Notes on Semigroups

Uday S. Reddy

March 14, 2014

Semigroups are everywhere. Groups are semigroups with a unit and inverses. Rings are "double semigroups:" an inner semigroup that is required to be a commutative group and an outer semigroup that does not have any additional requirements. But the outer semigroup must distribute over the inner one. We can weaken the requirement that the inner semigroup be a group, i.e., no need for inverses, and we have semirings. Semilattices are semigroups whose multiplication is idempotent. Lattices are "double semigroups" again and we may or may not require distributivity.

There is an unfortunate tussle for terminology: a semigroup with a unit goes by the name "monoid," rather than just "unital semigroup." This is of course confusing and leads to unnecessary divergence. The reason for the separate term "monoid" is that the unit is important. We should rather think of a semigroup as a "monoid except for unit," rather than the usual way of monoid as a "semigroup with unit."

Each kind of semigroup has actions. "Semigroup actions" and "group actions" are just that. Ring actions go by the name of modules (or, think "vector spaces"). Lattice actions don't seem to have their own name, yet.

There is more fun. A monoid, i.e., a semigroup with a unit, is just a one-object category. A monoid action is a functor from that category to an arbitrary category. This pattern repeats. A group is a one-object groupoid, i.e., a category with invertible arrows. A group action is again functor. A ring is a one-object semi-additive category (or, "additive category," depending on your terminology). A ring action, i.e., a module, is a functor from that category to another semi-additive category. So, monoids, groups and rings are degenerate forms of categories, and all their actions are degenerate forms of functors. So, most of Algebra is a study of degenerate categories of one form or another other. Or, to put it more positively, Category Theory is Algebra with multiple "types," if you see what I mean.

Monoid actions are automata, or "state machines." So, it would seem that group actions, modules and their cousins are equally "state machines." Now "state" is the business of Computer Science and of Physics. So, Computer Science and Physics have always been there, sitting inside Algebra, waiting for somebody to notice.

I want to say that the 20th century Algebra failed to bring out the fundamental unity of all these notions. I hope the 21st century will be different. This working notes is a rewrite of Algebra, bringing out the said fundamental unity, which might point us how to get there. The main texts I have used are [Eilenberg, 1974, Holcombe, 1982, Mac Lane and Birkhoff, 1967]. An advanced text [Clifford and Preston, 1961] is also used occasionally. The categorical perspective is obtained from [Street, 2007] and the ideas on categorification from [Mitchell, 1965].

Contents

1 Definitions

1.1 Basic notions A *groupoid* (also called a "magma") is a set with a binary operation (typically called "multiplication").

A *semigroup* is a groupoid whose multiplication operation is associative. It is called a *monoid* if the binary operation has a unit 1, which is necessarily unique whenever it exists. It is called a *group* if, in addition, every element has an inverse a^{-1} such that $a \cdot a^{-1} = a^{-1} \cdot a = 1$, which is again unique whenever it exists.

A *homomorphism* or *morphism* $f : A \to B$ of each of these structures preserves all the operations that exist. So, a semigroup morphism preserves multiplication, a monoid morphism preserves in addition the unit, and a group morphism preserves in addition the inverses. It turns out that groups have so much structure that we do not have to place the additional requirements, i.e., a semigroup morphism between groups is automatically a group morphism (§1.7).

A *logical relation* $R : A \leftrightarrow A'$ of these structures similarly respects all the operations that exist. More precisely, R is a logical relation of semigroups if it satisfies the requirement:

$$x \left[R\right] x' \wedge y \left[R\right] y' \implies xy \left[R\right] x'y' \tag{1.1}$$

That is, the multiplication operations of the two semigroups are related by $R \times R \to R$. A logical relation of monoids respects, in addition, the unit, i.e., the two units are related by R. A logical relation of groups respects, in addition, the inverses, i.e., the two inverse operations are related by $R \to R$. Be careful here. While a semigroup morphism between groups is automatically a group morphism, a semigroup logical relation between groups is not necessarily a logical relation of groups.

Other names used for logical relations are *relation of semigroups*, *compatible relation*, *regular relation* and *homogeneous relation*.

Every morphism $f : A \to B$ can be treated as a logical relation, by taking its function graph $\langle f \rangle : A \leftrightarrow B$, which is $\langle f \rangle = \{ (x, y) \mid f(x) = y \}$. The graph of the identity morphism is $\langle \mathrm{id}_A \rangle = I_A$ and composition of morphisms satisfies the "subsumption law:"

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ {\scriptstyle g}\downarrow & & \downarrow{\scriptstyle h} \\ A' & \xrightarrow{f'} & B' \end{array} \quad \Longleftrightarrow \quad \begin{array}{ccc} A & \xrightarrow{f} & B \\ {\scriptstyle \langle g \rangle}\updownarrow & & \updownarrow{\scriptstyle \langle h \rangle} \\ A' & \xrightarrow{f'} & B' \end{array} \tag{1.2}$$

The left hand side diagram is a commutative square of morphisms. The right hand side diagram says that f and f' map $\langle g \rangle$-related pairs to $\langle h \rangle$-related pairs.

The categories of semigroups, monoids and groups are denoted **SGrp**, **Mon** and **Grp** respectively. We also use the same notations for the corresponding reflexive graph-categories with logical relations as the edges.

1.2 Alternative formulations of logical relations In the semigroup theory literature, logical relations are treated using alternative notations but in a way equivalent to our definition. For example, [Eilenberg, 1976] defines a *relation of semigroups* as a relation $\varphi : A \to A'$ such that, treating φ as a map $\varphi : A \to \mathcal{P}A'$, we have

$$(x\varphi)(y\varphi) \subseteq (xy)\varphi \tag{1.3}$$

(Here, the evaluation of φ at x is written in "postfix" notation: $x\varphi$.)

$$\begin{array}{ccc} A \times A & \xrightarrow{\cdot} & A \\ {\scriptstyle \varphi \times \varphi}\downarrow & \subseteq & \downarrow{\scriptstyle \varphi} \\ A' \times A' & \xrightarrow{\cdot} & A' \end{array}$$

This means the same as a logical relation of semigroups $\varphi : A \leftrightarrow A'$. The relational formula $x \left[\varphi\right] x'$ says the same as $x' \in x\varphi$. Then

$$\begin{array}{rcl} (x\varphi)(y\varphi) & = & \{\, x'y' \mid x' \in (x\varphi) \wedge y' \in (y\varphi) \,\} \\ & = & \{\, x'y' \mid x \left[\varphi\right] x' \wedge y \left[\varphi\right] y' \,\} \end{array}$$

So, the requirement $(x\varphi)(y\varphi) \subseteq (xy)\varphi$ means that any $x'y'$ satisfying $x \left[\varphi\right] x' \wedge y \left[\varphi\right] y'$ also satisfies $xy \left[\varphi\right] x'y'$.

Eilenberg notes that, if φ is a function, it is a homomorphism of semigroups, which amounts to our subsumption law (1.2).

If φ is *total*, i.e., $x\varphi \neq \emptyset$ for all $x \in A$, then φ is called a *relational morphism* [Eilenberg, 1976, Ch. XII by Tilson] [Rhodes and Steinberg, 2009]. Relational morphisms are composable. If $\varphi : A \to A'$ and $\psi : A' \to A''$ are relational morphisms, so is $\varphi;\psi : A \to A''$. Tilson, Rhodes and Steinberg define a category of semigroups with relational morphisms.

1.3 Commutative semigroups A semigroup (monoid, group) is *commutative* if its binary operation is commutative. It is conventional to denote the commutative binary operation by $+$ instead of $\cdot$. The unit is written as 0 and the inverse as $-a$. Commutative groups are generally called "abelian groups."

The categories of commutative semigroups, monoids and groups are denoted **CSGrp**, **CMon** and **Ab** respectively.

1.4 Zero An element z of a groupoid A is called a *zero* if $za = z = az$ for all $a \in A$. It is a *left zero* if only $za = z$ holds. If a groupoid has a left zero and right zero then it must be the same, and hence a zero. A zero of a semigroup is unique if it exists.

A commutative semigroup may also have a "zero." We use some other symbol, say ε, so as to avoid conflict with 0, which is the unit of addition.

We use the notations $\mathbf{SGrp}_0$, $\mathbf{Mon}_0$, $\mathbf{Grp}_0$, $\mathbf{CSGrp}_0$, $\mathbf{CMon}_0$, $\mathbf{Ab}_0$ for the subcategories of their respective categories, whose objects have zero elements and morphisms preserve them.

In categories $\mathbf{Grp}_0$ and $\mathbf{Ab}_0$, only *nonzero* elements have inverses (because $zx = 1$ does not have any solutions for x). The morphisms preserve the inverses of all nonzero elements.

Note that, if $0_A = 1_A$ in a structure A, then the structure is *trivial* with 0 as its only element.

1.5 Initial, terminal and null objects In the category **SGrp**, the initial object **0** is the empty semigroup and the terminal object **1** is a one-element semigroup.

In **Mon** (and **Grp**), the initial object and the terminal object are the same, a one-element monoid (group). Such an object is called the *null object*, denoted conventionally as **0**. The

unique morphism $\mathbf{0} \to A$ maps the only element of $\mathbf{0}$ to the unit element 1_A. The unique morphism $A \to \mathbf{0}$ maps everything in A to the only element of $\mathbf{0}$.

A morphism $f : A \to B$ that factors through $\mathbf{0}$, i.e., $f = A \to \mathbf{0} \to B$ is called a *zero morphism*. There is evidently a unique zero morphism $A \to B$ (since $\mathbf{0}$ is both initial and terminal). In **Mon** and **Grp**, the zero morphisms are the constant functions given by $f(x) = 1_B$.

1.6 Duals If A is a semigroup, its dual semigroup A^{op} has the same underlying set, but its multiplication is reversed: $a \cdot b$ in A^{op} is $b \cdot a$ in A. If A is a monoid (or, group) its dual is similarly a monoid (group).

Due to the duality, most concepts in semigroup theory come in pairs. The dual of a concept χ in A is the concept χ in A^{op}.

A homomorphism $A^{\mathrm{op}} \to B$ is referred to as an *anti-homomorphism* from A to B.

1.7 Groups as special semigroups Groups can be regarded as special semigroups. If a semigroup A has unit, which is necessarily unique, and an inverse for every $x \in A$, which is again necessarily unique, then it is a group.

A semigroup homomorphism $h : A \to B$ beetween groups *A and B is automatically a group homomorphism.*

To see that h preserves the unit, suppose $h(1_A) = u$. Since $1_A = 1_A \cdot 1_A$, we obtain $h(1_A) = h(1_A) \cdot h(1_A)$, i.e., $u = uu$. By multiplying both sides by u^{-1}, we have $1_B = u$.

To see that h preserves the inverses (that exist in A), suppose $h(x) = y$ and $h(x^{-1}) = y'$. Since $xx^{-1} = 1_A$, we obtain $yy' = 1_B$. Similarly, $y'y = 1_B$. Thus y' is the two-sided inverse of y. ■

Hence **Grp** is isomorphic to a *full subcategory* of **SGrp**. In other words, the forgetful functor $G : \mathbf{Grp} \to \mathbf{SGrp}$ is full and faithful.

Since the above argument generalizes to all inverses that exist in A, we may also note that $\mathbf{Grp}_0$ is a full subcategory of $\mathbf{SGrp}_0$ and $\mathbf{Mon}_0$.

Grp ($\mathbf{Grp}_0$) is also a full subcategory of **Mon** ($\mathbf{Mon}_0$) in a similar fashion.

In contrast, **Mon** is *not* a subcategory of **SGrp**. A semigroup homomorphism $h : A \to B$ between monoids does not have to preserve the unit. However, the image $h[A] \subseteq B$ will be a monoid with its own "unit," i.e., a distinguished element $e = h(1_A)$ that satisfies $b \cdot e = b = e \cdot b$ for all $b \in h[A]$.

Some authors (implicitly) consider a category of semigroups, which we denote $\mathbf{SGrp}^{\bullet}$, whose arrows are semigroup homomorphisms $h : A \to B$ that preserve the units whenever present, i.e., if A has a unit then B has a unit and $h(1_A) = 1_B$. This gives rise to full subcategories $\mathbf{Grp} \subseteq \mathbf{Mon} \subseteq \mathbf{SGrp}^{\bullet}$.

1.8 Lemma (Dickson) *If a semigroup A has a left unit e and a left inverse for every element $x \in A$, then it is a group.* (In other words, e is also the right unit and each left inverse is also a right inverse.)

Let $a \in A$ be an arbitrary element. Suppose b is the left inverse of a, i.e., $ba = e$. Then, $bab = eb = b$. Multiplying on the left by a left inverse of b yields $ab = e$. Thus, b is also the right inverse of a.

For the same a, we have $ae = aba = ea = a$. Quantifying over all $a \in A$, we obtain the conclusion that e is a right unit as well. ■

Dually, we can assume a right unit and right inverses to conclude that a semigroup is a group.

1.9 Lemma (Weber-Huntington) *A semigroup A that satisfies the axiom: for all $a, b \in A$, there exist $x, y \in A$ such that $ax = b$ and $ya = b$, is in fact a group.*

Consider an arbitrary element $a \in A$. First there exists an element $e_a \in A$ such that $ae_a = a$. For any $b \in A$, there exists $y \in A$ such that $ya = b$. Hence, $b = ya = yae_a = be_a$. Thus, e_a is, in fact, a right unit of the semigroup. Write it as e. Second, since there is an x such that $ax = e$, we have a right inverse for a.

Thus we have a semigroup with a right unit and a right inverse for each element, which is then a group by the Dickson's axiom. ■

Using internal products (§2.10), this statement can be summarized as follows:

A semigroup A is a group if and only if $aA = A = Aa$ for every element $a \in A$.

Clearly, $aA \subseteq A$. If the reverse inclusion $A \subseteq aA$ is satisfied then, for every $b \in A$ there exists $x \in A$ such that $ax = b$. Similarly, $A \subseteq Aa$ implies that there exists $y \in A$ such that $ya = b$. Thus A is a group.

Conversely, if A is a group, it clearly satisfies $aA = A$ because $aa^{-1}b = b$.

1.10 Incidental groups When **Grp** and **SGrp** are regarded as reflexive graph-categories, the forgetful functor $G : \mathbf{Grp} \to \mathbf{SGrp}$ is *not* full. A semigroup logical relation $R : A \leftrightarrow B$ between groups has no reason to be a logical relation of groups. The full reflexive graph-subcategory of **SGrp** whose (vertex) objects are groups will be denoted $\mathbf{Grp}'$. Its objects will be called *incidental groups*. Fields (§9.3) are an example of incidental groups internal to **Ab**.

1.11 Monoid generated by a semigroup If A is a semigroup, the monoid generated by A is $A^I = A \uplus \{1\}$ with an additional formal element 1 and the multiplication of A is extended to A^I in the natural way. This operation is possible even if A is already monoid. In that case 1_A ceases to be the unit element of A^I and the new unit 1 takes its place. The evident functor that sends $A \mapsto A^I$ is left adjoint to the forgetful functor $G : \mathbf{Mon} \to \mathbf{SGrp}$.

In addition, the forgetful functor $\mathbf{Mon} \to \mathbf{SGrp}^\bullet$ has a left adjoint defined as follows. If A is a semigroup, but not a monoid, then we turn it into a monoid $A^\bullet = A \uplus \{1\}$ by adjoining a new a unit element. If A already has a unit element, then $A^\bullet = A$.

1.12 Group generated by a monoid If A is a monoid, the group generated by A is obtained by first considering the free group generated by $|A|$ and quotienting it by the group congruence relation of the multiplication table of A. This is called the *enveloping group* or the *group completion* of A.

For a semigroup S, the group generated by S is obtained by comibing these constructions. We first adjoin a unit element to obtain a monoid S^1 and then find the enveloping group of S^1. If S already has a unit element, it will collapse to the adjoined unit. Therefore, we could also describe the construction as the enveloping group of $S^\bullet$.

1.13 Examples: transformations For any set X, the set of functions $T(X) = [X \to X]$ is a semigroup under composition.[1] It is also a monoid because the identity function serves as the unit. The dual $[X \to X]^{\mathrm{op}}$ has the same elements, but its multiplication is sequential composition (";").

In any category, the set of endomorphisms of an object $\mathrm{End}(X) = \mathrm{Hom}(X, X)$ forms a monoid. The special cases of interest include the category of sets and partial functions, where $\mathrm{End}(X) = \mathbf{Pfn}(X, X)$ models computations that have a possibility of divergence, and the category of sets and relations, where $\mathrm{End}(X) = \mathbf{Rel}(X, X)$ models nondeterministic computations. Note that the empty partial function (relation) forms the *null object* in these monoids.

Similarly, the endomorphisms in the category of posets and monotone functions (**Poset**), that of cpo's and continuous functions (**CPO**) and so on are also examples of monoids. If the category is that of strict functions ($\mathbf{CPO}_\perp$), then the least element $\bar{\perp} \in \mathrm{End}(X)$ is the zero element.

For groups, one must restrict to $\mathrm{Aut}(X) \subseteq \mathrm{End}(X)$, which is the set of isomorphisms $X \to X$, also called *automorphisms*. (Arbitrary transformations in $\mathrm{End}(X)$ have no reason to have inverses, and so $\mathrm{End}(X)$ is not a group.) If X is a set, the automorphism are called *permutations*.

1.14 Examples: computations Let A denote the set of actions that can be performed by a state machine, i.e., each action $a \in A$ transforms the state of the machine. Then a natural operation to consider is that of *sequential composition*: doing two actions one after another, denoted $a; b$. We envisage that the actions of the machine are closed under sequential composition, and note that sequential composition should be associative. Thus A forms a semigroup. If we adjoin a unit element, typically written as **skip**, we obtain a monoid.

1.15 Examples: numerical The set of integers $\mathbb{Z}$ under addition forms a commutative group (and hence a semigroup, monoid and group). The set of integers modulo n, denoted, $\mathbb{Z}_n$, also forms a commutative group. The evident homomorphism $\mathbb{Z} \to \mathbb{Z}_n$ is in fact an epimorphism. Commutative groups (semigroups, monoids) whose binary operation is "addition" are referred to as "additive" groups (semigroups, monoids).

The set of natural numbers $\mathbb{N}$ under multiplication forms a commutative monoid. Every $\mathbb{Z}_n$ is similarly a commutative monoid under multiplication. The evident homomorphism $\mathbb{N} \to \mathbb{Z}_n$ is an epimorphism.

The set of real numbers $\mathbb{R}$ under addition forms a commutative group. The set $\mathbb{R}_+$ of positive real numbers forms a commutative group under multiplication. The function $\log_e : \mathbb{R}_+ \to \mathbb{R}$ is a homomorphism of groups because $\log_e(xy) = \log_e(x) + \log_e(y)$. It is in fact an isomorphism, with the inverse $x \mapsto e^x$, which is also a morphism of groups.

Note that $\mathbb{R}$ is not a group under multiplication because 0 does not have an inverse.

1.16 Examples: free structures The free semigroup generated by a set X is X^+, the set of all nonempty sequences over X. The free monoid is $X^* = \{\epsilon\} \cup X^+$, the set of all sequences over X.

[1] In semigroup theory, sequential composition is often used as the multiplication for transformations, which switches the left-right terminology.

The free semigroup with one generator is the set of positive natural numbers (which isomorphic to $\mathbf{1}^+$). The free monoid with one generator is $\mathbb{N}$. The free group with one generator is $\mathbb{Z}$, the set of integers. Note that they are all commutative. (With a single generator they are necessarily so.)

1.17 Examples: ordered A commutative semigroup whose binary operation is idempotent, i.e., $a + a = a$, is called a *semilattice* or a *commutative band*. It is conventional to write its binary operation as the "join" $a \vee b$, and the unit, if it exists, as $\bot$. The binary operation determines a partial order $a \leq b \iff a \vee b = b$. A unit $\bot$ would be the least element in the partial order. If the semilattice does not have a least element, we may freely adjoin one. (Cf. §1.11.) It is then a *unital semilattice*.

The "maximum" operation in any linearly ordered set of numbers ($\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$) gives an example of a semilattice. Any powerset $\mathcal{P}X$ is a semilattice under union. The induced partial order is the subset order.

One might also write the binary operation of a semilattice as the "meet" $a \wedge b$, and the unit, if it exists, as $\top$. The partial order induced by a meet is defined by $a \leq b \iff a \wedge b = a$. These are called *meet-semilattices* to distinguish them from the above *join-semilattices*. The "minimum" operation in linearly ordered sets of numbers and the intersection in powersets give examples of meet-semilattices.

A *lattice* is a set with both join-semilattice and meet-semilattice structures and, additionally, satisfying the following absorption laws:

$$a \wedge (a \vee b) = a = a \vee (a \wedge b)$$

The absorption laws ensure that both the semilattice structures determine the same partial order: $a \leq b \iff a \vee b = b \iff a \wedge b = a$.

1.18 Categories as generalized monoids A category is a generalization of a monoid where the elements have types of the form $a : X \to Y$. If $a : X \to Y$ and $b : Y \to Z$ then $a; b : X \to Z$, where we write "multiplication" as ";".

