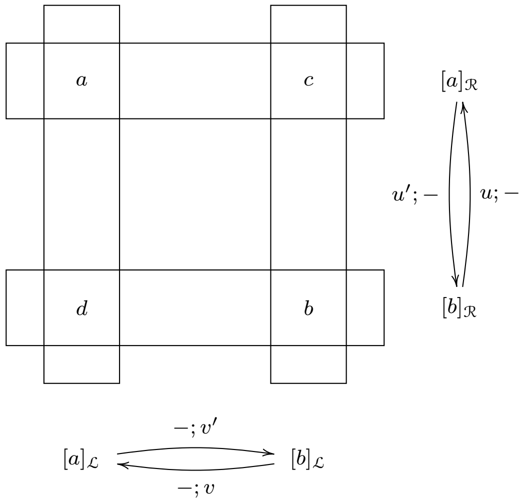
3.0.11 Proposition. *$\mathcal{R} \circ \mathcal{L}$ and $\mathcal{L} \circ \mathcal{R}$ always agree with $D = L \sqcap R$, and for categories where all hom-sets of the form $\mathscr{C}\langle X, X\rangle$ are finite, also agree with $\mathcal{J}$.*

Proof. We first show $\mathcal{L} \circ \mathcal{R} = \mathcal{R} \circ \mathcal{L}$. Suppose $a\mathcal{R}c\mathcal{L}b$, *cf.*. Diagrams (3.0-00) and (3.0-01). By definition, $c \in [a]_{\mathcal{R}} \cap [b]_{\mathcal{L}}$. There exist $\mathscr{C}$-morphisms $X \underset{u'}{\overset{u}{\rightleftarrows}} T$ and $Y \underset{v'}{\overset{v}{\rightleftarrows}} S$ with

$$a = v \circ c \quad , \quad c = a \circ v' \quad \text{and} \quad c = b \circ u \quad , \quad b = c \circ u'$$

Clearly, pre-composition with some morphism preserves $\mathcal{R}$, while post-composition preserves $\mathcal{L}$. Hence $v \circ -$ restrict to a function from $[c]_{\mathcal{L}} = [b]_{\mathcal{L}}$ to $[a]_{\mathcal{L}}$, while $v' \circ -$ maps in the opposite direction. In particular, we set $d := v \circ b \in [a]_{\mathcal{L}}$.

Similarly, $v' \circ -$ restricts to a function from $[a]_{\mathcal{L}}$ to $[c]_{\mathcal{L}} = [b]_{\mathcal{L}}$, while $- \circ u$ and $- \circ u'$ restrict to functions between $[b]_{\mathcal{R}}$ and $[a]_{\mathcal{R}}$. This is illustrated by the following “egg-crate” picture:



(3.0-03)

Claim. The functions $v' \circ -$ and $v \circ -$ are mutual inverses between $[a]_{\mathcal{L}}$ and $[b]_{\mathcal{L}}$, while $- \circ u'$ and $- \circ u$ are mutual inverses between $[a]_{\mathcal{L}}$ and $[b]_{\mathcal{L}}$.

We already know $v \circ v' \circ a = a$. Furthermore

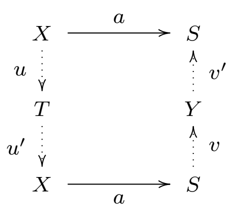
$$v' \circ d = v' \circ v \circ b = v' \circ v \circ c \circ u' = v' \circ a \circ u' = c \circ u' = b$$

Hence for any $y \in [b]_{\mathcal{L}}$ setting $x := v \circ b$ by the same argument yields $v' \circ x = y$. This settles the claim for $v \circ -$ and $v' \circ -$. The argument for $- \circ u$ and $- \circ u'$ is analoguous. This shows $\mathcal{L} \circ \mathcal{R} = \mathcal{R} \circ \mathcal{L}$.

Clearly, due to the reflexivity of $\mathcal{L}$ and $\mathcal{R}$ we have $L, R \subseteq \mathcal{L} \circ \mathcal{R} = \mathcal{R} \circ \mathcal{L}$. Consider any equivalence relation Q with $L, R \subseteq Q$. Transitivity shows $\mathcal{L} \circ \mathcal{R} = \mathcal{R} \circ \mathcal{L} \subseteq Q \circ Q \subseteq Q$. This establishes $\mathcal{L} \circ \mathcal{R} = \mathcal{R} \circ \mathcal{L}$ as the join $\mathcal{D}$ of $\mathcal{L}$ and $\mathcal{R}$.

Finally, consider the case where all hom-sets of the form $\mathscr{C}\langle X, X\rangle$ are finite. By Proposition 3.0.09 every idempotent in such a hom-set has an idempotent power.

By definition we have $L, R \subseteq J$ and hence $\mathcal{L} \circ \mathcal{R} = \mathcal{R} \circ \mathcal{L} \subseteq J$. Conversely, consider aJb, which means $a = v \circ b \circ u$ and $b = v' \circ a \circ u'$ for suitable $\mathscr{C}$-morphisms u, u', v, v'. Substituting a yields



For suitble $N > 0$ the endomorphism $(u \circ u')^N$ is idempotent. Then by Proposition 3.0.08 we have

$$a = a \circ (u \circ u')^N = a \circ u \circ u' \circ (u \circ u')^{N-1}$$

and hence $a \leq_L a \circ u$. But $a \circ u \leq_L a$ is true by default, and so we obtain $aL(a \circ u)$.

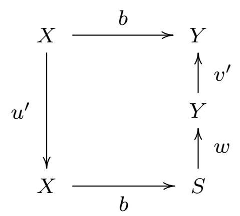
In similar fashion one can use an idempotent power of $v' \circ v$ to obtain $aR(a \cdot v)$. Since $\mathcal{R}$ is stable under pre-composition, this furthermore implies $(a \circ u)R(v \circ a \circ u) = b$, which combined with $aL(a \circ u)$ results in $a(\mathcal{R} \circ \mathcal{L}) = b$, and hence $J \subseteq \mathcal{R} \circ \mathcal{L}$. □

3.0.12 Corollary. *Every $\mathcal{D}$-class is partitioned in $\mathcal{L}$-classes of the same size, and in $\mathcal{R}$-clases of the same size. All $\mathcal{H}$ classes arise as intersections of $\mathcal{L}$-classes with $\mathcal{R}$-classes and are preserved by the isomorpnisms in Diagram (3.0-03).* □

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

$$J \cap \leq_L \subseteq L \qquad \text{and} \qquad J \cap \leq_R \subseteq R$$

Proof. aJb and $a \leq_R b$ due to the symmetry of $\mathcal{J}$ implies the existence of $\mathscr{C}$-morphisms u', v' and w such that $b = v' \circ a \circ u'$ and $a = w \circ b$, *i.e.*,



The endomorphisms u' and $v' \circ w$ have idempotent powers, in particular let $(v' \circ w)^N$ be idempotent with $N > 0$. Using Lemma 3.0.08 we get

$$b = (v' \circ w)^N \circ b = (v' \circ w)^{N=1} \circ v' \circ w \circ b$$

and hence $b \leq_R w \circ b = a$. By hypothesis this yields aRb, as desired. □

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.

3.0.14 Proposition. *In any category aDb impliies $[a]_{\mathcal{L}} \cap [b]_{\mathcal{R}} \neq \emptyset$. In case all endo-hom-sets are finite, one can use the equivalent hypothesis aJb.*

Proof. This is a simple consequence of $D = \mathcal{L} \circ \mathcal{R} = \mathcal{R} \circ \mathcal{L}$ and of $D = J$ in the case of finite endo-hom-sets. □

Green's classical lemma combines the results above in the case of monoids:

3.0.15 Lemma. (Green's Lemma) *Consider a finite monoid* $M = \langle M, \cdot, e \rangle$ *and elements* $a, b \in M$ *satisfying* aJb *(or equivalently,* aDb *).*

- ▷ *If* $a\mathfrak{R}c$ *, and this is realized by elements* $v, v' \in M$ *with* $a \cdot v' = c$ *, resp.* $b \cdot v = a$ *, then the corresponding right-multiplications* $- \cdot v'$ *and* $- \cdot v$ *restrict to inverse bijections between* $[a]_{\mathcal{L}}$ *and* $[b]_{\mathcal{L}}$ *that, moreover, preserve* $\mathcal{H}$*-classes.*
- ▷ *If* cLb *, and this is realized by elements* $u, u' \in M$ *with* $u' \cdot c = b$ *, resp.* $u \cdot b = c$ *, then the corresponding left-multiplications* $u' \cdot -$ *and* $u \cdot -$ *restrict to inverse bijections between* $[c]_{\mathcal{R}}$ *and* $[b]_{\mathcal{R}}$ *that, moreover, preserve* $\mathcal{H}$*-classes.* □

We now turn to idempotent endomorphisms.

3.0.16 Lemma. (Location Lemma (Clifford and Miller)**)** *In Diagram (3.0-03) we have*

$$c = b \circ a \text{ iff } d \text{ is idempotent.}$$

Proof.
($\Longrightarrow$) If $aR(b \circ a)$ for $v' = b$ there exists y with $a = v \circ (b \circ a)c$. Morover, $aR(b \circ a)L(b \circ a)$ implies $aD('8ab)$. Now by Proposition 3.0.11 (resp. Green's Lemma) $b \circ -$ and $v \circ -$ yield inverse bijections between $[a]_{\mathcal{L}}$ and $[b \circ a]_{\mathcal{L}}$. In particular, $d := v \circ b$ satisfies $b \circ d = b \circ v \circ b = b$ and hence $d = (v \circ b)^2$ is idempotent.
($\Longleftarrow$) Let $d = v \circ b \in [a]_{\mathcal{L}} \cap [b]_{\mathcal{R}}$ be idempotent. Since $a = d \circ u'$ and $b = v' \circ d$ by Proposition 3.0.08 $a = d \circ a$ and $b = b \circ d$. Now $b \circ -$ is an $\mathcal{H}$-class-preserving bijection from $[d]_{\mathcal{L}}$ to $[b]_{\mathcal{L}}$ that maps $a \in [d]_{\mathcal{L}}$ to $a \cdot b \in [b]_{\mathcal{L}}$. Similarly, $- \circ a$ is an $\mathcal{H}$-class-preserving bijection from $[d]_{\mathcal{R}} = [b]_{\mathcal{R}}$ to $[a]_{\mathcal{R}}$ that maps $b \in [d]_{\mathcal{R}}$ to $b \circ a \in [a]_{\mathcal{R}}$. □

3.0.17 Theorem. (Green's Theorem) *If* M *is a monoid and* $d \in M$ *is idempotent, then* $[i]_{\mathcal{H}}$ *is a local group in* M *in the sense of Definition 0.3.12. Hence every* $\mathcal{H}$*-class* Ξ *contains at most one idempotent element, and this is the case iff* $\Xi \cap \Xi^2 \neq \emptyset$ *.*

Proof. Let $\mathscr{C}$ have one object and hom-set M . For $a, b \in [d]_{\mathcal{H}}$ observe $[d]_{\mathcal{H}} = [a]_{\mathcal{L}} \cap [b]_{\mathcal{R}} = [a]_{\mathcal{R}} \cap [b]_{\mathcal{L}}$, hence $b \circ a \in [d]_{\mathcal{H}}$. Proposition 3.0.08 guarantees that i is neutral in $[d]_{\mathcal{H}}$.