For example, $n \times n$ matrices form a monoid under matrix multiplication with the identity matrix as the identity. More generally, $m \times n$ matrices for a category. The objects are positive integers and an $m \times n$ matrix a is given the type $a : n \to m$ (or, more suggestively written $a : m \leftarrow n$).

Monoids are categories with a unique object, which is normally written as $\star$. All the morphisms $a : \star \to \star$ are then the elements of the monoid. This approach of treating monoids as one-object categories is an instance of "categorification." (Cf. Sec. 16).

1.19 Monoids in monoidal categories Let $(\mathcal{C}, \otimes, I)$ be a monoidal category. A *monoid (object)* in $\mathcal{C}$ is an object M, equipped with morphisms $\mu : M \otimes M \to M$ and $\eta : I \to M$, satisfying the evident equational laws at the level of morphisms. Ordinary monoids are monoids in **Set**.

Monoids in **Poset** may be called *ordered monoids* and monoids in $\mathbf{CPO}_\bot$ may be called *complete ordered monoids*. Monoids in **CSLat**, the category of complete semilattices, are called *quantales*.

Monoids in the monoidal category (**Ab**, $\otimes_{\mathbb{Z}}$, $\mathbb{Z}$) are *rings* (§1.20). Note that multiplication $\cdot : A \otimes_{\mathbb{Z}} A \to A$ must be a morphism in **Ab**, i.e., it should be a bihomomorphism $A \times A \to A$, which gives the left and right distributivity laws: $x \cdot (y + z) = xy + xz$ and $(y + z) \cdot x = yx + zx$.

Monoids in the monoidal category (**CMon**, $\otimes_{\mathbb{N}}$, $\mathbb{N}$) are *semirings* (§1.20).

Similarly, monoids in the category (K-**Mod**, $\otimes_K$, K) are *K-algebras* (§7.31).

Monoids in the functor category ($[\mathcal{C}, \mathcal{C}]$, $\circ$, Id) are monads in $\mathcal{C}$.

1.20 Semirings and Rings A *semiring* A has two groupoid structures: $(+, 0)$ which forms a commutative monoid (the "additive" monoid) and $(\cdot, 1)$ which forms a monoid (the "multiplicative" monoid). In addition, multiplication distributes over addition. This distributivity includes the nullary case: $a \cdot 0 = 0 = 0 \cdot a$. Thus, the unit of the additive monoid is always a *zero element* for the multiplicative monoid.

If the additive monoid has inverses, making it a commutative *group*, then A is called a *ring*. If the additive monoid is idempotent, i.e., $a + a = a$ for all a, then it is called an *idempotent semiring* or *diod*.

Example: The set $\mathrm{End}(A)$ of of endomorphisms of a commutative monoid A forms a semiring, with pointwise addition, pointwise zero, composition and the identity endomorphism. If A is a commutative group then $\mathrm{End}(A)$ is a ring. If A is a semiring, its underlying commutative monoid $|A|$ gives rise to an endomorphism semiring $\mathrm{End}(|A|)$ in the same way.

Example: The set of $n \times n$ matrices over a semiring A is again a semiring, with pointwise addition, pointwise zero, matrix multiplication and the unit matrix.

1.21 Commutative semirings and rings Note that the additive monoid of a semiring or a ring is *always* commutative.

- If the multiplicative monoid is also commutative, then the structures are called *commutative semiring* and *commutative ring* respectively.
- If, in addition, the multiplicative monoid has inverses for all values except 0, then the structures are called *semifield* and *field* respectively.

In a sense that can be made precise using category theory a semiring is nothing but a "monoid with addition" and a ring is a "monoid with invertible addition."

Example: The set of natural numbers, $\mathbb{N}$, is a semiring under the natural addition and multiplication operations. The set of integers, $\mathbb{Z}$, is a ring. The set of rationals, $\mathbb{Q}$, the set of reals, $\mathbb{R}$, and the set of complex numbers, $\mathbf{C}$, are fields.

Example: A powerset $\mathcal{P}X$ is a semiring with the union regarded as "addition" and intersection as "multiplication." The distributivity law is: $a \cap (b \cup c) = (a \cap b) \cup (a \cap c)$. It is also a semiring with the roles of union and intersection reversed. A structure that is both a join-semilattice and a meet-semilattice with units and the two distributive laws hold is called a *lattice*.[2]

Example: For a commutative semiring K, a *polynomial* in a single variable x with coefficients from K is a formal sum $f = \sum_{i=0}^{\infty} a_i x^i$ with finitely many non-zero coefficients.

[2] This can't be right. Needs to be checked.

If $g = \sum_{i=0}^{\infty} b_i x^i$ is another such polynomial, addition is defined pointwise:

$$f + g \;=\; \sum_{i=0}^{\infty} (a_i + b_i) x^i$$

Multiplication is defined as the *convolution* product:

$$f * g \;=\; \sum_{k=0}^{\infty} \sum_{i+j=k} a_i b_j x^k$$

This gives a semiring of polynomials denoted $K[x]$. More formally, a polynomial is just a sequence $\mathbb{N} \to K$ which is "almost everywhere zero," i.e., non-zero for only finitely many positions.

If K is ring, i.e., has additive inverses, then $K[x]$ is a ring. The additive inverses are obtained by inverting the coefficients, $\sum_{i=0}^{\infty}(-a_i)x^i$.

Example: A *formal power series* is like a polynomial, but we do not insist that it has finitely many non-zero coefficients. Formal power series still form a semiring $K[\![x]\!]$, but their evaluation for particular values of x may not converge unless additional convergence axioms are assumed.[3]

Example: More generally, for any semiring A, *not necessarily commutative*, and a monoid M, the *monoid semiring* $A[M]$ consists of formal sums $f = \sum_{x \in M} a_x x$ with finitely many non-zero coefficients. (If A is commutative then $A[M]$ is commutative.) The convolution product is now defined as

$$f * g \;=\; \sum_{x \in M} \sum_{y \in M} a_x b_y xy \;=\; \sum_{z \in M} \sum_{xy = z} a_x b_y z$$

Once again, the formal treatment of such formal sums is as functions $M \to A$ which are almost everywhere zero. The examples of polynomials and formal power series are special cases of this with the monoid being $\mathbb{N}$ under addition, but viewed multiplicatively as formal powers x^n. Polynomials or formal power series with k variables also form special cases with the monoid being $\mathbb{N}^k$.

Example: The *max-plus algebra* has the carrier $\mathbb{R} \cup \{-\infty\}$ with the maximum operation ($\vee$) as the "addition" (with unit $-\infty$) and addition as the "multiplication" (with 0 as the unit). Note that addition distributes over maximum:

$$x + (y \vee z) \;=\; (x + y) \vee (x + z) \qquad (x \vee y) + z \;=\; (x + z) \vee (y + z)$$

The max-plus algebra is an idempotent semiring or a "dioid."

Similar algebras exist for other linearly ordered collections of numbers, including $\mathbb{Z}$, $\mathbb{Z}^+$, $\mathbb{Q}$, $\mathbb{Q}^+$ etc. Given any linearly ordered additive group G, one can adjoin a formal element ε, with the ordering $\varepsilon \leq x$ for all x, to obtain a max-plus algebra for $\overline{G} = G \cup \{\varepsilon\}$. *Min-plus* algebras (also called "tropical semirings") are defined in a dual fashion.

When using positive numbers, $\mathbb{Z}^+$, $\mathbb{Q}^+$, $\mathbb{R}^+$ etc., it is also possible to use ordinary multiplication as the "multiplicative" monoid because multiplication distributes over maximum:

$$x \cdot (y \vee z) \;=\; xy \vee xz \qquad (x \cdot y) \vee z \;=\; xz \vee yz$$

[3] Does K have to be a ring here?

In this case the number 0 can be used as the max-unit ε. The real interval $[0, 1]$ is also a tropical semiring with 0 as the max-unit and 1 as the multiplicative unit.

Have these examples already appeared above?

Example: The set of $n \times n$ matrices over a semiring A is again a semiring, with pointwise addition, pointwise zero, matrix multiplication and the unit matrix.

Example: The set $\mathrm{End}(A)$ of semiring endomorphisms forms a semiring, with pointwise addition, pointwise zero, composition and the identity endomorphism. Note that the multiplication of A does *not* play any role in this construction.

1.22 Bisemigroups and double semigroups A set with two associative binary operations $*_h$ and $*_v$ has been termed a *bisemigroup*. If each binary operation has a unit, then it is a *bimonoid*. (This terminology conflicts with that of "bialgebra," §7.33.)

A bisemigroup that satisfies the interchange law:

$$(a *_v b) *_h (c *_v d) = (a *_h c) *_v (b *_h d)$$

is called a *double semigroup*, in analogy with "double categories." If both the operations have units, they are forced to be the same. For example, by setting $a = 1_v$, $b = 1_h$ and $d = 1_v$, we obtain $1_v *_h c = c$, and similarly for other identities. The Eckman-Hilton argument [Eckmann and Hilton, 1962, Kock, 2007] shows that the two monoids are forced to be *commutative* as well!

Internal to **Poset**, it is possible to use a lax version of the interchange law due to Hoare [Hoare et al., 2011a]:

$$(a *_v c) *_h (b *_v d) \leq (a *_h b) *_v (c *_h d)$$

This does not force the above degeneracies. In particular, the two units have no reason to be identical. For example, using the valuation given above, we obtain an inequality $c \leq 1_v *_h c$. Likewise, one can obtain $a *_v 1_h \leq a$, $1_h \leq 1_v$ and $1_v \leq 1_v *_h 1_v$..

1.23 Concurrent and locality bimonoids We focus on the lax double monoids in **Poset** and assume that the "vertical" monoid is commutative. Interpret $*_v$ as a form of concurrent composition and $*_h$ as a form of sequential composition. A *concurrent monoid* is defined to be a lax double monoid where $1_h = 1_v$ [Hoare et al., 2011b].

An element a in a lax double monoid is said to be *local* if $a *_v 1_h = a$. A lax double monoid is said to be a *locality bimonoid* if 1_h is local, i.e., $1_h *_v 1_h = 1_h$.

The following properties are equivalent in a lax double monoid:

1. 1_v is local.
2. All elements are local.
3. 1_h is a unit of $*_v$.
4. $1_v = 1_h$.

Hence, any one of these properties turns a locality bimonoid into a concurrent monoid.

The map $a \mapsto a *_v 1_h$ in a locality bimonoid A is idempotent and is called the *localizer*. The image of the localizer, called the *local core* A_{loc}, is a concurrent monoid.

1.24 Representations For an object X in any category $\mathcal{C}$, the collection $\mathrm{End}(X) = \mathrm{Hom}(X, X)$ is a monoid. A morphism of semigroups (or monoids) $\delta : A \to \mathrm{End}(X)$ is called a *representation* of the semigroup (monoid) A by transformations of X.

A morphism $A \to \mathrm{End}(X)^{\mathrm{op}}$, or, equivalently, from $A^{\mathrm{op}} \to \mathrm{End}(X)$ is referred to as an *anti-representation*. This means essentially that the multiplication in A corresponds to the sequential composition of morphisms $f; g$ in $[X \to X]$.

For groups, similarly, a homomorphism $\delta : A \to \mathrm{Aut}(X)$ is called a *group representation*, and $\delta : A \to \mathrm{Aut}(X)^{\mathrm{op}}$ an *anti-representation*.

Categorification treats a monoid A as a one-object category. Then a representation of A in $\mathcal{C}$ is merely a *functor* $A \to \mathcal{C}$. Similarly, a group A is a one-object groupoid category and a group representation is a functor $A \to \mathcal{C}$. There is evidently a category of all such functors $[A, \mathcal{C}]$ or $\mathcal{C}^A$. The morphisms are natural transformations, which coincide with *A-linear maps* described below.

1.25 Actions and modules A representation in **Set**, viewed as the uncurried binary operation $\delta^\sharp : A \times X \to X$ is called an *A-action* (more specifically, a *left* A-action). The pair $X = \langle X, \delta^\sharp \rangle$ is called a *left A-module*. The binary operation satisfies the laws:

$$\begin{array}{rcl} \delta^\sharp(ab, x) & = & \delta^\sharp(a, \delta^\sharp(b, x)) \\ \delta^\sharp(1_A, x) & = & x \end{array}$$

In fact, these laws characterize actions *fully*, i.e., if $\delta^\sharp$ is any function satisfying these laws then it is the uncurrying of a representation.

The curried homomorphism $\delta^\sharp$ is often written as a binary operator $\cdot : A \times X \to X$ which makes the equations look more attractive:

$$\begin{array}{rcl} ab \cdot x & = & a \cdot (b \cdot x) \\ 1_A \cdot x & = & x \end{array}$$

In this context, the elements of A are called *scalars* and the operation "$\cdot$" is called *scalar multiplication* (as opposed to the unwritten operation in "ab" which is the ordinary multiplication of A). We use the notation ${}_AX$ to indicate the fact that we are treating X as a left A-module.

So, representations, actions and modules are the same concept. They also have many other names in the literature: *A-sets* (more commonly, *M-sets* for monoids M and *G-sets* for groups G), *A-operands* [Clifford and Preston, 1961], *A-acts* [Kilp et al., 2000], *A-polygons*, *A-systems* [Howie, 1976], *A-automata* and *transition systems*.

A homomorphism of A-modules $f : {}_AX \to {}_AY$, called an *A-linear map* or an *A-homogeneous map*, is a function $f : X \to Y$ that preserves the scalar multiplication:

$$f(a \cdot x) = a \cdot f(x)$$

A logical relation of A-modules $R : {}_AX \leftrightarrow {}_AY$ is called a *bisimulation relation*. It satisfies:

$$a \in A \wedge x \; [R] \; y \implies a \cdot x \; [R] \; a \cdot y$$

That is, the two scalar multiplications are related by $I_A \times R \to R$. The category of A-modules is denoted A-**Mod** and so is its corresponding reflexive graph category.

1.26 Right actions (right modules) Dually, an anti-representation $\delta : A^{\mathrm{op}} \to \mathrm{End}(X)$, viewed as a binary operation $\cdot : X \times A \to X$ is a *right A-action*, and $X_A = \langle X, \cdot \rangle$ is a *right A-module*. The binary operation satisfies the laws:

$$\begin{aligned} x \cdot ab &= (x \cdot a) \cdot b \\ x \cdot 1_A &= x \end{aligned}$$

Note that "left" vs "right" is not merely a matter of notation; it refers to the orientation of composition. In a left action, the multiplication of the semigroup is interpreted as standard composition $f \circ g$. In a right action, it is interpreted as sequential composition $f; g$.

A homomorphism of right A-modules, called a *right A-linear map* $f : X_A \to Y_A$ is a function $f : X \to Y$ that preserves the scalar multiplication on the right:

$$f(x \cdot a) = f(x) \cdot a$$

Similarly, a *bisimulation* of right A-modules $R : X_A \leftrightarrow Y_A$ is a relation that makes the two scalar multiplications related by $R \times I_A \to R$. The category of right A-modules is denoted **Mod**-A and so is its corresponding reflexive graph category.

1.27 Bimodules Given monoids A and B, an (A, B)-*bimodule* is a set X with an action of A on the left and an action of B on the right, satisfying:

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

We use the notation ${}_AX_B$ to signify the fact that we are treating X as an (A, B)-bimodule.

Using the notion of product monoids (§3.1), an (A, B)-bimodule may be seen to be the same as an $(A \times B^{\mathrm{op}})$-module, or a representation $\delta : A \times B^{\mathrm{op}} \to \mathrm{End}(X)$. The coherence condition above is merely an unraveling of the homomorphism condition of δ. Since $(a, 1_B) \cdot (1_A, b) = (a, b) = (1_A, b) \cdot (a, 1_B)$, we expect $\delta_{(a,1_B)}(\delta_{(1_A,b)}(x)) = \delta_{(1_A,b)}(\delta_{(a,1_B)}(x))$, i.e.,

$$\lambda_a(\rho_b(x)) = \rho_b(\lambda_a(x))$$

1.28 Group actions A *representation* of a group A is a semigroup homomorphism $\delta : A \to \mathrm{Aut}(X)$, which automatically becomes a group homomorphism. Thus, the uncurried homomorphism $\cdot : A \times X \to X$ satisfies the additional property:

$$a^{-1} \cdot (a \cdot x) = x$$

which follows from the laws of semigroup actions. Left-, right- and bi-modules of groups are now special cases of those of semigroups.

1.29 Regular representations and self-actions A semigroup A can be represented in the underlying set $|A|$ in a canonical way, by left multiplication $\lambda_a(x) = a \cdot x$. This gives a homomorphism $\lambda : A \to \mathrm{End}(|A|)$ because $\lambda_{a_1 a_2}(X) = (a_1 a_2)x = a_1(a_2 x) = (\lambda_{a_1} \circ \lambda_{a_2})(x)$. This is called the *left regular representation* of A. Viewed as an action $\cdot : A \times |A| \to |A|$, the scalar multiplication is nothing but left multiplication by the semigroup A. We write the resulting A-module as ${}_AA$ even though ${}_A|A|$ would be more accurate.

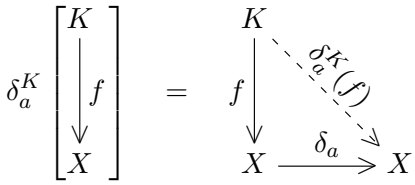
Similarly, right multiplication $\rho_a(x) = x \cdot a$ gives an *anti-representation* of A in $|A|$, where the multiplication is sequential composition: $\rho_{a_1 a_2}(x) = x(a_1 a_2) = (x a_1)a_2 = (\rho_{a_1}; \rho_{a_2})(x)$. (Note

that right multiplication does *not* give a representation $A \to \mathrm{End}(|A|)$. So, the distinction between left and right is unavoidable.)

Viewed in terms of modules, the underlying set of every semigroup (monoid, group) A automatically has a *bimodule structure* ${}_A|A|_A$, by just the restriction of multiplication..

A group A can be represented in $|A|$ by left multiplication as well. This means that the semigroup homomorphism $\lambda : A \to \mathrm{End}(|A|)$ cuts down to $A \to \mathrm{Aut}(|A|)$, i.e., λ_a is a bijection on $|A|$ with the inverse $\lambda_{a^{-1}}$. Secondly, since semigroup morphisms are automatically group homomorphisms, we obtain $\lambda_{1_A} = \mathrm{id}_A$ and $\lambda_{a^{-1}} = (\lambda_a)^{-1}$.

1.30 Higher-order representations (covariant) If a monoid A has a representation in X then it also has a representation in X^K for any fixed set K. Define $\delta^K : A \to \mathrm{End}(X^K)$ by $\delta^K_a = (\delta^K)_a = \delta_a \circ -$, or, more vividly,



We call it the *covariant extension* or *post-composition extension* of δ. The fact that it is a homomorphism is obvious from the diagram. (Imagine pasting two triangles corresponding to δ_{a_1} and δ_{a_2}.) Here is an explicit calculation:

$$\delta^K_{a_1 a_2} = \delta_{a_1 a_2} \circ - = \delta_{a_1} \circ \delta_{a_2} \circ - = \delta^K_{a_1}(\delta^K_{a_2}(-)) = \delta^K_{a_1} \circ \delta^K_{a_2}$$
$$\delta^K_1 = \delta_1 \circ - = \mathrm{id}_X \circ - = \mathrm{id}_{X^K}$$

Viewing the representation as a left action $\cdot : A \times X \to X$, we note that it extends to a left action on X^K, *viz.*, $\bullet : A \times X^K \to X^K$ given by

$$(a \bullet f)(k) = a \cdot f(k)$$

In other words, a left A-module ${}_AX$ gives rise to a left A-module ${}_A(X^K)$ for any set K.

In particular, we obtain a notion of *product modules* ${}_A(X^n) = {}_A(X \times \cdots \times X)$ with scalar multiplication given by $a \bullet (x_1, \ldots, x_n) = (a \cdot x_1, \ldots, a \cdot x_n)$.

Similarly, if A has an anti-representation in X (or a right action on X) then it has an anti-representation in X^K (or a right action on X^K). The definition is the same as the above, but we write it using sequential composition notation: $(\delta^K)_b = -; \delta_b$. The homomorphism property of $\delta^K_b = (\delta^K)_b$ is verified by:

$$\delta^K_{b_1 b_2} = -; \delta_{b_1 b_2} = -; \delta_{b_1}; \delta_{b_2} = \delta^K_{b_2}(\delta^K_{b_1}(-)) = \delta^K_{b_1}; \delta^K_{b_2}$$

Correspondingly, a right action $\cdot : X \times A \to X$ extends to a right action $\bullet : X^K \times A \to X^K$, given by

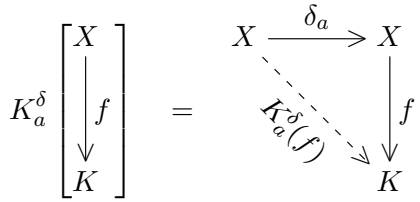
$$(f \bullet b)(x) = f(x) \cdot b$$

Combining the two features, if $A \times B^{\mathrm{op}}$ has a representation in X, making a bimodule ${}_AX_B$, then $A \times B^{\mathrm{op}}$ has a representaiton in X^K, giving a bimodule ${}_A(X^K)_B$. The two scalar multiplications are given by

$$\begin{aligned}(a \bullet f)(k) &= a \cdot f(k)\\ (f \bullet b)(k) &= f(k) \cdot b\end{aligned}$$

The coherence condition of X extends to that of X^K.

1.31 Higher-order representations (contravariant) If A has a representation in X then it gives rise to an *anti-representation* in K^X for any fixed set K. (Note that K^X is contravariant in X and, so, the sign of the representation is switched in moving from X to K^X.) The anti-representation is defined by $K^\delta_a = (K^\delta)_a = \delta_a; -$. Diagrammatically:



We call it the *contravariant extension* or *pre-composition extension* of δ.

To verify that it is a homomorphism, note:

$$K^\delta_{a_1a_2} = \delta_{a_1a_2}; - = (\delta_{a_1} \circ \delta_{a_2}); - = \delta_{a_2}; \delta_{a_1}; - = K^\delta_{a_2}(K^\delta_{a_1}(-)) = K^\delta_{a_1}; K^\delta_{a_2}$$

Correspondingly, a left action $\cdot : A \times X \to X$ determines a *right* action $\bullet : K^X \times A \to K^X$, given by

$$(f \bullet a)(x) = f(a \cdot x)$$

It is instructive to write the homomorphism condition in this notation and notice how the sign of the action changes in going from ${}_AX$ to $(K^X)_A$:

$$(f \bullet a_1 \bullet a_2)(x) = (f \bullet a_1)(a_2 \cdot x) = f(a_1 \cdot a_2 \cdot x)$$

Dually, an anti-representation of A in X (or a right action on X) determines a representation in K^X (or a left action on K^X). The definition is the same as the above. At the level of actions, the left action is given by

$$(b \bullet f)(x) = f(x \cdot b)$$

Combining the two features, if $A \times B^{\mathrm{op}}$ has a representation in X, making it a bimodule ${}_AX_B$, then K^X has an anti-representation of $A \times B^{\mathrm{op}}$ (or, equivalently, a representation of $A^{\mathrm{op}} \times B$). This gives a bimodule ${}_B(K^X)_A$. The two scalar multiplications are given by

$$\begin{array}{rcl} (b \bullet f)(x) & = & f(x \cdot b) \\ (f \bullet a)(x) & = & f(a \cdot x) \end{array}$$

1.32 Higher-order representations for groups A representation of a group A in X, being a homomorphism $\delta : A \to \mathrm{Aut}(X)$ has similar higher-order extensions:

$$\delta^K : A \to \mathrm{Aut}(X^K) \qquad K^\delta : A \to \mathrm{Aut}(K^X)$$

The *covariant extension* is given by $\delta^K_a = \delta_a \circ - : X^K \to X^K$. First, note that δ^K_a is an *automorphism*. If $\delta^K_a(f) = \delta_a \circ f = g$, we can recover f as $f = (\delta_a)^{-1} \circ g$ using the fact that δ_a is an automorphism of X. Secondly, δ is a homomorphism of groups and, so, $(\delta_a)^{-1} = \delta_{a^{-1}}$. We can write f as $f = \delta_{a^{-1}} \circ g = \delta^K_{a^{-1}}(g)$. Thus, the inverse of δ^K_a is $\delta^K_{a^{-1}}$, ergo, δ^K is a homomorphism of groups. Summarizing:

$$\delta^K_{a^{-1}} \;=\; \delta_{a^{-1}} \circ - \;=\; (\delta_a)^{-1} \circ - \;=\; (\delta^K_a)^{-1}$$

Note that a^{-1} is the inverse in A, $(\delta_a)^{-1}$ is the inverse in $\mathrm{Aut}(X)$ and $(\delta^K_a)^{-1}$ is the inverse in $\mathrm{Aut}(X^K)$.

The *contravariant extension* is given by $K^\delta_a = - \circ \delta_a$. Once again, K^δ_a is an automorphism, with the inverse of $g \in K^X$ given by $g \circ (\delta_a)^{-1}$. K^δ is then a group homomorphism because:

$$K^\delta_{a^{-1}} \;=\; - \circ \delta_{a^{-1}} \;=\; - \circ (\delta_a)^{-1} \;=\; (K^\delta_a)^{-1}$$

As in the case of monoids, we can write these representations as actions. The covariant extension gives a left action and the contravariant extensions gives a right action:

$$\begin{array}{rcl} (a \bullet f)(k) & = & a \cdot f(k) \\ (f \bullet a)(x) & = & f(a \cdot x) \end{array}$$

1.33 Endomorphisms The semigroup endomorphisms $\mathrm{End}(A)$ on a semigroup A have the structure of a monoid, as a special case of §1.13. In addition, $\mathrm{End}(A)$ has an *additional semigroup structure* inherited from A by pointwise multiplication $x \mapsto f_1(x) \cdot f_2(x)$. The two structures are coherent in a unique way giving rise the structure of "near semiring."

To treat this structure in an algebraic way, we write the elements of $\mathrm{End}(A)$ as $\alpha, \beta, \ldots$, composition as multiplication $\alpha \cdot \beta$ and pointwise multiplication inherited from A as $\alpha + \beta$. (Warning: there is no assumption of commutativity for $+$.) It is then convenient to think of the application of an endomorphism α to an element $x \in A$ as "exponentiation," which we write as ${}^\alpha x \stackrel{\mathrm{def}}{=} \alpha(x)$. Dually, the application of $\alpha \in \mathrm{End}(A)^{\mathrm{op}}$ is written as $x^\alpha \stackrel{\mathrm{def}}{=} \alpha(x)$.

The monoid structure of $\mathrm{End}(A)$ and $\mathrm{End}(A)^{\mathrm{op}}$ give the equations:

$$\begin{array}{rcl@{\qquad\qquad}rcl} {}^{\alpha\beta}x & = & {}^\alpha({}^\beta x) & x^{\alpha\beta} & = & (x^\alpha)^\beta \\ {}^1x & = & x & x^1 & = & x \end{array}$$

and the fact that α is a semigroup morphism says:

$$\begin{array}{rcl@{\qquad\qquad}rcl} {}^\alpha(xy) & = & ({}^\alpha x)({}^\alpha y) & (xy)^\alpha & = & (x^\alpha)(y^\alpha) \\ {}^\alpha 1_A & = & 1_A & (1_A)^\alpha & = & 1_A \end{array}$$

The pointwise operation $+ : \mathrm{End}(A) \times \mathrm{End}(A) \to \mathrm{End}(A)$ becomes:

$$\begin{array}{rcl@{\qquad\qquad}rcl} {}^{(\alpha+\beta)}x & \stackrel{\mathrm{def}}{=} & ({}^\alpha x)({}^\beta x) & x^{\alpha+\beta} & \stackrel{\mathrm{def}}{=} & x^\alpha x^\beta \\ {}^0x & \stackrel{\mathrm{def}}{=} & 1_A & x^0 & \stackrel{\mathrm{def}}{=} & 1_A \end{array}$$

Evidently, the semigroup $(\mathrm{End}(A), +)$ has a unit iff A has a unit, and it is commutative iff A is commutative. We can also see that multiplication in $\mathrm{End}(A)$ (or $\mathrm{End}(A)^{\mathrm{op}}$) distributes over addition on the *right* (*left*):

$$\begin{array}{rcl@{\qquad\qquad}rcl} (\alpha_1 + \alpha_2)\beta & = & \alpha_1\beta + \alpha_1\beta & \alpha(\beta_1 + \beta_2) & = & \alpha\beta_1 + \alpha\beta_2 \\ 0\beta & = & 0 & \alpha 0 & = & 0 \end{array}$$

However, only one distributivity law does holds, not the other. This form of a structure is referred to as a "near semiring."

If $\delta : B \to \mathrm{End}(A)$ is a representation of a monoid B, then the elements of B can be treated as exponents for the elements of A in the same way:

$${}^b x = {}^{\delta(b)}x$$

If $\delta : B \to \mathrm{End}(A)^{\mathrm{op}}$ is an anti-representation of B, then we use the notation:

$$x^b = x^{\delta(b)}$$

1.34 Automorphisms 1 If A is a group, the elements of $\mathrm{Aut}(A)$ and $\mathrm{Aut}(A)^{\mathrm{op}}$ can be thought of as exponents for the elements of A in the same way as in §1.33. In addition to the laws mentioned in §1.33, we obtain additional laws for inverses:

$$ {}^{\alpha}(x^{-1}) \;=\; ({}^{\alpha}x)^{-1} \qquad (x^{-1})^{\alpha} \;=\; (x^{\alpha})^{-1} $$

The (non-commutative) additive structure becomes a group, with inverses:

$$ {}^{(-\alpha)}x \;\stackrel{\mathrm{def}}{=}\; ({}^{\alpha}x)^{-1} \qquad x^{-\alpha} \;\stackrel{\mathrm{def}}{=}\; (x^{\alpha})^{-1} $$

We have a single distributive law as in §1.33, but now it works for additive inverses too:

$$ (-\alpha)\beta \;=\; -\alpha\beta \qquad \alpha(-\beta) \;=\; -\alpha\beta $$

This form of a structure is called a "near ring."

If $\delta : B \to \mathrm{End}(A)$ is a group representation, then B becomes a near ring as well.

1.35 Automorphisms 2 (Conjugation) The collection of group automorphisms $\mathrm{Aut}(A)$ is a group, as a special case of §1.24.

For any element $a \in A$, there is a canonical automorphism $C_a \in \mathrm{Aut}(A)$ called *conjugation*: $C_a(x) = axa^{-1}$. Using same intuitions as in §1.33, $C_a(x)$ is also written as x^a or ${}^a x$. To verify that conjugation is a group homomorphism, note that:

$$ \begin{array}{l} C_a(x_1x_2) \;=\; a(x_1x_2)a^{-1} \;=\; ax_1a^{-1}ax_2a^{-1} \;=\; C_a(x_1)\cdot C_a(x_2) \\ C_a(1_A) \;=\; a1_Aa^{-1} \;=\; 1_A \\ C_a(x^{-1}) \;=\; ax^{-1}a^{-1} \;=\; (a^{-1})^{-1}x^{-1}a^{-1} \;=\; (axa^{-1})^{-1} \;=\; C_a(x)^{-1} \end{array} $$

Second, conjugation C_a is an isomorphism, with the inverse $C_{a^{-1}}$:

$$ C_a(C_{a^{-1}}(x)) \;=\; a(a^{-1}xa)a^{-1} \;=\; (aa^{-1})x(aa^{-1}) \;=\; x $$

and similarly for $C_{a^{-1}}(C_a(x)) = x$.

Finally, the map $a \mapsto C_a$ of type $A \to \mathrm{Aut}(A)$ is a group homomorphism:

$$ \begin{array}{l} C_{(a_1a_2)}(x) \;=\; (a_1a_2)x(a_1a_2)^{-1} \;=\; a_1(a_2xa_2{}^{-1})a_1{}^{-1} \;=\; C_{a_1}(C_{a_2}(x)) \\ C_{1_A}(x) \;=\; 1_Ax(1_A)^{-1} \;=\; x \end{array} $$

$C_{a^{-1}}$ is the inverse of C_a as seen above.

All the automorphisms in $\mathrm{Aut}(A)$ of the form C_a are called *inner automorphisms*. The others are called *outer automorphisms*.

1.36 Center The *center* of a semigroup A, denoted $Z(A)$, is the set of all $x \in A$ that commute with every element of A, i.e., $\forall y \in A.\, xy = yx$. The elements that commute in this way are called the "central elements." It is clear that the center is a sub-semigroup of A. The same ideas apply to monoids and groups. In these cases, the center always contains at least 1_A and, hence, is nonempty.

2 Subobjects and Quotients

2.1 Lattice of subobjects The sub-semigroups of a semigroup A are partially ordered by the inclusion order. They form a complete lattice.

The inf's $\bigwedge$ are given by intersection. If $\{S_i\}_{i\in I}$ is a family of sub-semigroups, i.e., each S_i is closed under multiplication, then their intersection is evidently closed under multiplication as well. Similarly, submonoids of a monoid A and subgroups of a group A are also closed under intersection.

The sup $\bigvee$ of a family $\{S_i\}_{i\in I}$ of sub-semigroups is the sub-semigroup generated by their union $\bigcup_{i\in I} S_i$. The sups of submonoids and subgroups are similar. Consider a binary join $S \vee T$ as an example. It consists of all those elements of A that can be written as products $s_1 t_1 \cdots s_k t_k$ of $2k$ factors $s_i \in S$ and $t_i \in T$.

The least element of the complete lattice is the empty semigroup and the top element is the semigroup A itself. For monoids (groups), the least element of the complete lattice is the trivial monoid (group) $\{1\}$.

2.2 Submonoids and spans The submonoids S of a monoid A induce a certain structure on $|A|$. Note that every submonoid acts on the left as well as right by the restriction of multiplication $\cdot : S \times |A| \to |A|$ and $\cdot : |A| \times S \to |A|$. The resepctive spans of an element $x \in |A|$ (cf. §7.19) under these multiplications are denoted:

$$Sx = \{\, kx \mid k \in S \,\}$$
$$xS = \{\, xk \mid k \in S \,\}$$

We call Sx the *left span* of x under multiplication by S and xS the *right span* of x. Note that S itself is a span, *viz.*, the span of 1_A.

An alternative view is to think of the set Sx as a "translation" of S by an element $x \in A$, and hence it is the *right translate* of S by x. Similarly, xS is the *left translate* of S by x. Note that a "left span" becomes a "right translate" and *vice versa*. (This tussle between "left" and "right" is unfortunate, but inevitable. In an asymmetric composite Sx, we might either think of S acting on x from the left, or x acting on S from the right.)

2.3 Relative relations When $S \subseteq A$ is a submonoid, each span of S and, in fact, the entire monoid A, is preordered by the action of S:

$$x \preceq^L y \pmod S \iff \exists k \in S.\, kx = y$$
$$x \preceq^R y \pmod S \iff \exists k \in S.\, xk = y$$

(Multiplication in S gives transitivity and the unit gives reflexivity.) We call these relations the *left preorder* and *right preorder* relative to S. The left span of x is nothing but the set of all elements above x in the left preorder. The right span of x is the set of all elements above x in the right preorder. We can also express the preorders in terms of spans as follows:

$$x \preceq^L y \pmod S \iff Sx \supseteq Sy$$
$$x \preceq^R y \pmod S \iff xS \supseteq yS$$

These are called *(relative) Green's preorders* [Rhodes and Steinberg, 2009, App. A] but normally written with the opposite convention of the order from that used above.

If the monoid A is commutative then, for any submonoid S, the left and right preorders are the same, and the left and right spans of all elements are the same.

For example, consider the multiplicative monoid of natural numbers $\mathbb{N}$, and let S be the submonoid of even numbers. Then $x \preceq y$ holds iff y is an even multiple of x. The (left or right) span of x is the set of all even multiples of x.

2.4 Relative equivalence and orbits (cosets) For a submonoid $S \subseteq A$ and its left action on the monoid A, we define the *left equivalence* relation $\sim^L \subseteq |A| \times |A|$ generated by the rule:

$$x \sim^L y \pmod{S} \impliedby x \preceq^L y \vee x \succeq^L y \pmod{S}$$

More explicitly, the relation $\sim^L$ is the transitive closure $(\preceq^L \cup \succeq^L)^*$. The equivalence class of x under this equivalence relation, denoted ${}_S\lceil x \rceil$, is termed the *left orbit* of x under S (or a *right coset* of S).[4] Note that S itself is an orbit ${}_S\lceil 1_A \rceil$. The set of all left orbits of S is denoted $S\backslash A$. It is a quotient of the module ${}_S|A|$, with a trivial left action by S: $k \cdot {}_S\lceil x \rceil = {}_S\lceil kx \rceil = {}_S\lceil x \rceil$, since the orbit of x includes all the scalar multiples of x.

Similarly, we can define the right equivalence $\sim^R$ modulo S and *right orbits* $\lceil x \rceil_S$. The set of all right orbits of S is denoted A/S.

2.5 Relative Green's relations A submonoid $S \subseteq A$ also induces a *left Green's equivalence* relation $\approx^L \subseteq |A| \times |A|$ (traditional notation $\mathcal{L}^S$):

$$x \approx^L y \pmod{S} \iff x \preceq^L y \wedge x \succeq^L y \pmod{S} \iff Sx = Sy$$

Note that $\approx^L$ is smaller than $\sim^L$. We might regard it as an "inner" left equivalence compared to the "outer" left equivalence $\sim^L$. The equivalence class of an element $x \in A$ under this equivalence relation is a tighter orbit, which we denote as ${}_S\lfloor x \rfloor$, whereas the traditional notation is L_x (with S left implicit). It is called the *$\mathcal{L}$-class of x* under S. The set of all $\mathcal{L}$-classes under S is denoted $S \setminus\!\setminus A$.

Similarly, we have a *right Green's equivalence* relation $\approx^R$ (traditional notation: $\mathcal{R}^S$):

$$x \approx^R y \pmod{S} \iff x \preceq^R y \wedge x \succeq^R y \pmod{S} \iff xS = yS$$

The $\mathcal{R}$-class of x is denoted $\lfloor x \rfloor_S$ or R_x. The set of all $\mathcal{R}$-classes under S is denoted $A /\!\!/ S$.

The *relative Green's relation* $\mathcal{H}^S$ combines the previous two:

$$x \approx^H y \pmod{S} \iff x \approx^L y \wedge x \approx^R y \pmod{S}$$

The orbit of x under this equivalence relation is denoted ${}_S\lfloor x \rfloor_S$ or H_x.

2.6 Subgroups and orbits (cosets) If A is a group, then a subgroup $S \subseteq A$ directly induces two *equivalence relations* on $|A|$. (These are equivalence relations rather than preorders because inverses give symmetry.) Thus, the relations $\preceq^L$, $\succeq^L$, $\sim^L$ and $\approx^L$ all get identified. To spell out the equivalence aspect concretely, if there is $k \in S$ such that $kx = y$ then we also have $x = k^{-1}y$. Hence $x \preceq^L y$ if and only if $x \succeq^L y$. Moreover, we can express the witness $k \in S$ in $kx = y$ as $k = yx^{-1}$. So, one can use the following formulations:

$$\begin{array}{lll} x \sim^L y \pmod{S} & \iff & yx^{-1} \in S \iff xy^{-1} \in S \\ x \sim^R y \pmod{S} & \iff & x^{-1}y \in S \iff y^{-1}x \in S \end{array}$$

[4] Our terminology of left/right cosets is opposite to that of Mac Lane and Birkhoff [Mac Lane and Birkhoff, 1967] and the terminology for ideals in rings.

A left span Sx is now also a *left orbit* ${}_S[x]$, i.e., an *equivalence class* under $\sim^L$ modulo S. If $x \sim^L y \pmod S$, then $Sx = Sy$, i.e., ${}_S[x] = {}_S[y]$, $x \in Sy$ and $y \in Sx$. The notation $S\backslash A$ is used to denote the set of left orbits under S. These are also called the *right cosets* of S.

Dually, A/S denotes the set of right orbits (left cosets) $[x]_S = xS$.

The left and right cosets are in one-to-one correspondence ${}_S[x] \mapsto [x^{-1}]_S$.[5] To see that the correspondence is well defined, note that ${}_S[x] = Sx$. For every $ax \in Sx$, we have $(ax)^{-1} = x^{-1}a^{-1} \in x^{-1}S$.

For example, consider the multiplicative group P^* of all positive reals, and let S be the subgroup of positive rationals. Then $x \sim y \bmod S$ iff x and y are rational multiples of each other. The (left or right) orbit of x is the set of all rational multiples of x.

2.7 Order and index The *order* of a finite group A is the number of its elements. The *index* of a subgroup $S \subseteq A$ is the number of its distinct left or right orbits, written as $[A : S]$. (Since the left and right orbits are in one-to-one correspondence, the number of orbits is independent of the polarity.)

Every orbit of S has the same number of elements as S. If ${}_S[a] = Sa$ is a left orbit of S, notice that there is a bijection $S \cong Sa$. An element x is mapped to xa by right multiplication by a. Another right multiplication by a^{-1} recovers x. A symmetric argument applies to right orbits.

Since the orbits partition A and each orbit has the same cardinality as S, we obtain the formula:

$$|A| = [A : S] \cdot |S|$$

From this, it is also clear that:

- the order of a group is an integer *multiple* of the order of any of its subgroups (Lagrange's Theorem), and
- a group of a prime order has no nontrivial subgroups.

The *order* of an element $x \in A$ is the order of the (cyclic) subgroup $\langle x \rangle$ generated by it, i.e., it is the least natural number n such that $x^n = 1_A$ if such a number exists, or ∞ otherwise. If the order of x is $n < \infty$, then $x^i = x^{i+n}$ for all i. A simple corollary of the above result is:

- the order of a group is an integer multiple of the order of any of its elements. (Equivalently, the order of each element of a group A divides the order of A).