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

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 $d \in G$ neutral, hence idempotent in M . For each $a \in G$ there exists a "local inverse" a' with $a' \circ a = d = a \circ a'$, hence $d \leq_R a$ and $d \leq_L a$. But since $a \circ d = a = d \circ a$ we also have $a \leq_R d$ and $a \leq_L d$. Therefore dLa and dRa, hence dHa. This shows $G \subseteq [d]_{\mathcal{H}}$. □

Clearly, every local group in M is contained in an $\mathcal{H}$-class, the class of its neutral element, and the $\mathcal{H}$-classes with an idempotent element are precisely the maximal local groups.

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

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

Otherewise, $[d]_{\mathcal{H}}$ is a proper subset of $[d]_{\mathcal{D}}$. For $(X \xrightarrow{c} Y) \in [d]_{\mathcal{D}} - [d]_{\mathcal{H}}$ choose $(X \xrightarrow{a} S) \in [d]_{\mathcal{L}} \cap [c]_{\mathcal{R}}$. There exists $X \xrightarrow{u} S$ with $d \circ u = a$ and $S \xrightarrow{u'} X$ with $a \circ u' = d$. Therefore $u' \circ a \circ u' \circ a = u' \circ i \circ a = u' \circ i \circ d \circ u = u' \circ d \circ u = u' \circ a$ is idempotent on X.

Moreover, there exists $Y \xrightarrow{v} S$ with $v \circ c = a$ and $S \xrightarrow{v'} Y$ with $v' \circ a = c$. Therefore $u' \circ a = u' \circ v \circ c$ implies $u' \circ a \leq_R c$, while $c = v' \circ a = v' \circ d \circ u = v' \circ d \circ d \circ u = v' \circ d \circ a = v' \circ a \circ u' \circ a$ implies $c \leq_R u' \circ a$. This shows $u' \circ a \in [c]_{\mathcal{R}}$.

Similarly, with $(S \xrightarrow{b} Y) := (S \xrightarrow{d} S \xrightarrow{v} Y) \in [d]_{\mathcal{R}} \cap [c]_{\mathcal{L}}$ we obtain another idempotent $b \circ v$ on Y that belongs to $[c]_{\mathcal{L}}$. □

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

Proof. Consider different idempotents $S \xrightarrow{d} S$ and $X \xrightarrow{c} X$ satisfying $d\mathcal{D}c$. For $(S \xrightarrow{b} X) \in [d]_{\mathcal{R}} \cap [c]_{\mathcal{L}}$ we have $b \circ d = b$ and $b \circ u = c$. Hence the composition of $b \circ -$ and $- \circ u$ yields a bijection from $[d]_{\mathcal{H}}$ to $[c]_{\mathcal{H}}$ which maps d to $b \circ d \circ u = b \circ u = ac$.

In analogy to the preceding proof, $u \circ b \in [b]_{\mathcal{R}}$ turns out to be idempotent on S. Hence every $y \in [b]_{\mathcal{R}}$ satisfies $y \circ u \circ b = y$. For $x, y \in [d]_{\mathcal{H}}$ this yields

$$(b \circ x \circ u)(b \circ y \circ u) = b \circ (y \circ u \circ b) \circ x \circ u = b \circ y \circ x \circ u$$

which establishes the desired homomorphism. □

3.0.1 A proof of Schützenberger's result via Green's relations

The following is based on lecture notes of Kumar. [to be transfered from handwritten notes.]

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 langages, 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.

3.0.20 Definition. A language $L \subseteq \Sigma^*$ for finite alphabet Σ is called *star-free*, if it cn be described by a rational expression withour the Kleene star.

3.0.21 Example. Let Σ 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 [???] managed to characterize the star-free languages in terms of their syntactic monoids.

3.0.22 Definition. A monoid M is called *aperiodic*, if for every idenpotent 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.

3.0.23 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.

3.0.24 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 y with idempotent power y^N, $N > 0$. We claim show $y^{N+1} = y^N$. From $y^N = y^N y y^{N-1}$ we infer $y^N \mathcal{J} y^{N+1}$ (as well as similar relations for $\mathcal{L}$, $\mathcal{R}$, and $\mathcal{H}$). The hypothesis $|[y^N]_{\mathcal{J}}| = 1$ then yields $y^{N+1} = y^N$

Now consider a, b in some $\mathcal{H}$-class H. As there exist u, v with $a = b \circ u$ and $b = v \circ a$, we get $a = u^k a v^k$ for each $k > 0$. If y^N is idempotent, the first part shows $a = u^N a v^N = u^N a v^{N+1} = v \circ a = b$. □

4 Appendix: mathematical foundations

4.0 Categories, functors and natural transformations

It is useful to consider "all sets and functions" as well as "all monoids and homomorphisms" and other collections 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 connect the objects and can be assigned a "domain" or "source" and a "codomain" or "target". So far this specifies a directed graph, possibly large, *i.e.*, with a proper class of nodes. Categories arise when the arrows are equipped with a well-behaved composition operation on arrows where target and source match; this has to be associative and have neutral elements or "identities" for each object.

4.0.00 Definition. A *graph*

$$\mathscr{C} \ : \quad C_1 \underset{\partial_1}{\overset{\partial_0}{\rightrightarrows}} C_0$$

consists of (possibly large) sets C_0 of *objects* (or *vertices*, or *1-cells*), and C_1 of *morphisms* (or *edges* or *arrows* or *2-cells*), connected by two functions ∂_0 and ∂_1 assigning the *source* (or *domain*) and *target* (or *codomain*) to the arrows. The notation $A \xrightarrow{f} B$ indicates that $f \in C_1$ satisfies $\partial_0(f) = A$ and $\partial_1(f) = B$.

For these data to form a *category* $\mathscr{C}$, in addition one needs a family of distinguished *identity morphisms* $A \xrightarrow{id_A} A,\ A \in C_o$ and a partial composition on arrows

$$g \circ f \quad \text{is defined iff} \quad \partial_1(f) = \partial_0(g)$$

in diagrammatic form

$$A \xrightarrow{g \circ f} C \qquad \text{for} \qquad A \xrightarrow{f} B \xrightarrow{g} C$$

subject to the following requirements:

- ▷ Composition is comaptible with the source and target functions

$$\partial_0(g \circ f) = \partial_0(f) \quad \text{and} \quad \partial_1(g \circ f) = \partial_1(f)$$

- ▷ composition is associative, *i.e.*, $h \circ (g \circ f) = (h \circ g) \circ f$ provided one side is defined;
- ▷ identity morphisms are neutral with respect to composition, *i.e.*, $\partial_0(id_a) = A = \partial_1(id_A)$

A graph/category is called

- ▷ *small*, if C_o and C_1 are sets;
- ▷ *locally small*, of for any two objects $A, B \in \mathscr{C}$ the arrows $A \xrightarrow{f} B$ form a set.

We will mostly be dealing with locally small categories. ◁

4.0.01 Definition. A *graph morphism* $\mathscr{C} \xrightarrow{F} \mathscr{D}$ from

$$\mathscr{C} : \quad C_1 \underset{\partial_1}{\overset{\partial_0}{\rightrightarrows}} C_0 \qquad \text{to} \qquad \mathscr{D} : \quad D_1 \underset{\partial_1}{\overset{\partial_0}{\rightrightarrows}} D_0$$

consists of two functions $C_i \xrightarrow{F_i} D_i,\ i < 2$, that commute with the domain and codomain functions.

A graph morphims between categories is called a *functor*, if it preserves the identity morphisms and the composition of morphisms.

4.0.02 Examples.

(0) The paradigmatic category is ***set*** with sets as objects and functions as morphisms. ***set*** is a localy 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 4.0.10 below. Often these can be classified as *algebraic* in nature like the categories of

- ▷ ***mon*** monoids and monoid homomorphism;
- ▷ ***grp*** groups and group homomorphism;
- ▷ ***lat*** lattices and lattice homomorphisms;
- ▷ ***bool*** Boolean algebras and Boolean homomorphisms;
- ▷ ***ring*** rings and ring homomorphisms;

or topological in nature, like the categories of

- ▷ ***top*** topological spaces and continuous maps;
- ▷ ***met*** metric spaces and non-expanding maps;
- ▷ ***pre*** pre-ordered sets and order-preserving functions;
- ▷ ***cpo*** complete partial orders and continuous functions.

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

- ▷ ***rel*** has sets as objects, but binary relations as morphisms.
- ▷ ***prt*** has sets as objects, but partial functions as morphisms; clearly ***set*** is contained in ***prt***, which is contained in ***rel***.
- ▷ *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 cateories may be called $\bigsqcup$-***slat***, $\bigsqcap$-***slat*** and ***clat***, respectively.

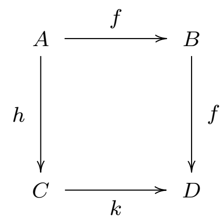
(3) The objects of a category do not need to be structured sets:

- ▷ 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 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-idenity arrow.