2.8 Green's lemma The generalization of the order and index for monoids is given by Green's lemma: *If $a \approx^L b \pmod S$ then the $\mathcal{R}$-classes $\lfloor a \rfloor_S$ and $\lfloor b \rfloor_S$ are bijective.*

Suppose $a \approx^L b$ is witnessed by factors $k, k' \in S$ such that $ka = b$ and $k'b = a$. Then we claim that we have a bijection $\lambda_k : \lfloor a \rfloor_S \cong \lfloor b \rfloor_S : \lambda_{k'}$.

First, any $x = al \in \lfloor a \rfloor_S$ has a corresponding element $kx \in \lfloor ka \rfloor_S = \lfloor b \rfloor_S$ because right equivalence is left compatible (cf. §2.12). Thus we have a mapping $\lambda_k : \lfloor a \rfloor_S \to \lfloor b \rfloor_S$ and, similarly, $\lambda_{k'} : \lfloor b \rfloor_S \to \lfloor a \rfloor_S$. It remains to show that these two are mutual inverses. Since $x = al = \rho_l(a)$, the corresponding $y = \lambda_k(x)$ is also given by $y = bl = \rho_l(b)$. (Note that $y =$

[5] Needs proof.

$\lambda_k(x) = kx = kal = bl$.) So, $\lambda_{k'}(\lambda_k(x)) = k'kx = k'kal = al = x$. Similarly, $\lambda_k(\lambda_{k'}(y)) = y$. Thus, we have established the claimed bijection. ■

Dually, *if* $a \approx^R b \pmod S$ *then the* $\mathcal{L}$*-classes* ${}_S\lfloor a \rfloor$ *and* ${}_S\lfloor b \rfloor$ *are bijective.*

2.9 Green index The *Green index* of a submonoid $S \subseteq A$ is defined as the number of $\mathcal{H}$-classes under S with an additional class for S itself, i.e.,

$$[A : S]_G = 1 + |(A - S)/\mathcal{H}_S|$$

This definition is due to [Gray and Ruškuc, 2008].

2.10 Internal products If A is a semigroup and S and T are subsets of A, then their *product* ST is the set of all elements ab where $a \in S$ and $b \in T$. Note that left and right translates are special cases of internal products, where we take the liberty to write a singleton set $\{a\}$ as just a.

2.11 Compatible relations A logical relation $R : A \leftrightarrow A$ between a semigroup A and itself is called a *compatible* relation. The multiplication operation is related to itself by $R \times R \to R$.

For relations $R : A \leftrightarrow A$ between a semigroup and itself, two other notions are used. A relation R is said to be *left compatible* if the binary operation satisfies $I_A \times R \to R$ and *right compatible* if it satisfies $R \times I_A \to R$[6] Note that left compatible means $a \left[R\right] b \Longrightarrow \forall x.\, xa \left[R\right] xb$.

To see the algebraic structure of these notions, regard A as a left A-module ${}_A|A|$. Then bisimulation relations $R : {}_A|A| \leftrightarrow {}_A|A|$ are precisely the *left compatible* relations. The multiplication operation is then a scalar multiplication and is related to itself by $I_A \times R \to R$. Similarly, viewing A as a right A-module, $|A|_A$, we obtain *right compatible* relations R that make the right (scalar) multiplication operation relate to itself by $R \times I_A \to R$.

2.12 Compatibility of relative relations If $S \subseteq A$ is a submonoid of a monoid A, we can view A as an (S, A)-bimodule ${}_SA_A$. The left preorder $a \preceq^L b \pmod S$ means $ka = b$ for some $k \in S$, which implies that, for any $r \in A$, we also have $kar = br$, i.e., $ar \preceq^L br \pmod S$. In other words, the left preorder modulo S is *right compatible.*

The *inverse* of the left preorder, $a \succeq^L b \pmod S$, is right compatible in the same way. We have $a = kb$ for some $k \in S$. Then $ar = kbr$ implying $ar \succeq^L br$. By iterating these relations, we can determine that the *left equivalence relation* $\sim^L$ is also right compatible.

Similarly, the right preorder modulo S and right equivalence modulo S are *left compatible.*

In groups, the left equivalence $\sim^L \pmod S$ is similarly right compatible and the right equivalence $\sim^R \pmod S$ is left compatible.

2.13 Congruences and Quotients A *congruence relation* in a semigroup A is a logical relation $\sim : A \leftrightarrow A$ that is also an equivalence relation. In short, it is a "logical equivalence relation," which also means a "compatible equivalence relation." The *quotient* of A with respect to $\sim$ consists of equivalence classes $[a]_\sim$ under $\sim$, along with the induced binary operation $[a]_\sim \cdot [b]_\sim = [ab]_\sim$ which is well-defined and associative. This gives a *quotient semigroup* $A/\sim$. Similarly, congruences on monoids and groups give rise to quotient monoids and quotient groups.

[6] Our terminology of left/right compatibility is consistent with [Clifford and Preston, 1961].

An equivalence relation $\sim : A \leftrightarrow A$ that is only left compatible with the binary operation is called a *left congruence* relation. Since left compatibility means a bisimulation relation of the left A-module ${}_A|A|$, a left congruence is a "bisimulation equivalence" of ${}_A|A|$.

A *right congruence* is similarly a right compatible equivalence relation.

If a relation is both a left congruence and a right congruence then it is a congruence. Suppose $a \sim b$ and $a' \sim b'$. Then $aa' \sim ba'$ by the fact that $\sim$ is a right congruence and $ba' \sim bb'$ by the fact that $\sim$ is a left congruence. Hence $aa' \sim bb'$.

Since, in groups, the left equivalence $\sim^L \pmod S$ is right compatible, it is a right congruence relation. Some authors call it *right congruence modulo* S and use the notation $\equiv_r \pmod S$. Note that this "congruence" is of the right A-module $|A|_A$ with an additional action of S on the left. So, $S\backslash A$, which is more accurately written $S\backslash|A|$, is a quotient of the module $|A|_A$ and continues to be a right A-module with the action ${}_S[x] \cdot a = {}_S[xa]$.

2.14 Congruences and normal subgroups 1 If $N \subseteq A$ is a subgroup of a group A then $N\backslash A$ and A/N are quotient modules of $|A|_A$ and ${}_A|A|$ respectively. We may ask the question, under what conditions do the quotients form groups, not merely modules?

For the quotient A/N to be a group, it must first be closed under multiplication, which is given by $[x]_N \cdot [y]_N = [xy]_N$. Since the orbits of subgroups are the same as spans, this amounts to requiring $xN \cdot yN = xyN$. A sufficient condition is to require $Ny = yN$. In that case, $xN \cdot yN = xNyN = xyNN = xyN$. A subgroup N satisfying

$$Ny = yN \quad (\forall y \in A) \tag{2.1}$$

is called a *normal subgroup*. We write $N \lhd A$ to denote the fact that N is a normal subgroup of A.

If N is a normal subgroup, $x \in Ny$ iff $x \in yN$, i.e., $x \in {}_N[y]$ iff $x \in [y]_N$. Thus the left and right equivalence relations modulo N are the same. So, we write the relation symmetrically as $\sim_N$. By virtue of being a left equivalence, $\sim_N$ is a right congruence and, by virtue of being a right equivalence, it is also a left congruence. Hence it is a *monoid congruence*. We claim that it is also a *group congruence*. If $x \in {}_N[y] = Ny$ then $x = ky$ for some $k \in N$ and $x^{-1} = y^{-1}k^{-1}$ and, so, $x^{-1} \in y^{-1}N = Ny^{-1} = {}_N[y^{-1}]$.

The quotient $A/N = N\backslash A$ is thus a group.

The condition (2.1) is often stated in a slightly different form. By multiplying both sides of the equation by y^{-1}, we obtain $N = yNy^{-1}$ for all $y \in A$. We say that N is *closed under conjugation* by elements of A.

Note that A itself is a normal subgroup of A and so is the singleton $\{1_A\}$. These two are considered *trivial* normal subgroups and others nontrivial. A group with no nontrivial normal subgroups is said to be *simple*. Simple groups are akin to prime numbers in the sense that they cannot be decomposed.

If A is a commutative group then *every subgroup* of A is normal. That is because there are no nontrivial conjugates: $aka^{-1} = aa^{-1}k = k$. Every subgroup is automatically closed under conjugation.

2.15 Congruences and normal submonoids Define *normal submonoids* $N \subseteq A$ using the same condition (2.1), i.e., $Ny = yN$ for all $y \in A$. For a normal submonoid N, we have

$$x \preceq^L y \pmod N \iff x \preceq^R y \pmod N$$

and, hence,

$$\begin{array}{lll} x \sim^L y \pmod N & \iff & x \sim^R y \pmod N \\ x \approx^L y \pmod N & \iff & x \approx^R y \pmod N \end{array}$$

Again, we write these relations symmetrically as $\sim_N$ and $\approx_N$. These are two-sided congruences, making $A/N = N\backslash A$ as well as $A /\!\!/ N = N \setminus\!\!\setminus A$ into monoids.

However, unlike in the case of groups, we have inclusions $\lfloor y \rfloor_N \subseteq yN \subseteq \lceil y \rceil_N$, which are not equalities in general. Hence, we do not get a complete representation of congruences by normal submonoids.[7]. An attempt to characterize the congruences representable by normal submonoids is in Appendix A.

2.16 Congruences and normal subgroups 2 As noted in §2.14 the equivalence relation $\sim_N$ of a normal subgroup N is a group congruence. Now we wish to argue that *every* group congruence is the equivalence relation of a normal subgroup.

If A is a group and $\equiv$ is a group congruence on A, the congruence class $N = [1]_\equiv$ forms a subgroup of A, because

- $k \equiv 1$ and $l \equiv 1$ implies $kl \equiv 1$, and
- $k \equiv 1$ implies $1 \equiv k^{-1}$.

The other congruence classes of $\equiv$ are orbits under N (or cosets of N). Using the formulas of §2.6, we note:

- $a \equiv b \iff 1 \equiv ba^{-1} \iff a \sim^L b \pmod N$.
- $a \equiv b \iff 1 \equiv a^{-1}b \iff a \sim^R b \pmod N$.

Thus, the congruence class of a is ${}_N[a] = [a]_N$. Note that it is *both a left orbit and a right orbit.* Since $Na = aN$ for all $a \in A$, N is a normal subgroup.

Equivalently, we can also note that the subgroup N is closed under *conjugation*, i.e., if $k \in N$ and a is any element of A, then $aka^{-1} \in N$. (Since $k \equiv 1$, $aka^{-1} \equiv aa^{-1} = 1$.) This is another way of saying that N is a normal subgroup.

2.17 Congruences and normal subgroups 3 *Normal subgroups are the same as congruence relations.*

- The left and right equivalence relations induced by a normal subgroup N coincide. If $a \sim^R_N b$, i.e., $a^{-1}b \in N$, then $ba^{-1} = ba^{-1}(bb^{-1}) = b(a^{-1}b)b^{-1} \in N$ by conjugation. So $a \sim^L_N b$. We write the relation symmetrically as $\sim_N$.
- The relation $\sim_N$ is a congruence relation. If $a_1 \sim_N b_1$ and $a_2 \sim_N b_2$ then ${a_1}^{-1}b_1 \in N$ and, by conjugation, ${a_2}^{-1}({a_1}^{-1}b_1)a_2 \in N$ and also ${a_2}^{-1}b_2 \in N$. Hence, ${a_2}^{-1}{a_1}^{-1}b_1a_2{a_2}^{-1}b_2 \in N$. But this element is nothing but $(a_1a_2)^{-1}(b_1b_2)$, showing that $a_1a_2 \sim_N b_1b_2$.

The equivalence class of 1 under $\sim_N$, i.e., the orbit of 1 under left or right multiplication by N, is precisely N. So, we have established a bijection between group congruences and normal subgroups.

The *quotient group* (also called a "factor group") A/N consists of all the orbits under N (which are precisely the equivalence classes under $\sim_N$) with multiplication defined as $[a]_N[b]_N = [ab]_N$. There is an evident epimorphism $p : A \to A/N$ given by $p(a) = aN$.

[7] Can we not define normal submonoids by the condition ${}_N\lceil y \rceil = \lceil y \rceil_N$?

2.18 The idea of normal subgroups Since normal subgroups seems a bit technical, we explain the intuition. A normal subgroup is a subgroup whose left and right equivalence relations are the same (and, hence, the left and right orbits are the same). Such a relation is then both a right congruence and a left congruence, i.e., *congruence*.

By definition:

$$\begin{array}{lll} a \sim^L b \pmod N & \iff & \exists l \in N.\, la = b \\ a \sim^R b \pmod N & \iff & \exists r \in N.\, ar = b \end{array}$$

identifying the two relations means that, for every la, there is some $r \in N$ so that $la = ar$, and *vice versa*. This is also said more easily as $aN = Na$.

Thinking of the group elements as *transformations*, what we mean is that the transformations of N are independent from all other transformations of A in the sense that they can be applied before or after those transformations to achieve the same effect.

This fact allows us to turn the quotient A/N into a group. We define multiplication by $[a]_N \cdot [b]_N = aN \cdot bN = a(Nb)N = a(bN)N = abN = [ab]_N$. This is possible if and only if $Nb = bN$ for all $b \in A$.

2.19 Transversals If $S \subseteq A$ is a subgroup then the quotient $S\backslash A$ consists of left orbits ${}_S[a]$ of elements $a \in A$. By picking one representative of each orbit, we obtain a *right transversal* of S. It is also called *complete system of right coset representatives* of S. (The standard terminology of "right" here is based on cosets rather than orbits.) A right transversal is bijective with the quotient $S\backslash A$.

Similarly, by picking respresentives of each right orbit $[a]_S$ in A/S, we obtain a *left transversal* of S, which will be bijective with A/S.

If S is a *normal* subgroup, then its left orbits and right orbits are the same. Therefore, every left transversal is also a right transversal. We simply refer to it as a *transversal* of S.

Transversals are not required to be closed under the multiplication of A, and hence do not make subgroups of A.

However, transversals form a weaker structure called "loops."

2.20 Decomposition of groups Let $N \subseteq A$ be a normal subgroup. Multiplication in A can be viewed as a two-step process consisting of multiplication in the quotient A/N followed by multiplication in N [Straubing, 1989].

Let $T = \{\nu_i\}_{i \in I}$ be a transversal of N. If $a \in A$, denote its chosen representative $\nu_i \in Na$ by a^T. Then a can be represented as $a = k\nu_i$ for some some $k \in N$ where $\nu_i = a^T$. This factorization of a is unique. Then the multiplication of $a = k_1\nu_i$ and $b = k_2\nu_j$ can be written as $(k_1\nu_i)(k_2\nu_j) = k\nu_l$ for some $k \in N$ where $\nu_l = (\nu_i\nu_j)^T$. The factor k can in fact be determined as $k = k_1(\nu_i k_2 {\nu_i}^{-1})$. Verify that $k_1(\nu_i k_2 {\nu_i}^{-1})\nu_i\nu_j = k_1\nu_i k_2\nu_j$.

The map $k \mapsto \nu_i k {\nu_i}^{-1}$ is an automorphism of N determined by $\nu_i \in T$. Thus, we have given a map $T \to \mathrm{Aut}(N)$. In fact, we have given a homomorphism $\phi : (A/N) \to \mathrm{Aut}(N)$ given by $\phi_a(k) = (a^T)k(a^T)^{-1}$.

This idea leads to the notion of wreath products (§3.11) ...

2.21 Kernel of a morphism If $h : A \to B$ is a morphism of semigroups, its *kernel congruence*, denoted $\sim_h$, is a subset of $A \times A$:

$$\sim_h = \{ (a, b) \mid h(a) = h(b) \}$$

To see that it is a congruence relation on A, note that, if $a_1 \sim_h b_1$ and $a_2 \sim_h b_2$ then $h(a_1) = h(b_1)$ and $h(a_2) = h(b_2)$ and, so, $h(a_1a_2) = h(b_2b_2)$, i.e., $a_1a_2 \sim_h b_1b_2$. The congruence classes under $\sim_h$ are denoted $[a]_h$. All the elements in a congruence class share the same image under h. The quotient $A/\sim_h$ is then isomorphic to the image of h: $A/(\sim_h) \cong \mathrm{Im}(h)$.

The kernel congruence of a group homomorphism can be equivalently expressed as a normal subgroup:

$$\ker(h) = \{ a \mid h(a) = 1_B \}$$

As mentioned above, normal subgroups are the same as congruence relations. So, the congruence relation $\rho(a, b).\, h(a) = h(b)$ determines a normal subgroup $[1_A]_h = \{ a \mid h(a) = h(1_A) \} = h^{-1}(1_B)$, which is called its *kernel*. The equivalence classes $[a]_h$ are nothing but the (left or right) orbits of $\ker(h)$, i.e., $[a]_h = {}_{\ker(h)}[a] = [a]_{\ker(h)}$. The quotient $A/\ker(h)$ is thus isomorphic to the image of h: $A/\ker(h) \cong \mathrm{Im}(h)$. This is referred to as the

First isomorphism theorem: *If $h : A \to B$ is a group isomorphism then the kernel of h is a normal subgroup of A and $A/\ker(h) \cong \mathrm{Im}(h)$.*

This simplification of the kernel congruence to the kernel is possible because groups have inverses. If $a \sim_h b$, i.e., $h(a) = h(b)$, then $h(a)(h(b))^{-1} = 1_B$. But this is nothing but $h(ab^{-1})$. Hence $ab^{-1} \in \ker(h)$. Secondly, the kernel is a *normal* subgroup, also facilitated by the presence of inverses. If $y = xk$ for some $k \in \ker(h)$ then $y = xkx^{-1}x$. Since $k' = xkx^{-1} \in \ker(h)$, every xk can be written as $k'x$ for some $k' \in \ker(h)$.

However, this simplification is *not* possible for semigroups and monoids. Semigroups are not required to have units. For monoids, the inverse image of 1_B is a submonoid of A, but it has no reason to be normal. Hence, it may not represent any congruence, let alone representing the induced congruence $\sim_h$.

2.22 Lemma *A morphism of groups $h : A \to B$ is a monomorphism if and only if* $\ker(h) = \{1\}$.

Proof: If h is a monomorphism, it is injective. So, there is exactly one element $x \in A$ such that $h(x) = 1$ and the condition of homomorphisms forces x to be 1.

Conversely, if $N = \ker(h) = \{1\}$, then the equivalence relation induced by N is $x \sim_N y \iff x = y$. Hence, $h(x) = h(y) \implies x = y$. ■

2.23 Image of a morphism If $h : A \to B$ is a morphism of semigroups, its *image*, denoted $\mathrm{Im}(h)$ or $h_!(A)$, is just the direct image of A under h. It is a *sub-semigroup* of B for, if $b_1, b_2 \in \mathrm{Im}(h)$, there exist $a_1, a_2 \in A$ such that $h(a_1) = b_1$ and $h(a_2) = b_2$. Hence, $b_1b_2 = h(a_1a_2) \in \mathrm{Im}(h)$.

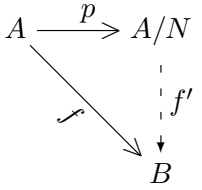
If $S \subseteq A$ is a subsemigroup, then its image under $h : A \to B$ is denoted $h_!(S)$.

If $h : A \to B$ is a morphism of monoids, then $\mathrm{Im}(h)$ is a submonoid of B because $h(1_A) = 1_B$. The same conclusion applies to submonoids $S \subseteq A$.

If h is a morphism of groups, then $\mathrm{Im}(h)$ is a subgroup of B because it is closed under inverses: $h(a^{-1}) = (h(a))^{-1}$. Again, the same conclusion applies to subgroups.

2.24 Universal property of quotient groups If $N \lhd A$ is a normal subgroup then the quotient group A/N along with the projection $p : A \to A/N$ satisfies the following universal property:

Every morphism $f : A \to B$ such that $f_*(N) = \{1_B\}$ uniquely factors through p.



If $f_*(N) = 1$ then $f(kx) = f(x)$ for any scalar multiple kx of $k \in N$. Hence f carries each coset $[x]_N$ to the same element in B. So, we can define the unique factor f' by $f'([x]_N) = f(x)$. This factor is a morphism because $f'([x]_N\,[y]_N) = f'([xy]_N) = f(xy) = f(x)\,f(y) = f'([x]_N)\,f'([y]_N)$.

Fact. $\mathrm{Im}(f') = \mathrm{Im}(f)$. It follows from the fact that p is surjective and $f = p; f'$.

Fact. $\ker(f') = \ker(f)/N$. Note that $K = \ker(f)$ is a normal subgroup of A. Since $f_*(N) = \{1_B\}$, $N \subseteq K$. Moreover, N is closed under conjugation by elements of K. So, $N \lhd K$ and K/N is a group. Secondly, every coset $[k]_N \in K/N$ is also a coset in A/N. So, $K/N \subseteq A/N$. We have $f'([k]_N) = 1_B \iff f(k) = 1_B \iff k \in K \iff [k]_N \in K/N$. Therefore, $\ker(f') = K/N$.

2.25 Corollary *If $f : A \to B$ is a morphism with kernel N, then there is a unique monomorphism $f' : A/N \rightarrowtail B$ such that $f' = p; f$.*

By assumption, $\ker(f) = N$. So, $\ker(f') = N/N = \{1_A\}$. This shows that f' is a monomorphism.