4.0.03 Remarks.

- ▷ For every category $\boldsymbol{C}$ and each $\mathscr{C}$-object C the hom-set $\mathscr{C}\langle C, C\rangle$ is automatically a monoid under composition, with id_C as neutral element.
- ▷ Replacing for a given category the non-empty homsets 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 diagrms":



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

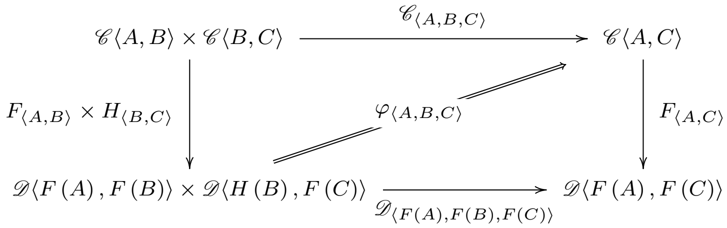
4.0.04 Remark. Alternatively, one can specify a locally small graph/category $\mathscr{C}$ in terms of local so-called hom-sets $\mathscr{C}\langle A, B\rangle$ (or $[A, B]$, when the relevant graph/category is clear), for $A, B \in C_0$, *i.e.*, a set-valued $C_0 \times C_0$-matrix. These hom-sets need not be pairwise disjoint, but we can reconstruct a global class C_1 of morphisms by taking their disjoint union. Conversely, the intersection of the ∂_0-preimage of A with the ∂_1-preimage of B specifies the hom-set $\mathscr{C}\langle A, B\rangle$.

Instead of a global partial composition operation we then have a family of local compositions $[A, B] \times [B, C] \longrightarrow [A, C]$ and of distinguished morphisms $1 \longrightarrow [A, A]$.

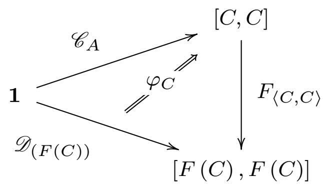
The advantage of this approach is the possibility to replace ***set*** as the category, where the hom-sets live, by some other suitable category $\mathscr{V}$. This must be equipped with a so-called *monoidal structure*, *i.e.*, a counterpart $\otimes$ for $\times$ that is associative and has a unit I. Allowing hom-objects to live in $\mathscr{V}$ esults in 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 $\mathbf{2}$. ◁

4.0.05 Definition. A *2-category* is a category enriched in ***Cat***, with cartesian product and $I = \mathbf{1}$ providing the required monoidal structure. A *2-functor* preserves i-cells, $i < 3$, as well as their composition and their units on the nose.

Besides 2-functors there are (at least) two further reasonable morphisms between 2-categories: *lax* and *oplax* functors $\mathscr{C} \xrightarrow{\langle F, \varphi \rangle} \mathscr{D}$, where the composition and units of the 1-cells do not need to be preserved on the nose, but only up to natural transformations of the form



and



in the lax case, and with reversed 2-cells in the oplax case. 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 $F(g) \circ F(f) \xrightarrow{\varphi\langle A,B,C\rangle\langle f,g\rangle} F(g \circ f)$ and $\mathscr{D}_{FC} \xrightarrow{\varphi_C} F\mathscr{C}_C$. 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 homoorphisms between ordered monoids.

4.0.06 Proposition. (Common constructions on categories)

- *The cartesian product $\mathscr{C} \times \mathscr{D}$ of two categories $\mathscr{C}$ and $\mathscr{D}$ is defined componentwise.*

- *Every subclass $\mathcal{A}$ of the object-class of a category $\mathscr{C}$ induces a so-called* full subcategories *of $\mathscr{C}$ with the same hom-sets $\mathscr{C}\langle A, B\rangle$ for all objects $A, B \in \mathcal{A}$.*

  Unless more is known about the elements of $\mathcal{A}$ there is little point in restricting the morphisms, but we will encounter such situations shortly.

- *The* arrow category *$\mathscr{C}^{\to}$ of $\mathscr{C}$ has morphisms of $\mathscr{C}$ as objects and commutative squares as morphisms.* □

In view of the Examles in (4) above, categories provide a common generalization of monoids and pre-ordered sets. This indicates that monoid homomorphisms and order-preserving functions ought to generalize as well to structure-preserving morphisms between categories.

4.0.07 Definition. A *functor* $\mathscr{C} \xrightarrow{F} \mathscr{D}$ between categories consists of

- a function $\mathscr{C}_0 \xrightarrow{F_0} \mathscr{D}_0$ between the object classes;
- either a function $\mathscr{C}_1 \xrightarrow{F_1} \mathscr{D}_1$ between the morphism classes such that sources and targets are preserved, *i.e.*,

$$\partial_0 F_1(f) = F_0 \partial_0(f) \quad \text{and} \quad \partial_1 F_1(f) = F_0 \partial_1(f)$$

such that composition and identity arrows are preserved, *i.e.*,

$$F_1(g) \circ F_1(f) = F_1 \quad \text{if } g \circ f \text{ is defined, and} \quad (g \circ f) F_1(id_A) = id_{F_0(A)} \quad \text{for each } A \in C_0 \quad \triangleleft$$

4.0.08 Remark. Equivalently, instead of F_1 one can specify a family of functions

$$\mathscr{C}\langle A, B\rangle \xrightarrow{F_{\langle A,B\rangle}} \mathscr{D}\langle F(A), F(B)\rangle$$

and require

$$\begin{array}{ccc}
\mathscr{C}\langle A, B\rangle \times \mathscr{C}\langle B, C\rangle & \xrightarrow{F_{\langle A,B\rangle} \times F_{\langle B,C\rangle}} & \mathscr{C}\langle F(A), F(B)\rangle \times \mathscr{D}\langle F(B), F(C)\rangle \\
{\scriptstyle \mathscr{C}_{\langle A,B,C\rangle}} \downarrow & & \downarrow {\scriptstyle \mathscr{D}_{\langle F(A),F(B),F(C)\rangle}} \\
\mathscr{C}\langle A, C\rangle & \xrightarrow[F_{\langle A,C\rangle}]{} & \mathscr{C}\langle F(A), F(C)\rangle
\end{array}$$

as well as

$$\begin{array}{ccc}
 & 1 & \\
{\scriptstyle id_A} \swarrow & & \searrow {\scriptstyle id_{FA}} \\
\mathscr{C}\langle A, A\rangle & \xrightarrow[F_{\langle A,C\rangle}]{} & \mathscr{C}\langle F(A), F(A)\rangle
\end{array}$$

This formulation of the notion of functor also works in the *enriched* setting, *cf.* Remark 4.0.04. $\triangleleft$

4.0.09 Examples.

(0) Each monoid homomorphism and each order-preserving function is a functor.

(1) Functor $\mathbf{1} \xrightarrow{C} \mathscr{C}$ 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.

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

- ▷ from ***set*** to itself;
- ▷ from ***set*** to the category $\bigvee$-***slat*** of $\bigvee$-semilattices;
- ▷ from ***mon*** to itself
- ▷ from ***mon*** to the category ***uqnt*** of unital quantales

4.0.10 Definition. A functor $\mathscr{A} \xrightarrow{F} \mathscr{B}$ is called *faithful/full*, if all F-components

$$\mathscr{A}\langle X, Y \rangle \xrightarrow{F\langle X,Y \rangle} \mathscr{B}\langle FX, FY \rangle$$

are injective/surjective. A faithful functor that also is injective on objects is called an *embedding*.

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

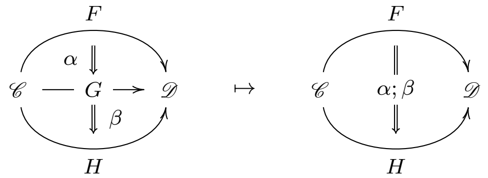
For general categories, the analogy with ordered sets can be carried further. Whenever we have order-preserving functions $\langle X, \leq \rangle \overset{f}{\underset{g}{\rightrightarrows}} \langle Y, \sqsubseteq \rangle$, they can be compared point-wise: $f \sqsubseteq g$ iff for every $a \in X$ we have $f(a) \sqsubseteq g(a)$. In other words, the set of order-preserving functions from $\langle X, \leq \rangle$ to $\langle Y, \sqsubseteq \rangle$ is itself an ordered set. Hence we would expect the set of functors from $\mathscr{C}$ to $\mathscr{D}$ to ba a category. We now specify the arrows in this category:

4.0.11 Definition. Given two functors $\mathscr{C} \overset{F}{\underset{G}{\rightrightarrows}} \mathscr{D}$, a *natural transformation* $F \overset{\alpha}{\Longrightarrow} G$ consists of a family of $\mathscr{D}$-arrows $F(A) \xrightarrow{\alpha_A} G(A)$, such that for any $\mathscr{C}$-arrow $A \xrightarrow{f} B$ we have

$$\begin{array}{ccc} F(A) & \xrightarrow{\alpha(A)} & G(A) \\ {\scriptstyle F(f)}\downarrow & & \downarrow{\scriptstyle G(f)} \\ F(B) & \xrightarrow[\alpha(B)]{} & G(B) \end{array} \tag{4.0-00}$$

There are two ways how natural transformations can be composed:

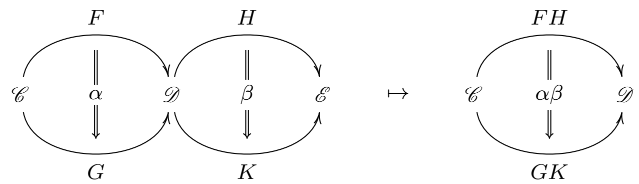
▷ *sequentially*:



the two possible composites

$\delta\beta \circ \gamma\alpha$ and $\delta \circ \gamma\beta \circ \alpha$ coincide. □

▷ *in parallel*:



which can be expressed as the sequential composition of $HF \xrightarrow{\beta F} KF$ with $KF \xrightarrow{K\alpha} KG$, or equivalently, of $HF \xrightarrow{H\alpha} KG$ with $HG \xrightarrow{\beta G} KG$. These latter obvious parallel composites of (identity natural transvormations of) functors with α, respectivley, β are known as *whiskerings*. ◁

4.0.12 Proposition. *For any two categories* $\mathscr{C}$ *and* $\mathscr{D}$ *the functors from* $\mathscr{C}$ *to* $\mathscr{D}$ *are the objects of a* functor category $[\mathscr{C}, \mathscr{D}]$, *which has the natural transformations between such functors as arrows. Any functors* $\mathscr{C}' \xrightarrow{F} \mathscr{C}$ *and* $\mathscr{D} \xrightarrow{H} \mathscr{D}'$ *induces a functor* $[\mathscr{C}, \mathscr{D}] \xrightarrow{[F,G]} [\mathscr{C}', \mathscr{D}']$ *that operates by pre- and post-composition,* i.e., $(\mathscr{C} \xrightarrow{G} \mathscr{D}) \mapsto (\mathscr{C}' \xrightarrow{FGH} \mathscr{D}')$. □

Notice that $[\mathscr{C}, \mathscr{D}]$ will be small, if both $\mathscr{C}$ and $\mathscr{D}$ are small, but need not be locally small, if $\mathscr{C}$ and $\mathscr{D}$ are. Since size will not be an issue in this course, we sidestep this question at the moment.

4.0.13 Examples.

(0) The family of inclusions $X \rightarrowtail X^*$ that map elements $a \in X$ to singleton words over X; recall that X is a subset of X^*.

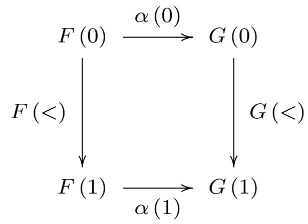
(1) The family of singleton-functions $X \overset{\{-\}}{\rightarrowtail} \mathbb{P}X$ that maps elements $a \in X$ to the singleton set $\{a\} \subseteq X$.

(2) The family of union maps $\mathbb{P}(\mathbb{P}X) \xrightarrow{\bigcup} \mathbb{P}(X)$ that forms the union of a set of subsets of X.

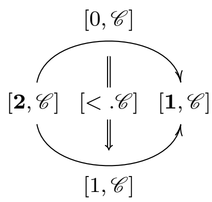
(3) If $\mathscr{C}$ is a monoid, a natural transformation $F \xrightarrow{\alpha} G$ is a single morphism $F* \xrightarrow{\alpha_*} G*$ in $\mathscr{D}$ that satisfies Diagram (**??**). While the F-and G-images of $\mathscr{C}$ form sub-monoids of $[F*, F*]$ and $[G*, G*]$, respectively, α_* in general does not induce a monoid homomorphism between these.

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

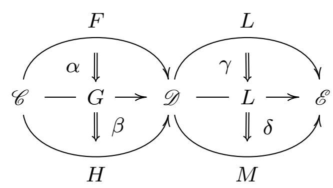
(5) For any category $\mathscr{C}$ the functor category $[\mathbf{2}, \mathscr{C}]$ is the same as the arrow-category $\mathscr{C}^{\rightarrow}$ of Proposition 4.0.06. Here natural transformations are pairs of morphisms that make a specific square commutative:



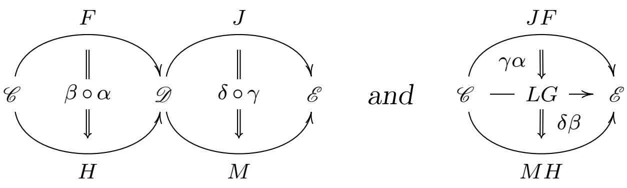
There are obvious domain-and codomain-functors from $[\mathbf{2}, \mathscr{C}]$ to $\mathscr{C}$ that map the square above to $F(0) \xrightarrow{\alpha_0} G(0)$, respectively, $F(1) \xrightarrow{\alpha_1} G(1)$. More precisely: when we identify $\mathscr{C}$ with $[\mathbf{1}, \mathscr{C}]$ and $\mathbf{2}$ with $[\mathbf{1}, \mathbf{2}]$, the relation $0 < 1$ produces a natural transformation



4.0.14 Theorem. *The sequential and parallel composition of natural transformations of Definition 4.0.11 satisfy the* middle interchange condition, i.e., *in the following situation*



the two possible composites agree:



4.0.15 Proposition. *Small categories as objects together with the functors as 1-cells and the natural transformations as 2-cells form a 2-category* ***cat*** , i.e., *a category enriched over itself,* cf. *Remark 4.0.04.*

4.0.16 Remark. Replacing hom-sets by hom-categories opens up another direction for generaiza-tion: the unit- and associativity laws for composition of 1-cells can be relaxed to only hold up to invertible morphisms (or isomorphism, see below), that are well-behaved in a technical sense known as "coherence". This relaxation results in "weak 2-categories", also known as "bi-categories". The only bicategory we will encounter in this lecture that is not a 2-category will probably be ***spn*** , *cf.*, Definition 4.7.00, a generalization of ***rel*** . ◁

4.1 Special morphisms

As the prototypical category is ***set*** , with functions as morphisms, it is no surprise that the notions of injectivity and surjectivity have been generalized to morphisms in arbitrary categories. As it turns out, the aspects of cancellability and invertibility lead to different generalizations.

4.1.00 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 $g \circ r = g \circ s$ impplies $r = s$ for any $\mathscr{C}$-morphisms $A \overset{r}{\underset{s}{\rightrightarrows}} B$;

▷ *section*, also known as *split mono*, if g has a *right inverse* $C \xrightarrow{h} B$ satisfying $h \circ g = id_B$.

Dual notions: *epimorphism*, *pre-cancellable*; *retraction* or *split epi*.

If g is invertible on both sides, it is called an *isomorphism*, or *iso* for short.

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 or course both epi and mono, as the pre- and post-compositin functions are bijective. Hovever, the converse need not be true:

4.1.01 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 $\mathbb{Z} \overset{f}{\underset{g}{\rightrightarrows}} M$. The images of $\mathbb{Z}$ under f and g automatically are groups, and hence f and g are uniquely determined by the values of $f \circ 1$ and $g \circ 1$, respectively, which by hypothesis are different elements of M . But then also $f \circ i \circ 1 \neq g \circ i \circ 1$, which implies $f \circ i \neq g \circ i$.

(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:

4.1.02 Definition. An epimorphism $A \xrightarrow{e} B$ is called

(0) *extramal epi*, if e does not factor through a proper mono, *i.e.*, whenever $e = m \circ f$ with m mono, then m is already iso;

(1) *strong epi*, if for every diagram of the form

$$\begin{array}{ccc} A & \xrightarrow{f} & C \\ {\scriptstyle e}\downarrow & & \downarrow{\scriptstyle m} \\ B & \xrightarrow[g]{} & D \end{array} \tag{4.1-00}$$

with m mono there exists a unique diagonal $B \xrightarrow{d} C$ making both triangles commute:

$$\begin{array}{ccc} A & \xrightarrow{f} & C \\ {\scriptstyle e}\downarrow & \nearrow^{d} & \downarrow{\scriptstyle m} \\ B & \xrightarrow[g]{} & D \end{array}$$

Dual Notions: *extremal mono*; *strong mono*.

4.1.03 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 (4.1-00) is split with left-inverse $B \xrightarrow{h} A$. We claim that $d := f \circ h$ is the desired diagonal. The lower triangle commutes because of

$$m \circ f \circ h = g \circ e \circ h = g$$

while the upper triangle commutes since m is mono and

$$m \circ f = g \circ e = m \circ f \circ h \circ e$$

(1) An iso is both mono and split epi, hence by (0) mono and strong epi. Conversely, set $e = m$ in Diagram (4.1-00), and use identities for the horizontal morphisms. Then d is left- and right-inverse to e.

(2) Suppose in Diagram (4.1-00) $D = B$ 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. □

4.1.04 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 wth non-empty domain are split mono: the inclusion of $\emptyset$ into an nonempty set cannot have a right-inverse. ◁

4.1.05 Proposition. *If* $B \xrightarrow{g} C \xrightarrow{h} D$ *is mono, so is* g.

Proof. If $A \underset{s}{\overset{r}{\rightrightarrows}} B$ are distinct, so are $h \circ g \circ r$ and $h \circ g \circ s$, which forces $g \circ r \neq g \circ s$. □

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 4.1.01 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.

4.1.06 Definition. Monos into an objcet C are calles *sub-objects* of C.
DUAL NOTION: *super-object.*

4.1.07 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 expicitly 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, otherse 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 4.7, we will need to consider regular epis.

4.1.08 Examples.

▷ In ***set*** subobjects 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 $\bigsqcup$-semilattices, which are, of course, $\sqcup$-semilattices, 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 4.6.

4.1.09 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*.

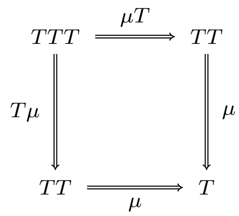
Dual notion: *sink*, *epi-sink*.

4.2 Monads

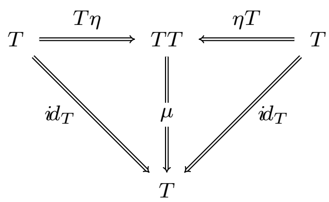
The key observation in Subsection 0.1 was that the free monoid functor and the power-set functor on ***set*** , carry extra structure that much resembles a monoid. Here come the official definitions.

4.2.00 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

▷ μ is associative, *i.e.*,



▷ η is left- and right-neutral with respect to μ , *i.e.*



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* $\nu_{\boldsymbol{Q}}$.

4.2.01 Examples.

(0) Further monads in the vein of the free monoid monad $(-)^*$ and the power-set monads $\boldsymbol{P}$ and $\boldsymbol{F}$ are indicated in Subsection 4.11.

(1) For every finitaly algebraic theory (defined by a finite signature Σ of function symbols with finite arity) the category of Σ-algebras arises as the the category of EM-algebras for a suitable monad.

(2) Algebraic theories with infinite signatures need not give rise to monads: *e.g.*, the category ***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***).

(3) Compact Hausdorff spaces arise as EM-algebras for the ultrafilter monad, another example for a signature with a function-symbol of infinite arity.

4.2.02 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 $\mathscr{C}^{\boldsymbol{T}}$.

4.2.03 Remark. Of course, the definition of monad makes sense in any 2-category:

▷ Monads in ***rel*** are just pre-orders, *i.e.*, reflexive and transitive relations on a set.

▷ 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.