2.26 Corollary *If $f : A \to B$ is an epimorphism with kernel N, then there is a unique isomorphism $f' : A/N \cong B$ such that $f' = p; f$.*

Since f is an epimorphism, so is f'. By the previous corollary, f' is also a monomorphism and, hence, an isomorphism.

2.27 Corollary *If $f : A \to B$ is a morphism with kernel N and image S, then it can be written uniquely as the composite*

$$A \xrightarrow{p} A/N \xrightarrow{\theta} S \xrightarrow{i} B$$

where p is an epimorphism, θ an isomorphism and the insertion i a monomorphism.

In fact, $\theta; i$ is itself a monomorphism. So, the corollary says that every morphism can be factored as an epimorphism followed by a monomorphism. This is the *epi-mono factorization* property.

2.28 Exact sequences An *exact sequence* of groups is a sequence of morphisms

$$A \xrightarrow{g} B \xrightarrow{h} C$$

such that $\mathrm{Im}(g) = \ker(h)$. Note that the $\mathrm{Im}(g;h) = \{1_C\}$. Longer sequences of morphisms $A_0 \xrightarrow{g_0} A_1 \xrightarrow{g_1} \cdots \xrightarrow{g_{n-1}} A_n$ can also be exact in this way.

The exact sequence

$$\mathbf{0} \longrightarrow A \xrightarrow{g} B$$

says that g is mono because the image of a morphism $\mathbf{0} \to A$ just has the unit element 1_A. Dually the exact sequence

$$B \xrightarrow{h} C \longrightarrow \mathbf{0}$$

says that h is epi because the image of h is required to be the entire C. The exact sequence

$$\mathbf{0} \longrightarrow A \xrightarrow{g} B \xrightarrow{h} C \longrightarrow \mathbf{0} \tag{2.2}$$

says that g is mono and h is epi. Such a sequence is called a *short* exact sequence. Then A is a subobject of B, isomorphic to the $\ker(h) \subseteq B$, and C is a quotient of B. *Merged text: In fact, exactness at B means that* $\mathrm{Im}(g) = \ker(h)$. *So,* $\mathrm{Im}(g) \lhd B$ *and* $C \cong B/\mathrm{Im}(g)$. In this case, we say that B is an *extension* of C by A. The extension is said to be *split* if h has a section (pre-inverse or right inverse) r such that $h \circ r = \mathrm{id}_C$.

For example, when $B = A \times C$ is the direct product (cf. §3.1), we have a short exact sequence. Thus $A \times C$ is an extension of C by A. It is also a *split* extension.

The *extension problem* for groups is to characterize all extensions of C by A, i.e., to determine all groups B such that the above sequence is exact.

2.29 Exact sequences of monoids More generally, in any regular category, a pair of morphisms $g_1, g_2 : A \to B$ is called a *kernel pair* for $h : B \to C$ if there is a pullback diagram

$$\begin{array}{ccc} A & \xrightarrow{g_1} & B \\ {\scriptstyle g_2}\downarrow & & \downarrow{\scriptstyle h} \\ B & \xrightarrow{h} & C \end{array}$$

The sequence:

$$A \underset{g_2}{\overset{g_1}{\rightrightarrows}} B \xrightarrow{h} C$$

is said to be a *short exact sequence* if (g_1, g_2) is a kernel pair for h and h itself is a coequalizer of (g_1, g_2) [**?**].

In monoids, the pullback is

$$A = \{\, (y_1, y_2) \mid h(y_1) = h(y_2) \,\}$$

with pointwise multiplication and units, g_1 and g_2 are projections. Since h forms a cocone for (g_1, g_2), all that is needed for it to universal is to be surjective. Hence, a short exact sequence always consists of surjective homomorphism and its kernel pair.

Normal submonoids

Several proposals for a notion of normal submonoids exist in the literature. The first paragraph is our attempt, which is yet unfinished.

2.30 Congruences and normal submonoids 2 As noted in §2.14, the equivalence relation $\sim_N$ of a normal submonoid N is a monoid congruence. Now we wish to argue that *every* monoid congruence is the equivalence relation of a normal submonoid.[8]

If A is a monoid and $\equiv$ is a monoid congruence on A, the congruence class $N = [1]_\equiv$ forms a submonoid of A, because

- $k \equiv 1$ and $l \equiv 1$ implies $kl \equiv 1$.

The following argument needs to be generalized to monoids:

The other congruence classes of $\equiv$ are orbits under N (or cosets of N). Using the formulas of §2.6, we note:

- $a \equiv b \iff 1 \equiv ba^{-1} \iff a \sim^L b \pmod N$.
- $a \equiv b \iff 1 \equiv a^{-1}b \iff a \sim^R b \pmod N$.

Thus, the congruence class of a is ${}_N\lceil a\rceil = \lceil a\rceil_N$. Note that it is *both a left and right orbit.* Since $Na = aN$ for all $a \in A$, N is a normal subgroup.

2.31 Grassmann Let $S \subseteq A$ be a submonoid. Consider the left outer equivalence $\sim^L \pmod S$. It is a right congruence: $x \sim^L y \pmod S \implies xa \sim^L ya \pmod S$ for all $a \in A$. Hence, ${}_S\lceil x\rceil \cdot a = {}_S\lceil xa\rceil$ and ${}_S\lceil x\rceil \cdot ab = ({}_S\lceil x\rceil \cdot a) \cdot b$.

A submonoid $N \subseteq A$ is called a *normal submonoid* in the sense of [Grassmann, 1979] if

$$ {}_N\lceil a\rceil = {}_N\lceil b\rceil \implies {}_N\lceil x\rceil \cdot a = {}_N\lceil x\rceil \cdot b \qquad (\forall x \in A) $$

In that case $N\backslash A$ is a monoid, with multiplication defined by ${}_N\lceil x\rceil\,{}_N\lceil y\rceil = {}_N\lceil x\rceil \cdot y$. The multiplication is associative:

$$ \begin{array}{rcl} ({}_N\lceil x\rceil\,{}_N\lceil y\rceil)\,{}_N\lceil z\rceil & = & ({}_N\lceil x\rceil \cdot y) \cdot z \;=\; {}_N\lceil x\rceil \cdot yz \;=\; {}_N\lceil xyz\rceil \\ {}_N\lceil x\rceil\,({}_N\lceil y\rceil\,{}_N\lceil z\rceil) & = & {}_N\lceil x\rceil\,({}_N\lceil y\rceil \cdot z) \;=\; {}_N\lceil x\rceil\,({}_N\lceil yz\rceil) \;=\; {}_N\lceil x\rceil \cdot yz \;=\; {}_N\lceil xyz\rceil \end{array} $$

[Grassmann, 1979] does not state if normality is symmetric, i.e., whether $N\backslash A = A/N$. However, he proves the third isomorphism theorem:

- *If $M \subseteq N \subseteq A$ and both M and N are normal submonoids of A, then M is normal in N and $M\backslash N$ is a submonoid of $N\backslash A$. Moreover, $M\backslash A \cong (M\backslash N)\backslash(N\backslash A)$.*

2.32 Ljapin's normal complex [Ljapin, 1974] introduced the following notion. A nonempty subset of a semigroup $U \subseteq A$ is a *normal complex* if, for all $u, u' \in U$

$$ xuy \in U \implies xu'y \in U $$

where x and y range over A^I.

Normal complexes of A include A itself, every singleton subset $\{a\} \subseteq A$, and every two-sided ideal of A. If $U \subseteq A$ is such that the intersections $UA \cap U$, $AU \cap U$ and $AUA \cap U$ are empty (the so-called two-sided "anti-ideals"), then U is a normal complex.

[8] It is unlikely that every monoid congruence can be represented as a normal submonoid. The representable ones need to be characterized.

Normal complexes are closed under non-empty intersection.

If A is a group and N is normal subgroup, then the orbits of N (cosets under N) are normal complexes. Suppose $ak, ak' \in U = [a]_N$. Then $xak'y = xakyy^{-1}k^{-1}k'y \in xakyU$. Therefore, $xaky \in U$ implies $xak'y \in U$.[9]

2.33 Ulam

2.1 Ideals

2.34 Ideals Recall that the underlying set of a semigroup A may be regarded as a left and right module of A, denoted ${}_A|A|_A$.

A *left ideal* of a semigroup A is a nonempty submodule U of the left-multiplication module ${}_A|A|$. In other words, a left ideal is closed under left multiplication $A \times U \to U$, a condition that can also be written as $AU \subseteq U$.

Similarly, U is a *right ideal* if it is closed under right multiplication $U \times A \to U$ or, equivalently, a submodule of the right module $|A|_A$.

A *two-sided ideal* U is closed under both left and right multiplication, or, equivalently, a submodule of the bimodule ${}_A|A|_A$.[10] So a two-sided ideal is both a left ideal and a right ideal. This condition can be represented by the formula $A^I U A^I \subseteq U$, where A^I is A with additional unit element adjoined (§1.11). Note that A itself is a two-sided ideal of A.

Left ideals (right ideals, two-sided ideals) of A are subsemigroups of A. However, ideals have no reason to contain the unit element of A (if it exists). So, they do not form submonoids of monoids.

Note that A itself is a two-sided ideal of A. So is the subset $\{0\}$ if A has zero. These two are considered "improper" ideals, and all other ideals are called *proper ideals*.

A semigroup A is said to be *simple* if it has no proper two-sided ideals, *left-simple* if it has no proper left ideals, and and *right-simple* if it has no proper right ideals. A semigroup with a zero is called 0-*simple* if $A^2 \neq \{0\}$ and has no proper ideals.

2.35 Ideals in rings If $h : A \to B$ is a morphism of rings, then the *kernel* of h is defined to be the inverse image of 0_B. Since h preserves the additive group structure, $\ker(h)$ is a (normal) subgroup under addition. Moreover, for any $a \in A$ and $k \in \ker(h)$, ak, $ka \in \ker(h)$. (Clearly, $h(ak) = h(a)h(k) = h(a) \cdot 0 = 0$.) So, $\ker(h)$ is closed under multiplication by elements of A, i.e., it is a two-sided ideal under multiplication. This motivates the following definition.

A *two-sided ideal* in a ring A is a subset U such that it is a subgroup of the additive group and closed under left- and right-multiplication by elements of A. A *left ideal* (*right ideal*) is a subgroup of the additive group that is only closed under left (right) multiplication by A.

2.36 Theorem *A semigroup is a group if and only if it is left simple and right simple.*

Proof: We first note that *a semigroup A is right simple iff $aA = A$ for every $a \in A$.* For if $aA \neq A$ then aA is a proper right ideal of A, so A would not be right simple. Conversely, if U is a proper right ideal of A containing a, then $aA \subseteq U \subset A$ and, so, $aA \neq A$.

[9] Is this right?

[10] A two-sided ideal is also called an "ideal." But we will not use this terminology. We reserve "ideal" to mean any species of ideal: left, right or two-sided.

But, to say that $aA = A$ for every a is the same as saying that, for every $a, b \in A$, there exists $x \in A$ such that $ax = b$. Combining this with the dual proposition, we obtain that there is also $x \in A$ such that $xa = b$. This means that A is a group, by Weber-Huntington axioms for groups. ■

2.37 Theorem *A nontrivial commutative ring is a field if and only if it has no proper two-sided ideals.*

The proof is the same as the above, but specialized to commutative monoids internal to **Ab**.

2.38 Internal products of ideals Note that the product of two left ideals U_1U_2 is in turn a left ideal: $AU_1U_2 \subseteq U_1U_2$ by virtue of U_1 being an ideal.

The product of two-sided ideals is again a two-sided ideal: $A^I U_1 U_2 A^I \subseteq A^I U_1 A^I A^I U_2 A^I \subseteq U_1U_2$. (We are able to add two copies of A^I in the middle because they contain the unit elements needed for the inclusion.)

2.39 Generated ideals Each species of ideals is closed under intersection. Therefore, for any subset $X \subseteq A$, there is a least ideal containing X. This is called the ideal *generated* by X. If A is a monoid, the left ideal generated by X is simply AX, the right ideal is XA and the two-sided ideal is AXA. If A is a semigroup, the left ideal generated by X is $A^I X$. The two-sided ideal generated by X is $A^I X A^I$.

The ideal generated by a singleton set $\{a\}$ is called a *principal* ideal.

For example, in the semigroup $\mathbb{N}$ of natural numbers under multiplication, a principal ideal is precisely the set of all multiples of some $n \in \mathbb{N}$. We write it as $n\mathbb{N}$. Other ideals are of the form $U\mathbb{N}$ where $U \subseteq \mathbb{N}$.

2.40 Theorem *If $h : A \to B$ is a monoid morphism then*

- *the inverse image of every ideal of B is an ideal of A.*
- *if h is surjective, the image of every ideal of A under h is an ideal of B.*

Proof: If J is a two-sided ideal of B, i.e., $BJB \subseteq J$, then

$$A\,h^{-1}(J)\,A \;\subseteq\; h^{-1}(B)\,h^{-1}(J)\,h^{-1}(B) \;\subseteq\; h^{-1}(BJB) \;\subseteq\; h^{-1}(J)$$

Let U be a two-sided ideal of A, i.e., $AUA \subseteq U$. Then the image under h is

$$h(AUA) \;=\; h(A)\,h(U)\,h(A) \;=\; B\,h(U)\,B$$

by the assumption of surjectivity of h. Since the direct image is a monotone function of subsets, we have $B\,h(U)\,B \subseteq h(U)$. Hence $h(U)$ is a two-sided ideal of B. ■

In particular, since $\{0\}$ is always an ideal of a monoid with a zero, the inverse image of $\{0_B\}$ is an ideal of A. The direct image of $\{0_A\}$ is an ideal of B.

2.41 Rees congruence If U is a two-sided ideal in a semigroup A, define a relation $\equiv_U$ that identifies all the elements of U and separates all the others. More formally,

$$a \equiv_U b \iff a = b \vee a, b \in U$$

This is a congruence relation. If $a \equiv_U b$ by virtue of a and b being in U, and $x \equiv_U y$, then ax, bx, ay, by are also in U. Hence, $ax \equiv_U by$. The dual argument shows that $xa \equiv_U yb$.

The congruence $\equiv_U$ is called the *Rees congruence modulo* U. The quotient $A/\equiv_U$, which is simply denoted A/U, is called the *Rees factor semigroup modulo* U. In this semigroup, the ideal U collapses into a single element and all other elements outside U retain their identity.

2.42 Green's relations The Green's equivalence relations $\mathcal{L}$, $\mathcal{R}$, and $\mathcal{J}$ equate the elements of a monoid when the principal ideals generated by them are the same

$$\begin{array}{lcl} x \left[\mathcal{L}\right] y & \iff & Ax = Ay \\ x \left[\mathcal{R}\right] y & \iff & xA = yA \\ x \left[\mathcal{J}\right] y & \iff & AxA = AyA \end{array}$$

If $x \left[\mathcal{L}\right] y$, we have $Ax = Ay$, which implies that $x \in Ay$ and $y \in Ax$. In other words, $y \preceq^L x \pmod A$ and $x \preceq^L y \pmod A$. Conversely, suppose $x \preceq^L y \pmod A$ and $y \preceq^L x \pmod A$. Then we obtain $Ay \subseteq Ax$ and $Ax \subseteq Ay$ and, hence, $Ax = Ay$. Therefore, we have the equations:

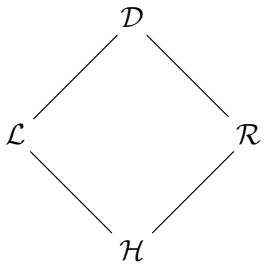
$$\begin{array}{rcl} \mathcal{L} & = & \preceq^L_A \cup \succeq^L_A \\ \mathcal{R} & = & \preceq^R_A \cup \succeq^R_A \end{array}$$

Moreover, $\mathcal{L}$ is a right congruence and $\mathcal{R}$ is a left congruence. The relation $\mathcal{J}$ does not have any corresponding properties.

We can combine $\mathcal{L}$ and $\mathcal{R}$ in two different ways:

$$\begin{array}{lcl} x \left[\mathcal{H}\right] y & \iff & x \left[\mathcal{L}\right] y \wedge x \left[\mathcal{R}\right] y \\ x \left[\mathcal{D}\right] y & \iff & x \left[\mathcal{L}\mathcal{R}\right] y \iff x \left[\mathcal{R}\mathcal{L}\right] y \end{array}$$

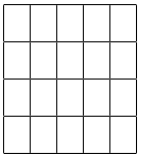
The last definition depends on the fact that $\mathcal{L}$ and $\mathcal{R}$ commute. Hence, $\mathcal{D}$ is the same as the join $\mathcal{L} \vee \mathcal{R}$ in the lattice of equivalence relations on A. Notice also that $\mathcal{H}$ is the same as the meet $\mathcal{L} \wedge \mathcal{R} = \mathcal{L} \cap \mathcal{R}$. Thus, we have the picture:



The equivalence classes under these equivalence relations are called $\mathcal{L}$-class, $\mathcal{R}$-class etc., and we use the notations L_x, R_x, J_x, H_x and D_x for the respective equivalence classes of an element x.

From the discussion above, it is clear that $\mathcal{L}$ and $\mathcal{R}$ each partition a $\mathcal{D}$-class D_x, and $\mathcal{H}$ partitions both $\mathcal{L}$-classes and $\mathcal{R}$-classes. Thus, for finite monoids, each $\mathcal{D}$-class can be presented as a matrix, called an "egg box" diagram, where each cell represents a $\mathcal{H}$-class, each row represents an $\mathcal{R}$-class and each column represents an $\mathcal{L}$-class. By Green's Lemma (§2.8),

$a \left[\mathcal{L}\right] b$ implies $|R_a| = |R_b|$ and, $a \left[\mathcal{R}\right] b$ implies $|L_a| = |L_b|$. Therefore, $a \left[\mathcal{D}\right] b$ implies $|H_a| = |H_b|$. Thus, each cell in the diagram contains the *same number* of elements.



2.43 Green's theorem A more striking consequence of the above facts is given by the Green's theorem, which we introduce in small increments.

Suppose an $\mathcal{H}$-class H_a contains a right multiple ax of a. That means $a \left[\mathcal{H}\right] ax$ and, in particular, $a \left[\mathcal{R}\right] ax$. Then $\rho_x : L_a \cong L_{ax}$ is a bijection. Retricting to R_a (which is the same as R_{ax}), the bijection becomes $\rho_x : H_a \cong H_{ax}$. But, $H_a = H_{ax}$. Therefore, ρ_x is a bijection of H_a onto itself.

Similarly, if H_a contains a left multiple xa of a, then λ_x is a bijection of H_a onto itself.

If an $\mathcal{H}$-class H contains $a, b \in H$ such that $ab \in H$. Then ρ_b and λ_a are bijections of H onto itself. Hence, $xb \in H$ and $ax \in H$ for every $x \in H$. That means that λ_x and ρ_x are also bijections of H onto itself. Hence $Hx = xH = H$ for every $x \in H$. This means that $H^2 = H$.

If H satisfies $Hx = xH = H$ for every $x \in H$, then H must be a group. We must have $ex = x$ for some $e \in H$. ... (to be completed).

3 Products

References: [Mac Lane and Birkhoff, 1967, Rotman, 1965, Grillet, 1995].

3.1 Direct product The *direct product* of semigroups A and B, denoted $A \times B$, consists of the cartesian product of the underlying sets. The multiplication is pointwise:

$$(a_1, b_1) \cdot (a_2, b_2) = (a_1 a_2,\ b_1 b_2)$$

This is a categorical product in the category **SGrp** of semigroups. The projections $\pi_1 : A \times B \to A$ and $\pi_2 : A \times B \to B$ are homomorphisms. If $f : Z \to A$ and $g : Z \to B$ are homomorphisms then $\langle f, g\rangle : Z \to A \times B$ is a homomorphism because $\langle f, g\rangle(z_1 z_2) = (f(z_1 z_2), g(z_1 z_2)) = (f(z_1) \cdot f(z_2),\ g(z_1) \cdot g(z_2)) = (f(z_1), g(z_1)) \cdot (f(z_2), g(z_2)) = \langle f, g\rangle(z_1) \cdot \langle f, g\rangle(z_2)$. The terminal object is any one-element semigroup $\mathbf{1}$. This gives a cartesian monoidal structure $(\mathbf{SGrp}, \times, \mathbf{1})$.

The same construction gives categorical products in **Mon** and **Grp**. $A \times B$ evidently has unit elements $1_{A\times B} = 1_A \times 1_B$. In **Grp**, $A \times B$ has inverses given by $(a, b)^{-1} = (a^{-1}, b^{-1})$. The one-element monoid (or group) is the *null* object, which we write as $\mathbf{0}$.

The isomorphism $A \times \mathbf{0} \cong A$ in **Mon** and **Grp** gives rise to injections $\iota_1 : A \to A \times B$ and $\iota_2 : B \to A \times B$ given by $\iota_1(x) = (x, 1_B)$ and $\iota_1(y) = (1_A, y)$. The injections are sections for the corresponding projections: $\iota_1; \pi_1 = \mathrm{id}_A$ and $\iota_2; \pi_2 = \mathrm{id}_B$.