4.2.04 Definition.

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

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

$$\begin{array}{ccc} T(X) & \xrightarrow{T(f)} & T(Y) \\ {\scriptstyle \xi}\downarrow & & \downarrow{\scriptstyle \zeta} \\ X & \xrightarrow[f]{} & Y \end{array} \tag{4.2-00}$$

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

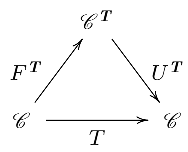
(1) In case of a monad $\boldsymbol{T} = \langle T, \eta, \mu \rangle$ on $\mathscr{C}$, an *Eilenberg-Moore algebra*, or *EM-algebra* for short, is an algebra $\langle X, TX \xrightarrow{\xi} X \rangle$ for T , subject to two compatibility conditions with η and μ , respectively:

$$\begin{array}{ccc} X & \xrightarrow{\eta_X} & T(X) \\ & {\scriptstyle id_X}\searrow & \downarrow{\scriptstyle \xi} \\ & & X \end{array} \qquad \text{and} \qquad \begin{array}{ccc} TT(X) & \xrightarrow{T(\xi)} & T(X) \\ {\scriptstyle \mu(X)}\downarrow & & \downarrow{\scriptstyle \xi} \\ T(X) & \xrightarrow[\xi]{} & X \end{array}$$

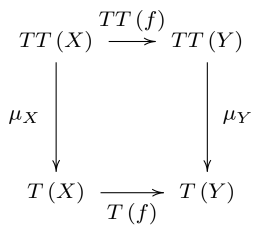
Notice that $\langle TX, TTX \xrightarrow{\mu_X} TX\rangle$ is always an EM-algebra, the so-called *free* EM-algebra over X. Furthermore, the structure map ξ of an EM-algebra $\langle X, \xi\rangle$ by definition is always an algebra homomorphism from the free EM-algebra over X into $\langle X, \xi\rangle$.

DUAL NOTIONS: *co-algebra* $\langle X, X \xrightarrow{\zeta} TX\rangle$; *coalgebra homomorphism*; *Eilenberg-Moore coalgebra* or *EM-coalgebra*.

4.2.05 Proposition. *The EM-algebras and the algebra-homomorphisms for a monad* $\boldsymbol{T}$ *over* $\mathscr{C}$ *form a category* $\mathscr{C}^{\boldsymbol{T}}$. *Moreover, the endofunctor* $\mathscr{C} \xrightarrow{T} \mathscr{C}$ *factors through* $\mathscr{C}^{\boldsymbol{T}}$ *by via*



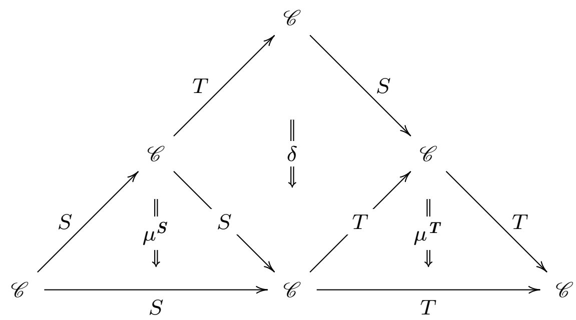
where $F^{\boldsymbol{T}}$ *maps* $X \xrightarrow{f} Y$ *in* $\mathscr{C}$ *to the algebra-homomorphism*



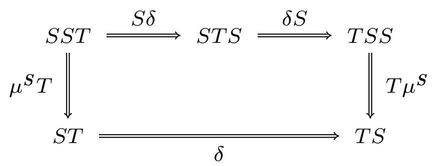
between the free algebras over X, *and* Y, *respectively. Conversely,* $U^{\boldsymbol{T}}$ *maps an algebra-homomorphism (4.2-00) to the* underlying $\mathscr{C}$*-morphism* $X \xrightarrow{f} Y$. □

4.3 Distributive laws

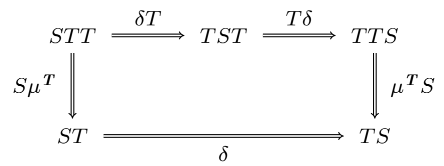
Given two monads $\boldsymbol{S} = \langle S, \eta^{\boldsymbol{S}}, \mu^{\boldsymbol{S}}\rangle$ and $\boldsymbol{T} = \langle T, \eta^{\boldsymbol{T}}, \mu^{\boldsymbol{T}}\rangle$ on the same category $\mathscr{C}$, the question arises, whether the composite functor $\mathscr{C} \xrightarrow{S} \mathscr{C} \xrightarrow{T} \mathscr{C}$ carries a monad structure as well. The problem is to define a meaningful multiplication $TSTS \Longrightarrow TS$. This can fail in general, but in the presence of a natural transformation $ST \overset{\delta}{\Longrightarrow} TS$ satisfying suitable axioms, the original multiplications $SS \overset{\mu^{\boldsymbol{S}}}{\Longrightarrow} S$ and $TT \overset{\mu^{\boldsymbol{T}}}{\Longrightarrow} T$ can be brought to bear, as indicated in the following diagram:



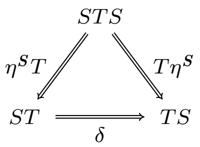
The axioms for δ are chosen such that the composite 2-cell above yields a monad together with the obvious unit $id_{\mathscr{C}} = id_{\mathscr{C}} id_{\mathscr{C}} \xrightarrow{\eta^S \eta^T} ST$, namely:



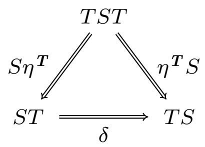
and



as well as



and



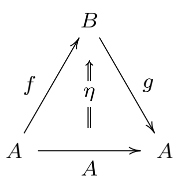
the reader is encouraged to check that the unit and multiplication above indeed provide a monad structure on TS .

4.4 Adjunctions

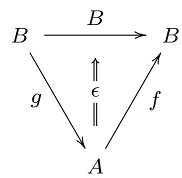
One of the fundamental notions of category theory is the notion of adjunction, due to Daniel Kan in 1958. It may be seen as a weakening of the notion of inverse functions in $\boldsymbol{set}$. The only sensible way that a function $X \xrightarrow{f} Y$ can have an inverse in $\boldsymbol{set}$ is to have a function $Y \xrightarrow{g} X$ such that $g \circ f = id_X$ and $f \circ g = id_Y$. 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 $g \circ f$ and id_X , respectively, $f \circ g$ and id_Y . Also recall the notion of a generalized inverse matrix M^{g} in linear algebra that satisfies $MM^{\mathrm{g}}M = M$. I

We start with the abstract concept available in any 2-category, which takes a diagrammatic form that is easy to memorize (we hope).

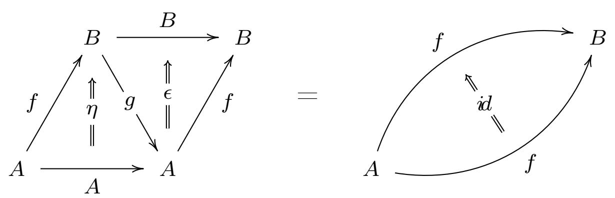
4.4.00 Definition. Two 1-cells $A \xrightarrow{f} B$ and $B \xrightarrow{g} A$ are called *adjoint*, if there exist 2-cells

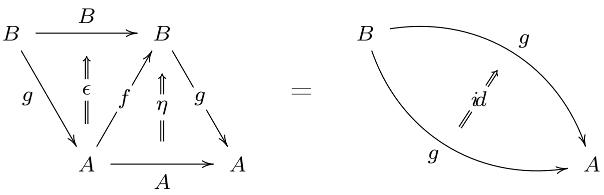


and



subject to the following axioms:

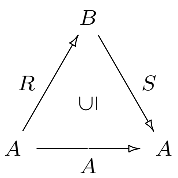




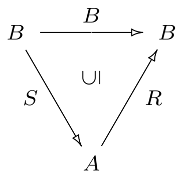
Notation: $f \dashv g$; f is called *left adjoint* and g *right adjoint*, while η and ϵ are called the *unit*, respectively *co-unit* of the adjunction.

4.4.01 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

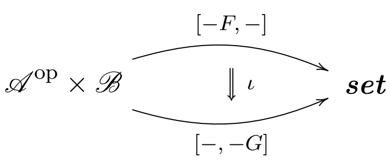


and



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 ***Cat*** can also be described in somewhat differernt 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 $\mathscr{C}\langle A, GB\rangle$ and $\mathscr{D}\langle FA, B\rangle$ 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 FA in $\boldsymbol{D}$, resp. the pre-images of the identities on GB in $\mathscr{C}$ yield the components $A \xrightarrow{\eta_A} GFA$ of the unit and $FGB \xrightarrow{\epsilon_B} B$ of the co-unit.

Conversely, given $\mathscr{C} \xRightarrow{\eta} GF$ and $A \in \mathscr{C}$ as well as $B \in \mathscr{D}$, every $\mathscr{C}$-morphism $A \xrightarrow{f} GB$

admits a unique $\mathscr{D}$-morphism $FA \xrightarrow{\tilde{f}} B$ such that $G\left(\tilde{f}\right)$ extends f along η_A:

$$\begin{array}{ccccc} A & \xrightarrow{\eta_A} & GFA & & FA \\ & \overset{\forall f}{\searrow} & \downarrow{\scriptstyle G\check{f}} & & \vdots\,{\scriptstyle \exists!\, \check{f}} \\ & & GB & & B \end{array} \tag{4.4-00}$$

Namely, $\check{f} := \epsilon_B \circ F(f)$.

Alternatively, starting with the co-unit $FG \xrightarrow{\epsilon} \mathscr{D}$ and objects A and B as above, every $\mathscr{D}$-morphism $FA \xrightarrow{g} B$ induces a unique $\mathscr{C}$-morphism $A \xrightarrow{\hat{g}} GB$ such that $F\hat{g}$ lifts g along ϵ_B:

$$\begin{array}{ccccc} A & & FA & & \\ \vdots\,{\scriptstyle \exists!\, \hat{g}} & & \downarrow{\scriptstyle F\hat{g}} & \overset{\forall g}{\searrow} & \\ GB & & FGB & \xrightarrow[\epsilon_B]{} & B \end{array} \tag{4.4-01}$$

Namely, $\hat{g} := G(g) \circ \eta_A$.

(2) Every monad $\boldsymbol{T} = \langle T, \eta^{\boldsymbol{T}}, \mu^{\boldsymbol{T}} \rangle$ on $\mathscr{C}$ induces an adjunction between the free functior $\mathscr{C} \xrightarrow{F^{\boldsymbol{T}}} \mathscr{C}^{\boldsymbol{T}}$ into the category of EM-algebras, and the forgetful functor $\mathscr{C}^{\boldsymbol{T}} \xrightarrow{U^{\boldsymbol{T}}} \mathscr{C}$.

4.5 Extensions and liftings

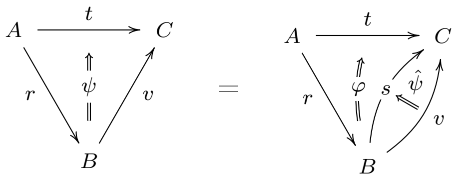
We have seen that many important notions of category theory, like monads and adjunctions, can in fact be formlized in any 2-category. Another such concept, *Kan-extensions* (due to Daniel M. Kan, 1960) can be abstracted in a similar fashion and provides a very economical, if slightly abstract approach, that subsumes many other notions. MacLane [**?**] 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 [**?**]. Consider a 2-cell in a 2-category $\mathcal{B}$.

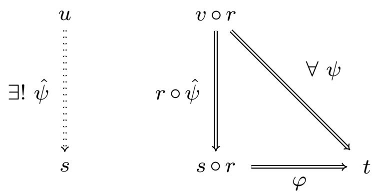
$$\begin{array}{ccc} A & \xrightarrow{t} & C \\ {\scriptstyle r}\searrow & \Uparrow \varphi & \nearrow{\scriptstyle s} \\ & B & \end{array} \tag{4.5-00}$$

4.5.00 Definition. Diagram (4.5-00) 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 φ at s is a bijection between the hom-sets $[A, C]\langle x \circ r, t\rangle$ and $[B, C]\langle x, s\rangle$. More precisely, any 2-cell $u \circ r \xRightarrow{\psi} t$ uniquely factors through φ



(4.5-01)

Alternatively, this universal property may be depicted as



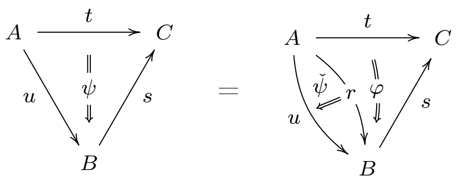
(4.5-02)

with a unique 2-cell $x \xRightarrow{\bar{\psi}} s$.

The right extension $\langle s, \varphi \rangle$ of t along r is called *absolut*, if for every 1-cell $C \xrightarrow{w} D$ the pair $\langle w \circ s, w \circ \varphi \rangle$ is a right-extension of $w \circ t$ along r .

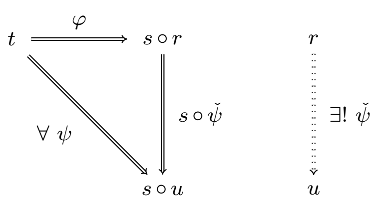
Dual notion: Diagram (4.5-00) exhibits the pair $\langle r, \varphi \rangle$ as a *right lifting* of $A \xrightarrow{t} C$ *through* $B \xrightarrow{s} C$; *absolute right lifting*.

Opposite notion: If the 2-cell φ in Diagram (4.5-00) 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



(4.5-03)

or, alternatively



(4.5-04)

Absolute left extensions/liftings are preserved by post/pre-cmposition with 1-cells.

4.5.01 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 (4.5-02) with Diagram (4.4-01) of Example 4.4.01(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. Another commen notation is $r \backslash -$.

(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.)

4.5.02 Definition. The 2-cateogory $\mathcal{B}$ 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.

4.5.03 Remark. In a closed bicategory $\mathcal{B}$ the bijective correspondences

$$\begin{array}{rcl} r & \Longrightarrow & t \triangleleft s \\ \hline\hline s \circ r & \Longrightarrow & t \\ \hline\hline s & \Longrightarrow & r \triangleright t \end{array}$$

automatically induces further adjunctions

$$[A,B] \underset{-\rhd t}{\overset{t\lhd -}{\underset{\top}{\leftrightarrows}}} [B,C]^{\mathrm{op}} \qquad \text{as well as} \qquad [B,C] \underset{t\lhd -}{\overset{-\rhd t}{\underset{\top}{\leftrightarrows}}} [A,B]^{\mathrm{op}} \tag{4.5-05}$$

that are known as *polarities*.

4.5.04 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

$$\begin{aligned} \langle b,c\rangle \in R \rhd T \quad &\text{iff} \quad \forall a \in A.\ \langle a,b\rangle \in R \Rightarrow \langle a,c\rangle \in T \\ \langle a,b\rangle \in T \lhd S \quad &\text{iff} \quad \forall c \in C.\ \langle a,c\rangle \in T \Leftarrow \langle b,c\rangle \in S \end{aligned}$$

In other words, we have adjunctions

$$[B,C] \underset{[R,C]}{\overset{R\rhd -}{\underset{\top}{\leftrightarrows}}} [A,C] \qquad \text{as well as} \qquad [A,B] \underset{[A,S]}{\overset{-\lhd S}{\underset{\top}{\leftrightarrows}}} [A,C] \tag{4.5-06}$$

and polarities

$$[A,B] \underset{-\rhd T}{\overset{T\lhd -}{\underset{\top}{\leftrightarrows}}} [B,C]^{\mathrm{op}} \qquad \text{as well as} \qquad [B,C] \underset{T\lhd -}{\overset{-\rhd T}{\underset{\top}{\leftrightarrows}}} [A,B]^{\mathrm{op}} \tag{4.5-07}$$

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 (4.5-06), by considering the dual relation $C \xrightarrow{S^{\mathrm{op}}} B$ we obtain a second adjunction in the opposite direction:

$$[1,B] \underset{[1,S]}{\overset{-\lhd S}{\underset{\top}{\leftrightarrows}}} [1,C] \qquad \text{as well as} \qquad [1,C] \underset{[1,S^{\mathrm{op}}]}{\overset{-\lhd S^{\mathrm{op}}}{\underset{\top}{\leftrightarrows}}} [1,B] \tag{4.5-08}$$

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

$$[1,B] \underset{[1,g]}{\overset{-\lhd g^{\mathrm{op}}}{\underset{\perp}{\overset{\perp}{\longleftarrow [1,g^{\mathrm{op}}] = -\lhd g \longrightarrow}}}} [1,C] \;\cong\; \mathbb{P}B \underset{g_\exists}{\overset{g_\forall}{\underset{\perp}{\overset{\perp}{\longleftarrow g^{-1} \longrightarrow}}}} \mathbb{P}C \;\cong\; [B,1] \underset{[g^{\mathrm{op}},1]}{\overset{g\rhd -}{\underset{\perp}{\overset{\perp}{\longleftarrow [g,1] = g^{\mathrm{op}}\rhd - \longrightarrow}}}} [C,1]$$

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

$$\begin{aligned} s_{\exists}[V] &:= \{ c \in C : \exists b \in B.\, b \in V \wedge s(b) = c \} = \{ c \in C : V \cap cs^{-1} \neq \emptyset \} \\ s_{\forall}[V] &:= \{ c \in C : \forall b \in B.\, s(b) = c \Rightarrow b \in V \} = \{ c \in C : cs^{-1} \subseteq V \} \end{aligned} \tag{4.5-09}$$

The set $s_{\exists}[V]$ 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? $\lhd$

4.5.05 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 coud be violated.

(1) Although ***set*** as a locally discrete category is not closed, the *suspension* ***set***-Σ of ***set*** is closed: this 2-category has a single object $*$, sets as 1-cells and functions as 2-cells. In particular, ***set***-Σ 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* $X \times [X,Y] \xrightarrow{\mathrm{ev}_X} Y$ that maps $\langle x, f \rangle$ to xf.

(2) 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.

(3) 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***.

(4) 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.)

(5) The operation of cartesian product can even be available in 2-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***.

(6) ***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 $\mathbf{1}$.

4.5.06 Theorem. *For 1-cells* $A \xrightarrow{f} B \xrightarrow{g} A$ *and a 2-cell* $f \circ g \xRightarrow{\epsilon} B$ *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) ϵ *is the co-unit of an adjunction* $f \dashv g$.

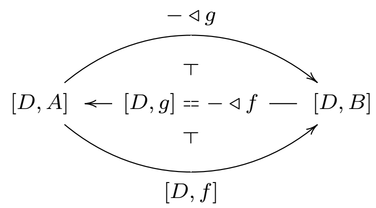
Proof.

(a) ⇒ (b): Trivial.

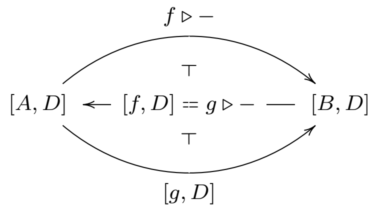
(b) ⇒ (c): Let $A \xRightarrow{\eta} g \circ f$ be the pre-image of the canonical 2-isomorphism $\bar{f}; A \Longrightarrow B; \bar{f}$ under pasting of $g \circ \varepsilon$ at $g \circ f$. Pasting $A; f \xRightarrow{(\eta f);(f\varepsilon)} f; B$ with ε at f modulo structural isos results in ε, which establishes ε and η as the counit and unit of an adjunction $f \dashv \bar{f}$.

(c) ⇒ (a): The desired bijection between 2-cells of the form $p \Longrightarrow f; q$ and 2-cells of the form $\bar{f}; p \Longrightarrow q$ is obtained by pasting with ε at f and with η at $\bar{f}$, respectively. □

4.5.07 Corollary. *The unit of an adjunction* $f \dashv g$ *is both an absolute left lifting and an absolute left extension, while the counit 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*



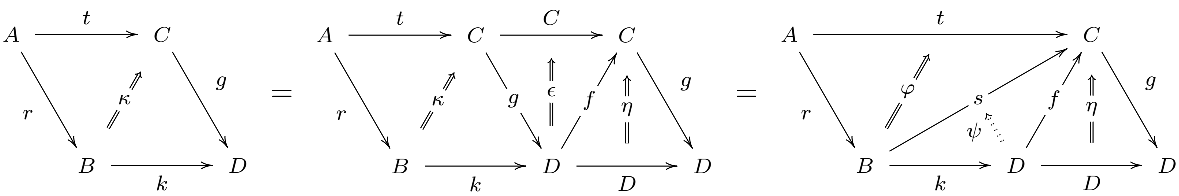
and



□

4.5.08 Theorem. *Post-composition with left/right adjoints preserves left/right extensions.*

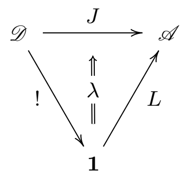
Proof. Suppose Diagram (4.5-00) exhibits $\langle s, \varphi \rangle$ as right extension of t along r, and $C \xrightarrow{g} D$ is right adjoint with left adjoint $D \xrightarrow{f} C$. We need to show that pasting with $g \circ \varphi$ at $g \circ s$ is a bijection between $\langle k, g \circ s \rangle[B, D]$ and $\langle k \circ r, g \circ t \rangle[A, D]$. Given a 2-cell $k \circ r \xRightarrow{\kappa} g \circ t$, we have



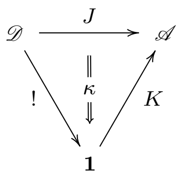
where ψ is the uniquely determined 2-cell by which $\eta t \circ f\kappa$ factors through φ. Hence $\hat{\kappa} := g\psi \circ \eta k$ is a candidate for the image of κ. Any other such candidate $k \xRightarrow{\omega} gs$ composed with $\epsilon \circ s$ must yield ψ, hence $\omega = \hat{\kappa}$. □

4.6 Limits and co-limits

4.6.00 Definition. In ***cat*** and ***Cat*** 2-cells of the form



resp.