3.2 Internal direct product The kernel congruence of $\pi_1 : A \times B \to A$ is a congruence relation $\sim_1$ on $A \times B$: $(a_1, b_1) \sim_1 (a_2, b_2) \iff a_1 = a_2$.

In the case of groups, the kernel is equivalently expressed as a normal subgroup $N_1 = \{\, (a, 1_B) \mid a \in A \,\}$, which is isomorphic to A. Similarly, $N_2 = \{\, (1_A, b) \mid b \in B \,\}$ is isomorphic to B. The group $A \times B$ itself can be expressed as the internal product $N_1 N_2 = \{\, p_1 p_2 \mid p_1 \in N_1,\ p_2 \in N_2 \,\}$. The subgroups N_1 and N_2 have the following properties:

1. Their intersection is trivial, i.e., equal to $\mathbf{0} \cong \{1_{A\times B}\}$.
2. $A \times B = N_1 N_2$.
3. Every element of N_1 commutes with every element of N_2, i.e., $p_1 p_2 =_{A\times B} p_2 p_1$ for all $p_1 \in N_1$ and $p_2 \in N_2$.

Together, these three properties completely determine the algebraic structure of the direct product.

3.3 Recognition theorem *If P is any group having subgroups N_1 and N_2 satisfying the above properties, then $P \cong N_1 \times N_2$.*

(Note that N_1 and N_2 are not required to be normal subgroups. It follows that they are normal.)

An equivalent formulation is the following: *If $P = N_1 N_2$ for normal subgroups N_1 and N_2 that have the trivial intersection, then $P \cong N_1 \times N_2$.*

Proof: We use letters a and b for the elements of N_1 and N_2 respectively. Every element $x \in P$ can be uniquely written as a product ab. Define a homomorphism $h : P \to N_1 \times N_2$ by $h(ab) = (a, b)$. It preserves multiplication: $h(a_1 b_1 a_2 b_2) = h(a_1 a_2 b_1 b_2) = (a_1 a_2, b_1 b_2)$. The unit

1 of P is the product $1{\cdot}1$. So, $h(1) = (1, 1)$. The inverse of ab in P is $(ab)^{-1} = b^{-1}a^{-1} = a^{-1}b^{-1}$. So, $h((ab)^{-1}) = (a^{-1}, b^{-1})$. The inverse $h^{-1} : N_1 \times N_2 \to P$ is evidently a homomorphism. So, h is an isomorphism.

To see that N_1 is normal, consider $k \in N_1$ and its conjugation by an arbitrary element $ab \in P$. We have $abk(ab)^{-1} = abkb^{-1}a^{-1} = abb^{-1}ka^{-1} = aka^{-1}$ which is a product within N_1. So, the fact that P is closed under conjugation implies N_1 is closed under conjugation. ■

When P has normal subgroups N_1 and N_2 as above, it is referred to as an *internal direct product*.

3.4 Group extensions Let K and Q be groups. An *extension* of Q by K is a group G such that

1. K is a normal subgroup of G and
2. $Q \cong G/K$.

A somewhat more general treatment of the situation is as follows.

Recall, from §2.28, that a short exact sequence

$$\mathbf{0} \longrightarrow A \xrightarrow{g} B \xrightarrow{h} C \longrightarrow \mathbf{0}$$

says that g is mono and h is epi. This means precisely that B is an *extension* of C by A. Every direct product gives rise to a short exact sequence:

$$\mathbf{0} \longrightarrow A \xrightarrow{p_1} A \times B \xrightarrow{\pi_2} B \longrightarrow \mathbf{0}$$

where p_1 sends each $a \in A$ to $(a, 1_B)$ and π_2 is the projection of the B component.

So, the above definition says that

$$\mathbf{0} \longrightarrow K \longrightarrow G \longrightarrow Q \longrightarrow \mathbf{0}$$

is a short exact sequence.

The *extension problem* (formulated by Hölder) is, given K and Q, to determine all extensions of Q by K. The direct product $K \times Q$ is one form of an extension of Q by K. Semidirect product $K \ltimes Q$, discussed next, is another. The general extension problem is still open.

3.5 Semidirect products of groups (internal) Suppose we have a homomorphism $f : A \to S$ to a subgroup $S \subseteq A$. The kernel $N = \ker(f)$ is a normal subgroup of A. Every orbit (coset) of N is bijective with N and f carries it to a single element $y \in S$. If we suppose further that y belongs to the orbit that is carried to y, i.e., $f(y) = y$ for all $y \in S$, then we can write every element of A as a product ky where $k \in N$ and $y \in S$. Note that $|A| = |N| \times |S|$. This leads to the idea of (internal) semidirect products.

If A is a group with a normal subgroup N and a subgroup S such that their intersection is trivial and $A = NS$, then A is a *semidirect product* of N and S (characterized internally). Symbolically, $A = N \rtimes S$ or $A = S \ltimes N$. Note that we have dropped the commutativity requirement of direct products as in §3.2.

One can also write $A = SN$ in the above, with the same effect. Since N is a normal subgroup, $SN = NS$.

This motivates the terminology of "semidirect product". A semidirect product becomes a direct product when both the factors are normal subgroups.

Note that an internal semidirect product is *not unique* (even up to isomorphism). Different groups A can have the same combination of N and S subgroups and be qualified to be written as $N \rtimes S$.

Example: [Aluffi, 2009, Sec. 2.1, Example 5.13] Consider the symmetry group S_3 on three elements $\{1, 2, 3\}$. We can write the elements of S_3 as

$$e, y, y^2, x, yx, y^2x$$

where

$$y = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \qquad x = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

It has a normal subgroup $C_3 = \{e, y, y^2\} = ((1\,2\,3))$ and a subgroup $C_2 = \{e, x\} = ((1\,2))\,((3))$ with trivial intersection $\{e\}$. Moreover, $S_3 = C_3C_2$. Therefore, $S_3 = C_3 \rtimes C_2$ is an internal semidirect product.

The direct product $C_3 \times C_2$ is C_6. Since the direct product is a special case of semidirect product, this means that we also have $C_6 = C_3 \rtimes C_2$. (Note that C_2 is normal in C_6 even though it is not normal in S_3.)

3.6 Observation Suppose $A = N \rtimes S$. Since N is a normal subgroup of A, it is closed under conjugation by elements of A. In particular, it is closed under conjugation by elements of S. So, there is a monomorphism $\phi : S \to \mathrm{Aut}(N)$ defined by $\phi_b(k) = bkb^{-1}$, similar to that in 1.34. There could be other homomorphisms $\phi : S \to \mathrm{Aut}(N)$. Each of them indicates how N is normal in A. Correspondingly, we have a range of semidirect products $N \rtimes_\phi S$ indexed by homomorphisms $\phi : S \to \mathrm{Aut}(N)$.

If $\phi : S \to \mathrm{Aut}(N)$ is the trivial homomorphism that takes each $x \in S$ to id_N then the semidirect product $N \rtimes_\phi S$ is the same as the direct product $N \times S$.

If $S = \mathrm{Aut}(N)$ and $\phi : S \to \mathrm{Aut}(N)$ is the identity morphism, then the resulting semidirect product is called the *holomorph* of N.

3.7 Semidirect products of groups (external) Let K and B groups and $\phi : B \to \mathrm{Aut}(K)$ be a homomorphism of groups. We can now use the elements of B as exponents for the elements of K as ${}^{\phi_b}k$. For conveience, we abbreviate the notation to ${}^{b}k$. See §1.34 and the identities listed there.

The *semidirect product* $K \rtimes_\phi B$ is defined to have the cartesian product $K \times B$ as its underlying set and multiplication defined by:

$$\langle k_1, b_1\rangle\langle k_2, b_2\rangle = \langle k_1({}^{b_1}k_2),\, b_1b_2\rangle = \langle k_1({}^{\phi_{b_1}}k_2),\, b_1b_2\rangle$$

(Intuition: To "move b_1 past k_2," as required for writing b_1b_2 in the second component, write b_1k_2 as $b_1k_2{b_1}^{-1}b_1$ and treat $b_1k_2{b_1}^{-1}$ as ${}^{b_1}k_2$. See §3.8 for further elaboration of the idea.)

We verify the various properties required.

Associativity:

$$\begin{aligned}
\big(\langle k_1, b_1\rangle\langle k_2, b_2\rangle\big)\,\langle k_3, b_3\rangle &= \langle k_1({}^{b_1}k_2),\, b_1b_2\rangle\langle k_3, b_3\rangle \\
&= \langle k_1({}^{b_1}k_2)({}^{b_1b_2}k_3),\, b_1b_2b_3\rangle \\
\langle k_1, b_1\rangle\,\big(\langle k_2, b_2\rangle\langle k_3, b_3\rangle\big) &= \langle k_1, b_1\rangle\langle k_2({}^{b_2}k_3),\, b_2b_3\rangle \\
&= \langle k_1({}^{b_1}k_2({}^{b_2}k_3)),\, b_1b_2b_3\rangle \\
&= \langle k_1({}^{b_1}k_2)({}^{b_1b_2}k_3),\, b_1b_2b_3\rangle
\end{aligned}$$

These two are equal.

Units:

$$\begin{aligned}\langle 1_K, 1_B\rangle\langle k, b\rangle &= \langle 1_K({}^{1_B}k),\, 1_B b\rangle = \langle k, b\rangle \\ \langle k, b\rangle\langle 1_K, 1_B\rangle &= \langle k({}^{b}1_K),\, b1_B\rangle = \langle k, b\rangle\end{aligned}$$

Inverse: The inverse of $\langle k, b\rangle$ is $\langle {}^{b^{-1}}(k^{-1}),\, b^{-1}\rangle$.

$$\begin{aligned}\langle k, b\rangle^{-1}\langle k, b\rangle &= \langle {}^{b^{-1}}(k^{-1}),\, b^{-1}\rangle\, \langle k, b\rangle \;=\; \langle ({}^{b^{-1}}(k^{-1}))({}^{b^{-1}}k),\, b^{-1}b\rangle \\ &= \langle {}^{b^{-1}}1_K,\, 1_B\rangle \;=\; \langle 1_K,\, 1_B\rangle \\ \langle k, b\rangle\langle k, b\rangle^{-1} &= \langle k, b\rangle\langle {}^{b^{-1}}(k^{-1}),\, b^{-1}\rangle \;=\; \langle k({}^{b}({}^{b^{-1}}(k^{-1}))),\, bb^{-1}\rangle \\ &= \langle k(b^{b^{-1}}(k^{-1})),\, 1_B\rangle \;=\; \langle k({}^{1_B}(k^{-1})),\, 1_B\rangle \;=\; \langle 1_K,\, 1_B\rangle\end{aligned}$$

The semidirect product $K \rtimes_\phi B$ has a *normal subgroup* isomorphic to K consisting of all elements of the form $\langle k, 1_B\rangle$ for $k \in K$. It also has a subgroup isomorphic to B, but it is not necessarily normal. To see that the former is a normal subgroup, note that the conjugation of $\langle k, 1_B\rangle$ by an arbitrary element of $K \rtimes_\phi B$ gives:

$$\begin{aligned}\langle l, b\rangle\, \langle k, 1_B\rangle\, \langle l, b\rangle^{-1} &= \langle l({}^{b}k),\, b\rangle\, \langle {}^{b^{-1}}(l^{-1}),\, b^{-1}\rangle \\ &= \langle l({}^{b}k)({}^{b}({}^{b^{-1}}(l^{-1}))),\, bb^{-1}\rangle \\ &= \langle l({}^{b}k)l^{-1},\, 1_B\rangle\end{aligned}$$

which is another element of the subgroup.

3.8 Recognition theorem for semidirect products *If $A = N \rtimes S$, i.e., $N \cap S = \mathbf{0}$ and $A = NS$, then there is a morphism $\phi : S \to \mathrm{Aut}(N)$ given by $\phi_b(k) = bkb^{-1}$ such that $A \cong N \rtimes_\phi S$.*

Proof: First, note that ϕ_b is an automorphism with inverse $\phi_{b^{-1}}$. As above, we write $\phi_b(k)$ as ${}^bk$. Every element of A can be written as kb for $k \in N$ and $b \in S$. Since $K \cap S = \mathbf{0}$, this representation is unique. Moreover,

$$(k_1b_1)(k_2b_2) = k_1b_1k_2({b_1}^{-1}b_1)b_2 = k_1(b_1k_2{b_1}^{-1})b_1b_2 = k_1({}^{b_1}k_2)b_1b_2$$

It is an easy step to write $k_1b_1 \in A$ as $(k_1, b_1) \in N \rtimes_\phi S$. ■

The equational steps displayed above give us insight into the external definition of semidirect products. We are trying to write the product $(k_1b_1)(k_2b_2)$ in the form kb. To do so, we need a way to move b_1 "past k_2." We achieve the result by introducing $({b_1}^{-1}b_1)$ in the middle, and then noticing that we obtain a conjugate of k_2. We generalize the conjugation to an arbitrary automorphism to obtain the official definition.

3.9 Semidirect products as split extensions Semidirect products of groups give rise to short exact sequences:

$$\mathbf{0} \longrightarrow K \xrightarrow{\iota_1} K \rtimes_\phi B \xrightarrow{\pi_2} B \longrightarrow \mathbf{0}$$

where ι_1 sends each $k \in K$ to $(k, 1_B)$ and π_2 is the projection of the B component. Thus $K \rtimes_\phi B$ is an *extension* of B by K.

It is in fact a *split extension*, *i.e.*, $\pi_2 : K \rtimes_\phi B \to B$ has a section $\iota_2 : B \to K \rtimes_\phi B$, *viz.*, the injection, satisfying $\iota_2; \pi_2 = \mathrm{id}_B$.

In fact, *semidirect products characterize all split extensions.* Whenever $K \xrightarrow{i} G \xrightarrow{p} B$ is a split extension with a section $j : B \to G$ for p, the image $N = i_*(K)$ is a normal subgroup of G and $S = j_*(B)$ is a subgroup. These subgroup satisfy the hypotheses of the recognition theorem (§**??**).

Old text: We can use the splitting j to define an action $\hat{j} : B \to \mathrm{Aut}(G)$ by $\hat{j}_b) : x \mapsto j(b)\, x\, j(b)^{-1}$. Note that $\hat{j}$ only uses the inner automorphisms of G and, so, $N = i_*(K)$ is closed under inner automorphisms by virtue of being a normal subgroup of G. So, the restriction of $\hat{j}_b$ to N gives an action $\phi : B \to \mathrm{Aut}(N)$. G is isomorphic to the semidirect product $N \rtimes_\phi B$.

Questions: Does this argument generalize to monoids? Is $K \rtimes_\phi B$ functorial in K and B?

3.10 Semidirect products of semigroups (external) Semidirect products of semigroups and monoids are defined in a similar way, using a homomorphism $\phi : B \to \mathrm{End}(A)$. (Cf. §1.33). We can use the elements of B as exponents for the elements of A: ${}^ba = {}^{\phi_b}a = \phi_b(a)$. In effect, we are providing a left action of B^{I} on A.

The semidirect product $A \rtimes_\phi B$ of semigroups A and B has the underlying set $A \times B$ and multiplication defined by:

$$\langle a_1, b_1\rangle \langle a_2, b_2\rangle = \langle a_1({}^{b_1}a_2),\, b_1b_2\rangle$$

The associativity and the presence of units (when A and B have units) follow the same way as in §3.7.[11]

[11] Is A a normal submonoid of $A \rtimes_\phi B$?

Dually, we can use a homomorphism $\phi : A \to \mathrm{End}(B)^{\mathrm{op}}$, and treat the elements of A as exponents for the elements of B: $b^a = b^{\phi_a} = \phi_a(b)$. In effect, we are providing a right action of A^{I} on B.

The semidirect product $A \ltimes_\phi B$ of semigroups A and B has the underlying set $A \times B$ and multiplication defined by:

$$\langle a_1, b_1\rangle \langle a_2, b_2\rangle = \langle a_1 a_2,\ (b_1^{a_2})b_2\rangle$$

Associativity:

$$\begin{array}{rcl}
\big(\langle a_1, b_1\rangle\langle a_2, b_2\rangle\big)\,\langle a_3, b_3\rangle & = & \langle a_1a_2,\ b_1^{a_2}b_2\rangle\,\langle a_3, b_3\rangle \\
& = & \langle a_1a_2a_3,\ (b_1^{a_2}b_2)^{a_3}b_3\rangle \\
& = & \langle a_1a_2a_3,\ b_1^{a_2a_3}b_2^{a_3}b_3\rangle \\
\langle a_1, b_1\rangle\,\big(\langle a_2, b_2\rangle\langle a_3, b_3\rangle\big) & = & \langle a_1, b_1\rangle\,\langle a_2a_3,\ b_2^{a_3}b_3\rangle \\
& = & \langle a_1a_2a_3,\ b_1^{a_2a_3}b_2^{a_3}b_3\rangle
\end{array}$$

Units (when A and B have units):

$$\begin{array}{rcl}
\langle 1_A, 1_B\rangle\,\langle a, b\rangle & = & \langle 1_A a,\ (1_B)^a b\rangle = \langle a, b\rangle \\
\langle a, b\rangle\,\langle 1_A, 1_B\rangle & = & \langle a1_A,\ (b^{1_A})1_B\rangle = \langle a, b\rangle
\end{array}$$

3.11 Wreath product of groups Let A and B be groups. The *wreath product* of A and B is defined as $A \wr B = A^{|B|} \rtimes_\phi B$ for the notions of $A^{|B|}$ and $\phi : B \to A^{|B|}$ specified below.

The set $K = A^{|B|}$ forms a group under component-wise multiplication:

$$(ff')(x) = f(x)\,f'(x) \qquad \overline{1}(x) = 1 \qquad (f^{-1})(x) = f(x)^{-1}$$

Define a "translation operator" ${}^b f$, for $b \in B$ and $f \in K$ by

$$({}^b f)(x) = f(xb)$$

(This is a translation operator in that it gives, at x, the value of f at a translated value xb. More vividly, if you think of the elements $f \in K$ as vectors $[a_x]_{x\in B}$ then ${}^b f$ is a translated vector giving $[a_{xb}]_{x\in B}$.)

The translation operator gives a homomorphism $\phi : B \to \mathrm{Aut}(A^{|B|})$ where $\phi_b(f) = {}^b f$. To verify that ϕ_b is an automorphism, note that

$$\begin{array}{rcl}
({}^b(f_1f_2))(x) & = & (f_1f_2)(xb) \;=\; f_1(xb)\,f_2(xb) \\
& = & ({}^b f_1)(x)\,({}^b f_2)(x) \;=\; (({}^b f_1)\,({}^b f_2))(x) \\
({}^b\overline{1})(x) & = & \overline{1}(xb) \;=\; 1 \;=\; \overline{1}(x) \\
({}^b(f^{-1}))(x) & = & (f^{-1})(xb) \;=\; (f(xb))^{-1} \;=\; (({}^b f)(x))^{-1} \;=\; ({}^b f)^{-1}(x)
\end{array}$$

To verify that ϕ is a homomorphism, note that

$$\begin{array}{rcl}
({}^{b_1b_2}f)(x) & = & f(xb_1b_2) \;=\; ({}^{b_2}f)(xb_1) \;=\; ({}^{b_1}({}^{b_2}f))(x) \\
({}^1 f)(x) & = & f(x1) \;=\; f(x) \\
({}^b({}^{b^{-1}}f))(x) & = & ({}^{b^{-1}}f)(xb) \;=\; f(xbb^{-1}) \;=\; f(x)
\end{array}$$

The last equation means $\phi_b \circ \phi_{b^{-1}} = \mathrm{id}$ showing that ϕ_b and $\phi_{b^{-1}}$ are mutually inverse.

3.12 General wreath product of groups Let A and B be groups and Ω a set with a left B-action $\cdot : B \times \Omega \to \Omega$. Then the elements of A^Ω form a group under component-wise multiplication:

$$(ff')(x) = f(x)\, f'(x) \qquad \overline{1}(x) = 1 \qquad (f^{-1})(x) = f(x)^{-1}$$

Equivalently, A^Ω is the direct product $\prod_{x\in\Omega} A$ with elements $[a_x]_{x\in\Omega}$ and component-wise multiplication.

The B-action on Ω extends to a left B-action on A^Ω, in fact, to a homomorphism $\phi : B \to \mathrm{Aut}(A^\Omega)$, given by

$$({}^b f)(x) = ({}^{\phi_b} f)(x) = f(b^{-1} \cdot x)$$

(Once again, this is a translation operator that sends $[f_x]_{x\in\Omega}$ to $[f_{b^{-1}\cdot x}]_{x\in\Omega}$.)

The proof that ϕ_b is an automorphism is similar to the previous paragraph. To verify that ϕ is a homomorphism, note that

$$\begin{array}{rcl} ({}^{b_1 b_2} f)(x) & = & f((b_1 b_2)^{-1} \cdot x) \;=\; f({b_2}^{-1} {b_1}^{-1} \cdot x) \;=\; f({b_2}^{-1} \cdot {b_1}^{-1} \cdot x) \\ & = & ({}^{b_2} f)({b_1}^{-1} \cdot x) \;=\; ({}^{b_1}({}^{b_2} f))(x) \end{array}$$

This gives a homomorphism $\phi : B \to \mathrm{Aut}(A^\Omega)$ defined by $\phi_b(f) = b \cdot f$. The *general wreath product* of A and B is $A \operatorname{Wr}_\Omega B = A^\Omega \rtimes_\phi B$.