(4.6-00)

are called *cones/co-cones* of the functor $\mathcal{J}$. If they have the universal property of a right/left extension, they are called *limits/colimits*. The category $\mathscr{A}$ is called *complete/co-complete*, if it has all small(!) limits/co-limits.

4.6.01 Remarks.

(0) The composition and units of $\mathscr{D}$ are irrelevant for these notion; 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 "upper bounds" and of co-cones as "lower bounds" of the diagram $\mathfrak{J}$.

(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 ordererd pair.

(3) Associated with every cone/co-cone is a source/sink (*cf.*, Definition 4.1.09), which results from pre-composing $\mathfrak{J}$ 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 acceppts 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/colimits).

The following useful result follows directly from the definition:

4.6.02 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:

4.6.03 Definition. A category is called *finitely complete*, if it has all limits of finite diagrams, and *complete*, if it has all limits of small diagrams.

Requiring (co-)limits of larger, *i.e.*, class-sized diagrams to exist causes the category to collapse:

4.6.04 Theorem. *A locally small category with all limits or all colimits is a complete lattice. The same is true for small categories with all small limits or all small colimits.* □

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 colimits:

4.6.05 Examples.

(0) 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 thet the souce of projections always is a mono-source.

(1) 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 $J(0)$, and this is always mono, while for coequalizers the morphism out of $J(1)$ is always epi.

(2) If $\mathscr{D}$ is a cospan $1 \xrightarrow{f} 0 \xleftarrow{g} 2$, we obtain *pullbacks* as limits, but no interesting colimits (why?). The case of $J(f) = J(g)$ will be of interest in conncection with congruences and such pullbacks are also known as *kernel pairs*.

(3) If $\mathscr{D}$ is a span $1 \xleftarrow{f} 0 \xrightarrow{g} 2$, we obtain *push-outs* as colimits, but no interesting limits. Push-outs in case of $J(f) = J(g)$ are called *co-kernel pairs*.

(4) If $\mathscr{A}$ is an ordered set, then limits/colimits turn into infima/suprema of the image of $\mathscr{D}$, *i.e.*, a certain subset.

The fact that coequalizers are epi is the basis to define an important class of epimorphisms:

4.6.06 Definition. An epimorphism $B \xrightarrow{g} C$ is called *regular*, if it arises as coequalizer of some parallel pair $A \overset{r}{\underset{s}{\rightrightarrows}} B$ of morphisms.
Dual Notion: *regular mono*.

4.6.07 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 $e \circ id_A = e = e \circ (s \circ e)$. We claim that e is a coequalizer of $A \overset{id_A}{\underset{s \circ e}{\rightrightarrows}} A$. Consider $A \xrightarrow{f} C$ with $f \circ id_A = f = f \circ (s \circ e)$. By hypothesis e factors through f via $s \circ e$. Suppose $f = g \circ e$ for some $B \xrightarrow{g} C$. Now $g \circ e = f$ implies $g = g \circ e \circ s = f \circ s$, hence $f \circ s$ is the only possibility for e to factor through f.

(1) Suppose e in Diagram (4.1-00) is a coequalizer of $Z \overset{u}{\underset{v}{\rightrightarrows}} A$. Since m is mono, we have $f \circ u = f \circ v$, hence the universal property of the coequalizer induces a diagonal d that makes the upper triangle commute. By Proposition 4.6.02 e is epi and hence the lower triangle commutes as well.

(2) If $\mathscr{C}$ has pullbacks, form the pullback of Diagram (4.1-00). Then e factores through the pullbck 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. □

4.6.08 Proposition. *In any category $\mathscr{C}$ the following are equivalent:*

(a) $A \xrightarrow{f} B$ *is mono;*

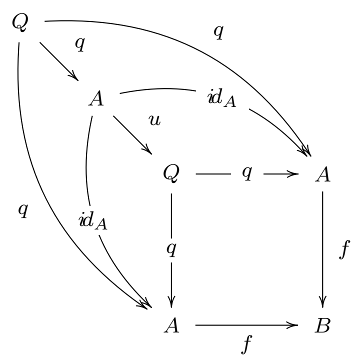
(b) f *has a kernel pair of the form* $\langle q, q \rangle$;

(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-inferse to $Q \xrightarrow{q} A$. Therefore u is split mono and q is split epi. As the outer triangles commute trivially, $u \circ q$ 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. □

4.6.09 Examples.

▷ ***set*** is complete and cocomplete.

▷ ***rel*** has all small products, and by self-duality all small coproducts, 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 colimits often are more complicated. This applies, for instance, to categories of EM-algebras for monads on ***set***. In particuar, the coproduct 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 $\mathcal{M}$ nd in $\mathcal{N}$.

4.6.10 Proposition. *Left/right adjoint functors in* ***Cat*** *preserve colimits/limits, while left/right adjoint order-preserving functions in* ***pos*** *preserve suprema/infima.*

Proof. This is an immediate consequence of Theorem 4.5.08. □

4.6.11 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 $\boldsymbol{set}\langle A, C\rangle$, 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 ***set*** to $\boldsymbol{set}^{\mathrm{op}}$, and right adjoint from $\boldsymbol{set}^{\mathrm{op}}$ to ***set***. Both properties imply that co-limits in ***set*** are mapped to limits in ***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 differernt notation.

The third law arises from the adjunction $A \times - \dashv [A, -]$, since the co-unit $A \times [A, C] \xrightarrow{\text{ev}} C$ 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 wrt. $\times$ 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. ◁

4.7 Congruences

Any function $B \xrightarrow{g} C$ in ***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 equivelence 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 subobject g-***img*** of C;
- ▷ any bijection into the set $|g\text{-}\boldsymbol{img}|$ admits a lifting to an isomorphisms with codomain g-***img***; this proerty is sometimes called *transprotability*.

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.

4.7.00 Definition. Let $\mathscr{C}$ be a category with pullbacks. The bi-category ***spn*** $(\mathscr{C})$ 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 cospan $R \xrightarrow{r_1} B \xleftarrow{s_0} S$, 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 ***csp*** $(\mathscr{C})$ with *co-spans* $\boldsymbol{R} = (A \xrightarrow{r_0} R \xleftarrow{r_1} B)$ as 1-cells.

4.7.01 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 multipliction and addition are replaced by cartesianproduct and disjoint union, respecively.

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 4.5.02.

(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 η and μ 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:

4.7.02 Definition. The *image* of a $\mathscr{C}$-morphism $A \xrightarrow{f} B$, if it exists, is the smallest subobject of B through which f factors. It will be denoted by $\boldsymbol{img}(f)$.

4.7.03 Proposition. *If* $A \xrightarrow{f} B$ *has an image-factoization* $A \xrightarrow{e} \boldsymbol{img}(f) \xrightarrow{m} B$, *then* $A \xrightarrow{e} \boldsymbol{img}(f)$ *is an extremal epi.* □

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 a relation, due to its universal property as a pullback is trivially reflexiv and symmetric, its transitivity initially takes the somewhat strange form that the composite span(!) factors throuth 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.

4.7.04 Definition. A mono-span $\boldsymbol{R} = (A \xleftarrow{r_o} R \xrightarrow{r_1} A)$ on a $\mathscr{C}$-object A is called a *congruence* or *internal equivalence relation*, if the composit span exists and the identity span, the opposite span and the composit 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.

The following definition will ensure that image factorizations exist by requiring that congruences built via kernel pairs do have quotients, analogous to sets of equivalence classes equipped with the relevant structure.

4.7.05 Definition. A finitey complete category $\mathscr{C}$ is called *regular*, if

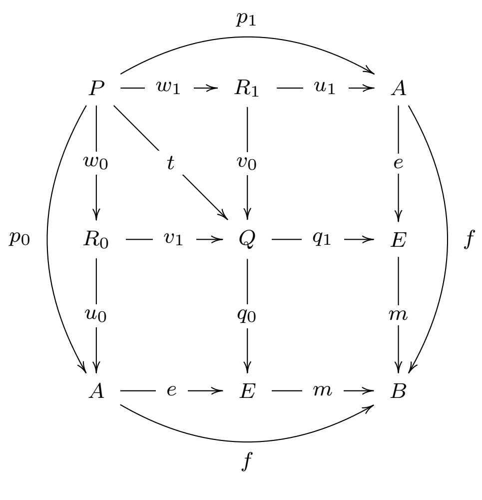
- ▷ coequalizers of kernel pairs exist in $\mathscr{C}$;
- ▷ regular epis are stable under pullback;

and *exact* or *Barr-exact*, if in addition

- ▷ every concruence is a kernel pair.

4.7.06 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 coequalizer $A \xrightarrow{e} E$. Since $f \circ p_0 = f \circ p_1$ there exists a unique morphism $E \xrightarrow{m} B$ satisfying $f = m \circ e$. 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 paricular implies that t as the composition of two epis is epi. Now

$$q_0 \circ t = e \circ p_0 = e \circ p_1 = q_1 \circ t$$

shows that $q_0 = q_1$, hence by Proposition 4.6.08 m is mono. □

4.7.07 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:

4.7.08 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.*

4.8 Comma Categories

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

4.8.00 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 $f \in \mathscr{A}\langle F(C), G(D)\rangle$;
- ▷ as morphisms from $\langle C, f, D\rangle$ to $\langle C', f', D'\rangle$ pairs of morphisms $\langle u, v\rangle \in \mathscr{C}\langle C, C'\rangle \times \mathscr{D}\langle D, D'\rangle$ subject to the condition

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

4.8.01 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 $\mathscr{C}\langle C, C' \rangle$.

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 $F\partial_0 \xrightarrow{F \triangleright G} G\partial_1$ 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 φ factors as $id_V(F \triangleright G)$.

4.8.02 Proposition. *The category* $1 \downarrow \boldsymbol{set}$ *of* pointed sets, *whose objects are sets with distinguished base point and whose morphisms are base-point preserving functions, 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.*

Proof. For a set X we write X_* for the one-pont 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 $X \rightharpoonup Y$. Then we obtain an "evaluation"

$$X_* \wedge \lfloor X_*, Y_* \rfloor_* \xrightarrow{ev} Y_*$$

specified by

$$\langle x, f \rangle \mapsto f(x) \quad , \quad \langle *, f \rangle \mapsto * \quad , \quad \langle x, * \rangle \mapsto * \quad \text{and} \quad \langle *, * \rangle \mapsto *$$

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. □

4.8.03 Remark. One-sided variants like "left zeros" and "right zeros", respevively, intended to absorb only when pre- or postcomposed, are incompatilbe with the associativity of composition. For the same reason one-sided variants of the smash product fail to be associative.

4.9 Profunctors

The standard morphisms between categories are functors, see Definition 4.0.01, 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 4.5.04).

It seems reasonable to generalize functors between categories to more general morphisms in much the same way as relations generlize functions. The corresponding "profunctors" have already been mentioned in Example 4.5.05(5).

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

We may indicate the functional assignment $A \mapsto F(A)$ by introducing new arrows from the $\mathscr{C}$-object A to the $\mathscr{D}$-object $F(A)$. 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 $A \longrightarrow F(A)$, 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 $A \longrightarrow F(A)$ has to be composed with all $\mathscr{C}$-arrows with codomain A, and with all $\mathscr{D}$-arrows with domain $F(A)$ in suitably associative fashion. Furthermore, one then would expect to obtain commutative squares of the form

$$\begin{array}{ccc} A & \longrightarrow & F(A) \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle F(f)} \\ B & \longrightarrow & F(B) \end{array}$$

Abstracting from the functor $\mathscr{C} \xrightarrow{F} \mathscr{D}$ one arrives at

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

Interpretation: Consider the elements of $R\langle C, D\rangle$ as new arrows from the $\mathscr{C}$-object C to the $\mathscr{D}$-object D. Funtoriality 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 supercategory of the disjoint union $\mathscr{C} + \mathscr{D}$ with potentially extra arrows joining $\mathscr{C}$-objects with $\mathscr{D}$-objects, a so-called "glueing" of $\mathscr{C}$ with $\mathscr{D}$.

4.9.01 Remark. The composition of profunctors is somewhat problematic: given $\mathscr{A} \xrightarrow{R} \mathscr{B} \xrightarrow{S} \mathscr{C}$, the set $(S \circ R)\langle A, C\rangle$ ought to be the disjoint union of the sets $R\langle A, B\rangle \times S\langle B, C\rangle$ modulo equalities resulting from $\mathscr{B}$-morphisms. This type of colimit 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 diappears, if we restrict ourselves to small categories.

4.9.02 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 **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 $R\langle C, D\rangle$ are independent from $\mathscr{C}$ and $\mathscr{D}$, the compositions ρ_L and ρ_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 ρ_L and/or ρ_R.

4.9.03 Example. Consider a cospan of functors $\mathscr{C} \xrightarrow{F} \mathscr{A} \xleftarrow{G} \mathscr{D}$ between locally small categories. Any sub-functor $\mathcal{R}$ of $\mathscr{C}^{\mathrm{op}} \times \mathscr{D} \xrightarrow{F^{\mathrm{op}} \times G} \mathscr{A}^{\mathrm{op}} \times \mathscr{A} \xrightarrow{\hom_{A}} \boldsymbol{set}$ yields a profunctor $\mathscr{C} \xrightarrow{R} \mathscr{D}$. Here by "sub-functor" we mean that $R\langle C, D\rangle \subseteq \mathscr{A}\langle F(C), G(D)\rangle$.

4.10 Ideals

A special case of Example 4.9.03 above deserves a special name:

4.10.00 Definition. Profunctors that arise as subfunctors of hom-functors[0] are called *2-sided ideals* or just *i*deals of $\mathscr{C}$.

4.10.01 Example. If a profunctor $\mathscr{C} \xrightarrow{R} \mathbf{1}$, *i.e.*, $\mathscr{C}^{\mathrm{op}} \times \mathbf{1} \xrightarrow{R} \boldsymbol{set}$ satisfies $R\langle C, 1\rangle \subseteq C/\mathscr{C}$, we call $\mathcal{R}$ a *left ideal*.

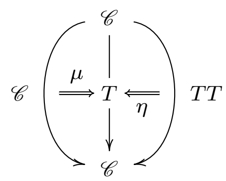
Similarly, if a profunctor $\mathbf{1} \xrightarrow{R} \mathscr{D}$, *i.e.*, $\mathbf{1}^{\mathrm{op}} \times \mathscr{D} \xrightarrow{R} \boldsymbol{set}$ satisfies $R\langle 1, D\rangle \subseteq \mathscr{D}/D$, we call $\mathcal{R}$ a *right ideal*.

[0] There has to be a natural transformation into the hom-functor that pointwise is an inclusion.

4.10.02 Remark. Profunctors $\mathscr{C} \xrightarrow{R} \mathbf{1} \xrightarrow{S} \mathscr{D}$ always compose to a profunctor $\mathscr{C} \xrightarrow{S \circ R} \mathscr{D}$.

If $\mathscr{C} = \mathscr{D}$, $\mathcal{R}$ is a left ideal and S is a right ideal, the cartesian product of $R\langle C, 1\rangle \times S\langle 1, C'\rangle$ always contains a subset of $\mathscr{C}\langle C, C'\rangle$. **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 4.2.03 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



subject to the ususal identity and associativity laws.

4.10.03 Definition. An monomorphism $I \xRightarrow{\iota} T$ is called a *left ideal*, if the composite 2-cell $IT \xRightarrow{\iota T} TT \xRightarrow{\mu} T$ factors through ι by means of a 2-cell $IT \xRightarrow{\lambda} I$.

Dual notion: *right ideal*, where $TI \xRightarrow{T\iota} TT \xRightarrow{\mu} T$ factors through ι by menas of a 2-cell $TI \xRightarrow{\rho} I$.

A *2-sided ideal*, or just *ideal*, is both a left- and a right-ideal.

4.10.04 Examples.

- ▷ Ordinary monoids are monads in the suspension of $\boldsymbol{set}$, 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 $\boldsymbol{ab}$ 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 $\boldsymbol{spn}$ of spans over sets, *cf.* Remark 4.7.01(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 neccessary 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 generalizaions 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 4.0.03 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.

4.10.05 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.*

4.10.06 Remarks.

- ▷ Since $\mathscr{C}$ may have several connected components, catgories of the form $\mathscr{C}(A)$ need not be enriched in $1 \downarrow$ ***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 intersectthe 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.

4.11 Sets, bags and tuples

Fundamentally, here are three ways how to form collections of elements of a set:

- ▷ as *sets*, where neither the order nor the (positive) multiplicity matters;
- ▷ as *bags* or *multisets*, 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 of finite, or unconstrained collections of elements of a given set. 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 12 monads.

Power-sets: We have already seen the (unconstrained) power-set monad $\boldsymbol{P} = \langle \mathbb{P}, \{-\}, \bigcup \rangle$ and its restriction $\boldsymbol{F}$ to finite sub-sets. The EM-algebras are $\bigsqcup$-semi-lattices, which are in fact complete, but where only superma need to be preserved by the morphisms in the infinite case, and $\sqcup$-semilattices in the finite case. The latter may be thought of as idempotent commutative monoids. Requiring the sub-sets in question 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 $\perp$ has to be dropped.

Power-bags: The singleton- and union-operations carry over from sets to multisets. But rather than a supremum-operation, where multiple occurrences of eleemts 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 free monoid monad $(-)^*$, resp., the free semigroup monad $(-)^+$, of non-empty tuples are considered. 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.]

As seen in the first sections, the free monoid 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.

These distributive laws carry over to the finite and infinite power-bag monads, where they take the form of distributivity of multiplication over addition.

4.11.00 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.

4.12 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.

4.12.00 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 *linarly 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 ervery $a \in A$. The set of all upper bounds af A is denoted by $A^{\uparrow}$. A *least upper bound*, or *supremum* $\bigsqcup A$ is an element of $A^{\uparrow}$ satisfying $\bigsqcup A \sqsubseteq A^{\uparrow}$.

Dual notions: *lower bound*; $A_{\downarrow}$; *greatest lower bound* or *infimum* $\bigsqcap A$.

(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 $f(x) \leq f(z)$ for all $x, y \in P$.

(3) Relations $P \xrightarrow{R} Q$ between pre-ordered sets are called *order-ideals*, if they satisfy $\sqsubseteq R\, le \subseteq R$, which bey the reflexivity of the orders is equivalent to the equality). Notice that order-preserving functions usually are not order-ideals, unless both orders are discrete.

4.12.01 Remark. Notice that the supremum, respectivley, infimum of $\emptyset \subseteq P$, if it exists, has to be a smallest, resp., largest element element of P, usually denotec by $\bot$, resp. $\top$.

4.12.02 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 $\sqcap$-*semilattice*. The slightly stronger notion of $\sqcap$-*semilattice* results from requiring all finite subsets to have a supremum, in particular the empty set.

Dual notion: $\sqcup$-*semilattice*.

(1) A poset that is both a $\sqcup$-and a $\sqcap$-semilattice is called a *lattice*.

(2) A lattice $\langle L, \sqsubseteq \rangle$ is called *complete*, if every subset $A \subseteq L$ has a supremum $\bigsqcap A$, or equivalently, every subset B has an infimum $\bigsqcap B$ (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 arbitray infima, and vice versa.

4.12.03 Examples.

- ▷ For every set X the power-set $\mathbb{P}(X)$ 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 $I\!N = \{0, 1, 2, \dots\}$ 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 infinum and supremum are given by $\gcd$ and lcd, respectively, while infinite suprema/infima do not exist.

4.12.04 Remark. The notions of $\sqcup$-semilattice 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 $\perp = @(\emptyset)$. 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 forumla $@\{x, y\} = y$. Of course, then @ has to be interpreted as the infimum operation for $\sqsupseteq$. Using a symbol like $\sqcup$ oer $\bigsqcup$ 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

Index

Jürgen Koslowski (koslowj@tu-bs.de)
Theoretical Computer Science
TU Braunschweig
Mühlenpfordstr. 23
D-38106 Braunschweig
Germany