The standard wreath product can be seen as a special case of this, by choosing the B-action $\cdot : B \times |B| \times |B|$ given by ...

The group B acts on itself via left multiplication: $b \cdot y = by$, which extends to an action on A^B by $(b \cdot f)(y) = f(b^{-1}y)$.

3.13 Wreath product of monoids Let A and B be monoids. The set $K = A^B$ forms a monoid under component-wise multiplication:

$$(ff')(x) = f(x)\, f'(x) \qquad \overline{1}(x) = 1$$

Define a "translation operator" ${}^b f$, for $b \in B$ and $f \in K$ by

$$({}^b f)(x) = f(xb)$$

The translation operator gives a homomorphism $\phi : B \to \mathrm{End}(A^B)$ where $\phi_b(f) = {}^b f$. To verify that ϕ_b is an endomorphism, note that

$$\begin{array}{rcl} ({}^b(f_1 f_2))(x) & = & (f_1 f_2)(xb) \;=\; f_1(xb)\, f_2(xb) \\ & = & {}^b f_1(x)\, {}^b f_2(x) \;=\; ({}^b f_1\, {}^b f_2)(x) \\ ({}^b \overline{1})(x) & = & \overline{1}(xb) \;=\; 1 \;=\; \overline{1}(x) \end{array}$$

To verify that ϕ is a homomorphism, note that

$$\begin{array}{rcl} ({}^{b_1 b_2} f)(x) & = & f(xb_1 b_2) \;=\; {}^{b_2} f(xb_1) \;=\; {}^{b_1}({}^{b_2} f)(x) \\ ({}^1 f)(x) & = & f(x1) \;=\; f(x) \end{array}$$

The *wreath product* of A and B is defined as $A \wr B = A^B \rtimes_\phi B$. Expanding out the definition of semidirect product involved, $A \wr B$ has $A^B \times B$ as the carrier and the multiplication operation given by:

$$(f_1, b_1)\,(f_2, b_2) = (f_1({}^{b_1} f_2),\, b_1 b_2)$$

3.14 General wreath product of monoids Let A and B be monoids and Ω a set with a right B-action $\cdot : \Omega \times B \to \Omega$. The set A^Ω forms a monoid under component-wise multiplication.

$$(ff')(x) = f(x)\, f'(x)$$

Moreover, the right-action of B on Ω extends to a left action on A^Ω (cf. §1.31) given by

$$({}^b f)(x) = ({}^{\phi_b} f)(x) = f(x \cdot b)$$

To verify that ϕ_b is an endomorphism, note that

$$\begin{array}{rcl} ({}^b f_1 f_2)(x) & = & (f_1 f_2)(x \cdot b) \;=\; f_1(x \cdot b)\, f_2(x \cdot b) \\ & = & ({}^b f_1)(x)\, ({}^b f_2)(x) \;=\; (({}^b f_1)\, ({}^b f_2))(x) \\ ({}^b \overline{1})(x) & = & \overline{1}(x \cdot b) \;=\; 1 \;=\; \overline{1}(x) \end{array}$$

The general wreath product $A \operatorname{Wr}_\Omega B$ is defined as the semidirect product $A^\Omega \rtimes_\phi B$. So, the multiplication in $A \operatorname{Wr}_\Omega B$ is given by

$$(f_1, b_1)\,(f_2, b_2) = (f_1({}^{b_1} f_2),\, b_1 b_2)$$

The standard wreath product $A \wr B$ is defined as the semidirect product $A^B \rtimes_\phi B$ using the fact that B has a right-action on itself.

Dually, if Ω has a left B-action $\cdot : B \times \Omega \to \Omega$, we can obtain a semidirect product $B \ltimes_\phi A^\Omega$ where $\phi : B \to \operatorname{End}(A^\Omega)^{\mathrm{op}}$ defined by:

$$f^b(x) = f^{\phi_b}(x) = f(b \cdot x)$$

4 Commutative monoids

4.1 Commutative monoids A commutative monoid K satisfies $xy = yx$ for all $x, y \in K$. In this case, it is conventional to write the binary operation as $+$, and call K an "additive monoid." A commutative group is called an "abelian group."

Given any monoid A, an element $k \in A$ is called a *central element* if $kx = xk$ for all $x \in A$. The submonoid of all central elements of A is called the *center* of A. The center is a commutative monoid.

A homomorphism $h : K \to L$ between commutative monoids must satisfy:

$$h(x + y) = h(x) + h(y) \qquad h(0) = 0$$

We will refer to such homomorphisms as *additive maps*. (One might also called them "linear maps," but that terminology conflicts with "linear" in the sense of preserving scalar multiplication. Cf. §7.5.) The collection of additive maps $\mathrm{Hom}(K, L)$ has a commutative monoid structure inheriting that of L:

$$(h_1 + h_2)(x) \stackrel{\text{def}}{=} h_1(x) + h_2(x)$$

Thus the category **CMon** is a "closed category." We note in §4.6 that it is in fact a symmetric monoidal closed category.

4.2 Commutative monoids are $\mathbb{N}$-modules By writing an n-fold sum $x + x + \cdots + x$ as a multiple $n \cdot x$, we can regard a commutative monoid as having an action of the semiring $\mathbb{N}$. In other words, commutative monoids are (left or right) $\mathbb{N}$-modules.

Similarly, abelian groups are (left or right) $\mathbb{Z}$-modules.

4.3 Endomorphism semirings The set of endomorphisms $\mathrm{End}(K)$ of a commutative monoid K becomes a *semiring* under pointwise addition $(f + g)(x) = f(x) + g(x)$ and composition $f \circ g$. The latter distributes over addition: $(f_1 + f_2) \circ g = f_1 \circ g + f_2 \circ g$ and $f \circ (g_1 + g_2) = f \circ g_1 + f \circ g_2$. If K is a commutative group, with additive inverses $-x$ then the endomorphisms form a *ring*.

Corresponding to the "exponent" notation used in §**??**, we think of the elements of $\mathrm{End}(K)$ as "scalars" α (with composition as multiplication and the pointwise addition as addition) and regard their application to elments of K as scalar multiplication $\alpha \cdot x$. The laws mentioned in §1.33 are now written as:

$$\begin{aligned} \alpha\beta \cdot x &= \alpha \cdot (\beta \cdot x) \\ 1 \cdot x &= x \end{aligned}$$

and the fact that α is an additive map says:

$$\begin{aligned} \alpha \cdot (x + y) &= (\alpha \cdot x) + (\alpha \cdot y) \\ \alpha \cdot 0_K &= 0_K \end{aligned}$$

The pointwise operation $+ : \mathrm{End}(K) \times \mathrm{End}(K) \to \mathrm{End}(K)$ becomes:

$$\begin{aligned} (\alpha + \beta) \cdot x &\stackrel{\text{def}}{=} (\alpha \cdot x) + (\beta \cdot x) \\ 0 \cdot x &\stackrel{\text{def}}{=} 0_K \end{aligned}$$

The multiplication in $\mathrm{End}(K)$ now distributes over addition on *both sides*:

$$\begin{array}{rclcrcl} (\alpha_1+\alpha_2)\beta & = & \alpha_1\beta+\alpha_1\beta & \qquad & \alpha(\beta_1+\beta_2) & = & \alpha\beta_1+\alpha\beta_2 \\ 0\beta & = & 0 & & \alpha 0 & = & 0 \end{array}$$

We can treat the elements of $\mathrm{End}(K)^{\mathrm{op}}$ as scalar multiples on the *right* in exactly the same way.

If $h : L \to \mathrm{End}(K)$ is a representation of a monoid L, then the elements of L can be treated as scalars for the elements of K in the same way:

$$b \cdot x = h(b) \cdot x$$

4.4 Biadditive Maps A *biadditive map* $h : K \times L \to M$ of additive monoids is a function on the underlying sets that is additive in each argument:

$$\begin{array}{rclcrcl} h(x_1+x_2,\, y) & = & h(x_1,y)+h(x_2,y) & \qquad & h(0,\, y) & = & 0 \\ h(x,\, y_1+y_2) & = & h(x,y_1)+h(x,y_2) & & h(x,\, 0) & = & 0 \end{array}$$

Note that a biadditive map is *not* an additive map because

$$h(x_1+x_2, y_1+y_2) = \sum_{i=1,2}\sum_{j=1,2} h(x_i,y_j)$$

and this is not the same as $h(x_1,y_1)+h(x_2,y_2)$. It is possible to define a tensor product $K \otimes L$ such that additive functions $K \otimes L \to M$ are the same as biadditive functions $K \times L \to M$. This we do, in the next paragraph.

4.5 Tensor product The tensor product of additive monoids K and L is another additive monoid $K \otimes L$ along with a biadditive map $u : K \times L \to K \otimes L$ such that every biadditive map $K \times L \to O$ to an additive monoid O factors through u. The construction is the same as that of the tensor product $K \otimes_{\mathbb{N}} L$ of semiring-modules (§8.8). But, we describe it explicitly.[12]

To construct $K \otimes L$, we first construct the *free additive monoid* F with $K \times L$ as the generators. Take F to be the set of all finite multisets over $K \times L$, which may be viewed as "multiplicity" functions $\varphi : K \times L \to \mathbb{N}$ with only a finite number of non-zero values. We define the addition operation of $K \times L$ as the multiset union and the 0 as the empty multiset. In terms of the multiplicity funcitons, the operations are:

$$\begin{array}{rcl} (\varphi_1+\varphi_2)(x,y) & = & \varphi_1(x,y)+\varphi_2(x,y) \\ 0(x,y) & = & 0 \end{array}$$

For $x \in K$ and $y \in L$, let $\{(x,y)\}$ denote the function corresponding to the singleton multiset, i.e., the funciton in F that is 1 for (x,y) and 0 everywhere else. Then every $\varphi \in F$ can be written as a finite linear combination:

$$\varphi = \sum_{x,y} \varphi(x,y) \cdot \{(x,y)\}$$

Define a function $u : K \times L \to F$ by $u(x,y) = \{(x,y)\}$. Every biadditive function $h : K \times L \to O$ can be expressed as $h = s \circ u$ for an additive function $s : F \to O$, given by

$$s(\varphi) = \sum_{x,y} \varphi(x,y) \cdot h(x,y)$$

[12] Is the tensor product the categorical coproduct in **CMon**?

In particular, $s\{(x, y)\} = h(x, y)$.

Consider the congruence relation on F generated by the equivalences:

$$\begin{array}{rclcrcl} \{(x_1 + x_2,\, y)\} & \equiv & \{(x_1,\, y)\} + \{(x_2,\, y)\} & \qquad & \{(0,\, y)\} & \equiv & 0 \\ \{(x,\, y_1 + y_2)\} & \equiv & \{(x,\, y_1)\} + \{(x,\, y_2)\} & & \{(x,\, 0)\} & \equiv & 0 \end{array}$$

Now, $K \otimes L$ is the quotient $F/\equiv$. We write the equivalence class of $\{(x, y)\}$ as $[x, y]$, even though the conventional notation is $x \otimes y$. The projection $(x, y) \mapsto [x, y]$ of type $K \times L \to K \otimes L$ is evidently a biadditive map. Every biadditive map $h : K \times L \to O$ uniquely factors through $K \otimes L$ as $h = K \times L \xrightarrow{u} K \otimes L \xrightarrow{h^*} O$ with the definition:

$$h^*[x, y] = h(x, y)$$

4.6 Symmetric monoidal closed structure The tensor unit is $\mathbb{N}$, which may regarded as the free additive monoid generated by a singleton set. This makes $\langle \mathbf{CMon}, \otimes, \mathbb{N}\rangle$ into a symmetric monoidal category.

The set $\mathrm{Hom}(L, M)$ of additive maps between additive monoids is in turn an additive monoid, with pointwise addition and zero. We state that there is an adjunction:

$$\mathrm{Hom}(K \otimes L,\, M) \;\cong\; \mathrm{Hom}(K,\, \mathrm{Hom}(L, M))$$

An additive map $K \otimes L \to M$ is the same as a biadditive map $K \times L \to M$. Given such a biadditive map h, we can hold the K argument fixed and obtain an additive map $L \to M$ given by $y \mapsto h(x, y)$. Call this map $h^*(x)$. Now, h^* itself is an additive function in its x argument. Hence, $h^* \in \mathrm{Hom}(K, \mathrm{Hom}(L, M))$.

4.7 Tensor product as coproduct The tensor product is the categorical coproduct in the category **CMon**. The "injections" $K \xrightarrow{i} K \otimes L \xleftarrow{j} L$ are given by $x \mapsto [x, 0_L]$ and $y \mapsto [0_K, y]$. (Note that they may not be injective in the set-theoretic sense.) Given another cocone $K \xrightarrow{f} O \xleftarrow{g} L$, we define $[f, g] : K \otimes L \to O$ by

$$[f, g]([x, y]) = f(x) + g(y)$$

This is a valid definition because $[x, y] = [x, 0_L] + [0_K, y]$

4.8 Abelian groups Biadditive Maps and tensor products for abelian groups are defined similarly. The only difference is that the free abelian group generated by $K \times L$ is a collection of functions $K \times L \to \mathbb{Z}$ that are 0 for all but finitely many elements. The tensor unit is $\mathbb{Z}$, giving a symmetric monoidal category $\langle \mathbf{Ab}, \otimes, \mathbb{Z}\rangle$. It is symmetric monoidal closed in the same way as **CMon**.

5 Actions

Much of this section is duplicated elsewhere.

5.1 Actions and modules A *left action* of a semigroup A on a set X is a function $\cdot : A \times X \to X$ such that the multiplication of A is respected:

$$(a_1 a_2) \cdot x = a_1 \cdot (a_2 \cdot x)$$

We also say that X is a left *A-module*. (Other common names for the concept are *A-operands* [Clifford and Preston, 1961], *A-acts* [Kilp et al., 2000], *A-sets* [Lambek and Scott, 1986], *A-polygons*, *A-systems* [Howie, 1976], *A-automata* and *transition systems*.)

A *right action* is given by a function $\cdot : X \times A \to X$ such that the multiplication of A is respected on the right:

$$x \cdot (a_1 a_2) = (x \cdot a_1) \cdot a_2$$

Note that a right action of A is nothing but a left action of A^{op}.

A *monoid action* $\cdot : A \times X \to X$ is a semigroup action that also satisfies $1_A \cdot x = x$. A *group action* $\cdot : A \times X \to X$ is just a monoid action. But, since A has inverses, we obtain that $a^{-1} \cdot x = y \iff a \cdot y = x$. (Group actions have also been called *vector systems* by Hoehnke.)

5.2 Self-action Every monoid A has an action on "itself," or more precisely, its own underlying set, given by the multiplication:

$$a \cdot x = ax$$

This is a left action. Similarly, a right action can be defined by:

$$x \cdot a = xa$$

These actions are *transitive*. We can go from any x to any y by the sequence $x \xleftarrow{x}_L 1_A \xrightarrow{y}_L y$. If A is a group, the transformation is more simply left multiplication by yx^{-1}.

5.3 Cayley's theorem *Every monoid A is isomorphic to a submonoid of the monoid of homomorphisms $[A \to A]$.*

Define $h : A \to [A \to A]$ by $h(a)(x) = ax$. $h(1_A) = id_A$. $h(ab) = \lambda x.\, abx = (\lambda y.\, ay) \circ (\lambda x.\, bx) = h(a) \circ h(b)$. h is injective. If $h(a) = h(b)$ then, since $h(a)(1_A) = a$ and $h(b)(1_A) = b$, $a = b$.

A similar result also holds for semigroups. If A is a semigroup, construct monoid A^I by adjoining a unit. It is isomorphic to a submonoid of $[A^I \to A^I]$. ...

5.4 Homomorphisms A *homomorphism* of semigroup actions $h : (X, A, \cdot) \to (Y, B, *)$ is a pair consisting of a function $h_0 : X \to Y$ and semigroup morphism $h_1 : A \to B$ such that

$$h_0(x \cdot a) = h_0(x) * h_1(a)$$

Even though this is a general definition, it is more common to consider homomorphisms where the semigroup is fixed. A *homomorphism of A-actions* $h : (X, A, \cdot) \to (Y, A, *)$ is a function $h : X \to Y$ such that

$$h(x \cdot a) = h(x) * a$$

Keeping the monoid fixed has its counterpart in the theory of modules and vector spaces, which also represent actions of algebraic structures.

A homomorphism of A-actions $h : X \to Y$ induces an equivalence relation on X:

$$x \sim_h x' \iff h(x) = h(x')$$

This is a congruence relation in that

$$x \sim_h x' \Longrightarrow a \cdot x \sim_h a \cdot x'$$

for all $a \in A$.

There are some applications of homomorphisms in state minimization of automata. However, there are not widely useful.

5.5 Relations *Relations* of semigroup actions $R : (X, A, \cdot) \leftrightarrow (X', A', \cdot')$ can be defined in a similar way, as pairs $(R_0 : X \leftrightarrow X',\ R_1 : A \leftrightarrow A')$ satisfying:

$$x \left[R_0\right] x' \wedge a \left[R_1\right] a' \Longrightarrow x \cdot a \left[R_0\right] x' \cdot' a'$$

As a special case, *relations of A-actions* keep the semigroup fixed and involve a single relation R_0 between the set components. Note that a congruence relation of A-actions is indeed a relation of A-actions.

5.6 Transformation semigroups A useful special case of semigroup actions is that of *transformation semigroups*, which are pairs (Q, T) where Q is a set and $T \subseteq [Q \to Q]$ is a subsemigroup of the semigroup of transformations on T. A transformation semigroup has an implicit action: $q \cdot a = a(q)$.

Since every semigroup action can be viewed as a semigroup morphism $h : T \to [Q \to Q]$ and the quotient semigroup $T/\sim_h$ has an isomorphic image in $[Q \to Q]$, the transformation semigroups have all the generality of semigroup actions.

5.7 Direct product The direct product of transformation semigroups is defined by:

$$(Q_1, T_1) \times (Q_2, T_2) = (Q_1 \times Q_2,\ T_1 \times T_2)$$

with the associated action:

$$(q_1, q_2) \cdot (a_1, a_2) = (a_1(q_1),\ a_2(q_2))$$

Note that this is really a *semigroup action* rather than a transformation semigroup because $T_1 \times T_2$ is not a subset of $[Q_1 \times Q_2 \to Q_1 \times Q_2]$.

The direct product corresponds to parallel composition of semiautomata. The two machines have internal states Q_1 and Q_2 and action sets T_1 and T_2 respectively.

5.8 Wreath product Recall from §3.13 that the wreath product $A \operatorname{Wr}_\Omega B$ of two monoids or semigroups is defined as the semidirect product $A^\Omega \rtimes_\phi B$ with the multiplication:

$$\langle f_1, b_1 \rangle \langle f_2, b_2 \rangle = \langle (b_2 \uparrow f_1) f_2,\ b_1 b_2 \rangle$$

where $(b \uparrow f)(x) = f(x \cdot b)$.

If $Q = (Q, A)$ and $X = (X, B)$ are transformation semigroups, their *wreath product* is given by:

$$Q \circ X = (Q \times X,\ A^X \rtimes_\phi B)$$

with the associated action:

$$(q, x) * \langle f, b\rangle = (q \cdot f(x),\ x \cdot b)$$

In effect, we have picked $\Omega = X$ and used the wreath product of the semigroups as the second ("upstream") semigroup.

We verify the identity of the semigroup actions:

$$\begin{array}{rcl}
(q,x) * (\langle f_1, b_1\rangle \langle f_2, b_2\rangle) & = & (q,x) * \langle (b_2 \uparrow f_1) f_2,\ b_1 b_2\rangle \\
& = & (q \cdot ((b_2 \uparrow f_1) f_2)(x),\ x \cdot b_1 b_2) \\
& = & (q \cdot (b_2 \uparrow f_1)(x) \cdot f_2(x),\ x \cdot b_1 \cdot b_2) \\
& = & (q \cdot f_1(x \cdot b_2) \cdot f_2(x),\ x \cdot b_1 \cdot b_2) \\
(q,x) * \langle f_1, b_1\rangle * \langle f_2, b_2\rangle & = & (q \cdot f_1(x),\ x \cdot b_1) * \langle f_2, b_2\rangle \\
& = & (q \cdot f_1(x) \cdot f_2(x \cdot b_1),\ x \cdot b_1 \cdot b_2) \\
& = & (f_1, b_1) \cdot (f_2(x)(q),\ b_2(x)) \\
& = & (f_1(b_2(x))(f_2(x)(q)),\ b_1(b_2(x)))
\end{array}$$

The wreath product corresponds to serial composition of semiautomata. The actions for the downstream machine Q are generated by the upstream machine X. So, the input action for the downstream machine depends on the state of the upstream one. This motivates the structure of the actions for the composite machine, of the form (f, b). When two such actions (f_1, b_1) and (f_2, b_2) are composed, the first action on the downstream machine is f_2 which depends on the state of the downstream machine. However, the second action depends also on the previous action on the upstream machine b_2 because that would have affected the next state of the downstream machine. Thus the second action is $\phi_{b_2}(f_1)$ which takes the action b_2 into account.

6 Automata

6.1 Semiautomata A *semiautomaton* is a triple $(Q, \Sigma, \cdot)$ where Q is a set (of "states"), Σ is a set (of "input events") and $\cdot : Q \times \Sigma \rightharpoonup Q$ is a partial function called the *transition function*. We also say that the pair $(Q, \cdot)$ is a Σ-*semiautomaton*. In case $\cdot$ is a total function, the semiautomaton is said to be *complete*.

The action of the semiautomaton can be extended to sequences of input events in the obvious way:

$$q \cdot a_1 a_2 \cdots a_n \;\simeq\; q \cdot a_1 \cdot a_2 \cdot \cdots \cdot a_n$$

We then obtain a *right monoid action* $(Q, \Sigma^*, \cdot)$ where Σ^* is the free monoid generated by Σ. Note that such a monoid action is nothing but a monoid morphism $\Sigma^* \to \mathbf{Pfn}(Q, Q)^{\mathrm{op}}$.

Monoid actions generalize the concept of semiautomata by using general monoids. The effect is to treat the input events in an "abstract way" so that sequences of events that have the same effect can be identified in the monoid.

6.2 Nondeterministic semiautomata A *nondeterministic semiautomaton* is is a triple $(Q, \Sigma, \cdot)$ where $\cdot : Q \times \Sigma \leftrightarrow Q$ is a relation. It is convenient to regard the transition relation as a function of type $\cdot : Q \times \Sigma \to \mathcal{P}(Q)$. If the relation is total, i.e., $q \cdot a \neq \emptyset$, the semiautomaton is said to be *complete*.

Once again, nondeterministic semiautomata can be seen as right monoid actions or, equivalently, monoid morphisms $\Sigma^* \to \mathbf{Rel}(Q, Q)^{\mathrm{op}}$.

6.3 Mealy machines A *Mealy machine* (which is called an *automaton* in [Ginzburg, 1968]) is a semiautomaton $A = (Q, \Sigma, \cdot)$ along with two further components $(\Gamma, *)$ where Γ is a set (of "output symbols") and $* : Q \times \Sigma \rightharpoonup \Gamma$ is a partial function (the "output function") such that

$$q * a \text{ defined} \implies q \cdot a \text{ defined}$$

We also call the triple $(Q, \cdot, *)$ a *Mealy machine of type* $\Sigma \Rightarrow \Gamma$.

The output function is extended to sequences $Q \times \Sigma^* \to \Gamma^*$ by defining:

$$\begin{aligned} q * \epsilon &= \epsilon \\ q * ax &= (q * a)(q \cdot a * x) \end{aligned}$$

[Ginzburg and Yoeli, 1965] use the notation $\bar{a} : Q \rightharpoonup Q$ for the partial function induced by a symbol $a \in \Sigma$ under the action of $\cdot$, and $\bar{a}_* : Q \rightharpoonup \Gamma$ for the partial function under the action of $*$.

A more abstract treatment of Mealy machines is to view them as monoid morphisms $\alpha : \Sigma^* \to [Q \rightharpoonup Q \times \Gamma^*]$ where $[Q \rightharpoonup Q \times \Gamma^*]$ is a monoid under the composition:

$$g \circ f = \lambda q.\, \mathbf{let}\ (q_1, y_1) \Leftarrow f(q);\ (q_2, y_2) \Leftarrow g(q_1)\ \mathbf{in}\ (q_2,\ y_1 y_2)$$

The unit of the monoid is $\lambda q.\, (q, \epsilon)$. Now, note that $\bar{a} = \pi_1 \circ \alpha_a$ and $\bar{a}_* = \pi_2 \circ \alpha_a$ or, equivalently, $\langle \bar{a}, \bar{a}_* \rangle = \alpha_a$.

6.4 Labelled transition systems Labelled transition systems offer an alternative view of semiautomata and Mealy machines.

A *labelled transition system* is a triple $(Q,\ \Sigma,\ \{\xrightarrow{a} \mid a \in \Sigma\})$ where Q is a state of ("states"), Σ is a set (of "labels") and each $\xrightarrow{a}$ is a binary relation on states. Such a system is equivalent to a nondeterministic semiautomaton $(Q, \Sigma, \cdot)$ where $\cdot : Q \times \Sigma \to \mathcal{P}Q$ is defined by $q \cdot a = \{\, q' \mid q \xrightarrow{a} q' \,\}$. If each $\xrightarrow{a}$ is a single-valued relation (but possibly partial) we get a deterministic semiautomaton.

A nondeterministic Mealy machine $(Q, \Sigma, \cdot, \Gamma, *)$ can be modelled as labelled transition system with labels from $\Sigma \times \Gamma$ (with the input symbol written above the arrow and the output symbol below the arrow):

$$q \xrightarrow[t]{a} q_1 \iff \big(q_1 \in q \cdot a\big) \wedge \big(t \in q * a\big)$$

Again, this turns into a deterministic Mealy machine if $\cdot$ and $*$ are single-valued.

When multiple transition systems are involved, we give them labels $A, B, \ldots$ and write $q \xrightarrow[t]{a} q_1 \pmod A$, $q \xrightarrow[t]{a} q_1 \pmod B$ etc.

6.5 Homomorphisms Given two Mealy machines A and B of type $\Sigma \Rightarrow \Gamma$, a homomorphism $h : A \to B$ is a function $h : Q_A \to Q_B$ such that, for all $a \in \Sigma$, (i) $\bar{a}^A; h \subseteq h; \bar{a}^B$ and (ii) $\bar{a}^A_* \subseteq h; \bar{a}^B_*$.

$$\begin{array}{ccc} Q_A & \xrightarrow{\bar{a}^A} & Q_A \\ {\scriptstyle h}\downarrow & \supseteq & \downarrow{\scriptstyle h} \\ Q_B & \xrightarrow{\bar{a}^B} & Q_B \end{array} \qquad \begin{array}{ccc} Q_A & \xrightarrow{\bar{a}^A_*} & \Gamma \\ {\scriptstyle h}\downarrow & \supseteq & \| \\ Q_B & \xrightarrow{\bar{a}^B_*} & \Gamma \end{array}$$

Homomorphisms of semiautomta are obtained by ignoring the second condition.

In the notation of labelled transition systems, these conditions can be stated more simply as [Sangiorgi, 2009, Sec. 4.1]:

$$q \xrightarrow[t]{a} q_1 \pmod A \implies h(q) \xrightarrow[t]{a} h(q_1) \pmod B$$

Since h is a total function, the only partiality involved is for the transition relations. So, the diagram on the left has the effect of saying that for every transition of $\bar{a}^A$ there is a corresponding transition of $\bar{a}^B$ for the images under h. The diagram on the right says that the corresponding output symbols are equal.

6.6 Weak homomorphisms Given two Mealy machines A and B of type $\Sigma \Rightarrow \Gamma$, a *weak homomorphism* in the sense of [Ginzburg and Yoeli, 1965, Ginzburg, 1968] is a binary relation $R : Q_A \leftrightarrow Q_B$, *total* on both Q_A and Q_B, such that, for all $a \in \Sigma$, (i) $R^\smile; \bar{a}^A \subseteq \bar{a}^B; R^\smile$ and (ii) $R^\smile; \bar{a}^A_* \subseteq \bar{a}^B_*$.

$$\begin{array}{ccc} Q_A & \xrightarrow{\bar{a}^A} & Q_A \\ {\scriptstyle R^\smile}\uparrow & \subseteq & \uparrow{\scriptstyle R^\smile} \\ Q_B & \xrightarrow{\bar{a}^B} & Q_B \end{array} \qquad \begin{array}{ccc} Q_A & \xrightarrow{\bar{a}^A_*} & \Gamma_A \\ {\scriptstyle R^\smile}\uparrow & \subseteq & \| \\ Q_B & \xrightarrow{\bar{a}^B_*} & \Gamma_B \end{array}$$

Using the notation of labelled transition systems, these conditions can be stated as:

$$q \,\big[R\big]\, q' \ \wedge\ q \xrightarrow[t]{a} q_1 \pmod A \implies \exists q_1' \in Q_B.\ q' \xrightarrow[t]{a} q_1' \pmod B \ \wedge\ q_1 \,\big[R\big]\, q_1'$$

Once again, since the relation R is *total* on both Q_A and Q_B, the only partiality involved is for the transition relations. Despite the apparent reversal of the partial order $\subseteq$ in the diagrams, they still say that, for every transition of $\bar{a}^A$ there is a corresponding transition of $\bar{a}^B$. However, since we are now dealing with many-to-many correspondences R, the existence of a transition of $\bar{a}^B$ also involves the existence of a *target state* q_1' and this target state must be related to the target state q_1 of the transition of $\bar{a}^A$.

Note that the bitotality of the relation R is not required in the relational formulation. This seems to be a side effect of the "evil" practice of composing correspondences with transitions, which play very different conceptual roles.

To make sense of the existential quantifier for q_1', we return to the abstract view of Mealy machines as monoid actions $\Sigma^* \to [Q \rightharpoonup Q \times \Gamma^*]$ and amend it to $\Sigma^* \to [Q \rightharpoonup_{\subseteq} Q \times \Gamma^*]$ using a new type operator "$\rightharpoonup_{\subseteq}$". The relation operator $\rightharpoonup_{\subseteq}$ is defined as follows:

$$f \left[R \rightharpoonup_{\subseteq} S\right] f' \iff \forall x, x'.\, x \left[R\right] x' \wedge f(x) \neq \emptyset \implies f'(x') \neq \emptyset \wedge f(x) \left[S\right] f'(x')$$

The relation $[R \rightharpoonup_{\subseteq} S]$ can also be viewed as $[R \to \mathcal{P}^1_{\subseteq} S]$ where $\mathcal{P}^1_{\subseteq} S$ is defined by:

$$u \left[\mathcal{P}^1_{\subseteq} S\right] u' \iff \forall x \in u.\, \exists x' \in u'.\, x \left[S\right] x'$$

Now a weak homomorphism $R : A \leftrightarrow B$ is nothing but a relation $R \subseteq Q_A \times Q_B$ satisfying:

$$\alpha_A \left[I_{\Sigma^*} \to [R \rightharpoonup_{\subseteq} R \times I_{\Gamma^*}]\right] \alpha_B$$

6.7 Coverings A *covering* [Holcombe, 1982, Sec. 2.4] of a semiautomaton $\mathcal{M}_A = (Q_A, \Sigma_A, \cdot)$ by another semiautomaton $\mathcal{M}_C = (Q_C, \Sigma_C, \cdot)$, denoted $\mathcal{M}_A \prec \mathcal{M}_C$, represents the idea that $\mathcal{M}_C$ implements the behaviour represented by $\mathcal{M}_A$. It is given by pair

$$\begin{array}{cc} Q_C & \Sigma_C \\ \big\downarrow{\scriptstyle h_0} & \big\uparrow{\scriptstyle h_1} \\ Q_A & \Sigma_A \end{array}$$

where h_0 is a *surjective* partial function and h_1 is a function such that

$$h_0(q') \cdot a \subseteq h_0(q' \cdot h_1(a))$$

Diagrammatically:

$$\begin{array}{ccc} Q_C & \xrightarrow{h_1(a)} & Q_C \\ {\scriptstyle h_0}\big\downarrow & \subseteq & \big\downarrow{\scriptstyle h_0} \\ Q_A & \xrightarrow[a]{} & Q_A \end{array} \qquad \begin{array}{ccccc} & & Q_C \times \Sigma_C & \xrightarrow{\cdot} & \\ & \nearrow{\scriptstyle Q_C \times h_1} & & & Q_C \\ Q_C \times \Sigma_A & & \subseteq & & \big\downarrow{\scriptstyle h_0} \\ & \searrow{\scriptstyle h_0 \times T_A} & & & \\ & & Q_A \times T_A & \xrightarrow{\cdot} & Q_A \end{array} \tag{6.1}$$

Note that h_0 and h_1 run in *opposite* directions.

The idea is that the semigroup action $\mathcal{M}_C = (Q_C, \Sigma_C, \cdot)$ can "simulate" (or *cover*) the semigroup action $\mathcal{M}_A = (Q_A, \Sigma_A, \cdot)$. So, a state of Q_A may be represented by one or more states in Q_C and all states of Q_A must be represented in this way. Hence, we require $h_0 : Q_C \to Q_A$ to be *surjective*. However, not all states of Q_C may participate in the representation. So, we allow h_0 to be *partial*. Every transformation in Σ_A should have a corresponding transformation in Σ_C so that its behaviour can be simulated. If the state q' represents a state $q \in Q_A$, and we

can run a transformation a on q to obtain a resulting state $q \cdot a$, then running the corresponding transformation $h_1(a)$ on q' should give a state that represents the same result.

The rationale for the given direction of the partial order is that, whenever a state $q' \in Q_C$ is in the domain of h_0 and the action a is defined for its image, $h_1(a)$ should be defined for q' and h_0 should be defined for the resulting state. In this sense, we ensure that the covering machine is defined in all cases that the covered machine is defined.

When there is such a pair (h_0, h_1), we say that $\mathcal{M}_C$ *covers* $\mathcal{M}_A$ and write $\mathcal{M}_A \prec \mathcal{M}_C$.

6.8 Coverings of semigroup actions A *covering* [Holcombe, 1982, Sec. 2.4] of a semigroup actions $\mathcal{M}_A = (Q_A, T_A, \cdot)$ by another action $\mathcal{M}_C = (Q_C, T_C, \cdot)$, denoted $\mathcal{M}_A \prec \mathcal{M}_C$ is given by a *single* surjective partial function

$$\begin{array}{c} Q_C \\ \big\downarrow h_0 \\ Q_A \end{array}$$

such that for every $a \in T_A$ there exists $a' \in T_C$ satisfying:

$$h_0(q') \cdot a \subseteq h_0(q' \cdot a'$$

Diagrammatically:

$$\begin{array}{ccc} Q_C & \xrightarrow{a'} & Q_C \\ {\scriptstyle h_0}\big\downarrow & \subseteq & \big\downarrow{\scriptstyle h_0} \\ Q_A & \xrightarrow{a} & Q_A \end{array}$$

Unlike for covering of semiautomata, it is not required to have a monoid homomorphism $T_C \to T_A$.

6.9 Eilenberg notions of covering Eilenberg [Eilenberg, 1976, I.4-5] uses a slightly more general definitions of covering. Let $(Q_A, T_A, \cdot)$ and $(Q_C, T_C, \cdot)$ be semigroup actions. Given a relation $\phi : Q_C \leftrightarrow Q_A$, a transformation $a \in T_A$ is said to be *covered* by a transformation $a' \in T_C$ if

$$\phi(q') \cdot a \subseteq \phi(q' \cdot a')$$

where $\phi(q')$ etc denote sets of elements and we use function application notation to mean direct image.

1. If, for each $a \in T_A$, there is an $a' \in T_C$ that covers it, then ϕ is said to be a "relation" of semigroup actions.
2. If ϕ is surjective, then it is said to be a *relational covering*.
3. If ϕ is a partial surjective function, then ϕ is said to be a *covering*.

The difference from the previous definition is that there is no semigroup morphism as a part of the covering, only a *mapping* whose *existence* is all that matters.

6.10 Substitution property A relation $R : \mathcal{M} \leftrightarrow \mathcal{M}$ on a semiautomaton $\mathcal{M} = (Q, \Sigma, \alpha)$ is said to satisfy the *substitution property* [Hartmanis and Stearns, 1966, Yeh, 1968] iff it is a logical relation of semiautomata. The condition is often written as

$$(\forall a \in \Sigma) \quad {\alpha_a}^{-1}; R; \alpha_a \ \subseteq \ R$$

Note that it means the same as $\alpha_a \left[R \rightharpoonup R\right] \alpha_a$. Such relations are also called SP-relations.

The relation R is said to satisfy the *dual substitution property* if

$$(\forall a \in \Sigma) \quad \alpha_a; R; {\alpha_a}^{-1} \ \subseteq \ R$$

Note that it means the same as ${\alpha_a}^{-1} \left[R \to \mathcal{P}R\right] {\alpha_a}^{-1}$ where the inverse transition function is treated as a function of type $Q \to \mathcal{P}Q$.

6.11 Congruence relation A *congruence relation* of a semiautomaton $\mathcal{M} = (Q, \Sigma, \alpha)$ is a logical equivalence relation (or an equivalence relation with the substitution property). Given such a congruence relation, we can partition Q into congruence classes such that each congurence class $[q]$ is mapped by α_a into another congruence class $[a \cdot q]$. A partition π of Q satisfying this property is called an *SP-partition* [Hartmanis and Stearns, 1966] or an *admissible partition* [Ginzburg, 1968]. SP-partitions and congruence relations are one-to-one with each other.

6.12 Generalized congruence relation A *generalized congruence relation* $R : \mathcal{M} \leftrightarrow \mathcal{M}$ in the sense of [Yeh, 1968, Yeh, 1970] is a logical relation that is reflexive and symmetric (but not necessarily transitive). Given such a relation, we can decompose Q into a collection of subsets π, where each subset contains states R-related to a particular state q, having the substitution property. Such a collection of subsets is called a *set system* [Hartmanis and Stearns, 1966], *admissible decomposition* [Ginzburg and Yoeli, 1965] or *SP cover* [Yeh, 1968]. A decomposition of π of Q is admissible if and only if:

1. $\bigcup_{C \in \pi} C = Q$.
2. for all $a \in \Sigma$ and $C \in \pi$, there exists $C' \in \pi$ such that $\alpha_a[C] \subseteq C'$.

Every admissible decomposition π determines a unique generalized congruence relation R_π:

$$q \left[R_\pi\right] q' \iff \exists C \in \pi.\ \{q, q'\} \subseteq C$$

However, multiple decompositions may represent the same generalized congruence relation.

6.13 Cascade product Given complete Mealy machines $\mathcal{M}' = (X, \cdot, *)$ of type $\Gamma' \Leftarrow \Sigma'$ and $\mathcal{M} = (Q, \cdot, *)$ of type $\Gamma \Leftarrow \Sigma$, where $\Sigma' = \Gamma$, their *cascade product* is a Mealy machine $\mathcal{M}' \circ \mathcal{M} = (X \times Q, \cdot, *)$ of type $\Gamma' \Leftarrow \Sigma$ where

$$\begin{array}{rcl} (x, q) \cdot a & = & (x \cdot (q * a),\ q \cdot a) \\ (x, q) * a & = & x * (q * a) \end{array}$$

We can run the cascade product machine on a sequence of input symbols by:

$$\begin{array}{rcl} (x, q) \cdot \epsilon & = & (x,\ q) \\ (x, q) * \epsilon & = & \epsilon \\ (x, q) \cdot as & = & (x \cdot (q * a),\ q \cdot a) \cdot s \ = \ (x \cdot (q * a) \cdot (q \cdot a * s),\ q \cdot a \cdot s) \\ (x, q) * as & = & (x \cdot (q * a),\ q \cdot a) * s \ = \ x \cdot (q * a) * (q \cdot a * s) \end{array}$$

The general formulas can be written as:

$$\begin{array}{rcl}(x,q)\cdot a_1a_2\cdots a_n & = & (x\cdot(q*a_1)\cdot(q\cdot a_1*a_2)\cdots(q\cdot a_1\cdots a_{n-1}*a_n),\ q\cdot a_1\cdots a_n)\\ (x,q)*a_1a_2\cdots a_n & = & x\cdot(q*a_1)\cdot(q\cdot a_1*a_2)\cdots(q\cdot a_1\cdots a_{n-2}*a_{n-1})*(q\cdot a_1\cdots a_{n-1}*a_n)\end{array}$$

6.14 Wreath product Given complete Mealy machines $\mathcal{M}' = (X, \cdot, *)$ of type $\Gamma' \Leftarrow \Sigma'$ and $\mathcal{M} = (Q, \cdot, *)$ of type $\Gamma \Leftarrow \Sigma$, where $\Sigma' = \Gamma$, their *wreath product* is a Mealy machine $\mathcal{M}' \wr \mathcal{M} = (X \times Q,\ \cdot,\ *)$ of type $\Gamma' \Leftarrow \Sigma'^{Q} \times \Sigma$ where

$$\begin{array}{rcl}(x,q)\cdot(f,a) & = & (x\cdot f(q),\ q\cdot a)\\ (x,q)*(f,a) & = & x*f(q)\end{array}$$

Wreath product is a generalization of the cascade product. To obtain the behaviour of the cascade product $\mathcal{M}' \circ \mathcal{M}$ on an input symbol a, run the wreath product $\mathcal{M}' \wr \mathcal{M}$ on the input pair $(\lambda q.\, q * a,\ a)$.

To run the wreath product machine on a sequence of input symbols, we use:

$$\begin{array}{rcl}(x,q)\cdot(f_1,a_1)(f_2,a_2)\cdots(f_n,a_n) & = & (x\cdot f_1(q)\cdot f_2(q\cdot a_1)\cdots f_n(q\cdot a_1\cdots a_{n-1}),\ q\cdot a_1\cdots a_n)\\ (x,q)*(f_1,a_1)(f_2,a_2)\cdots(f_n,a_n) & = & x\cdot f_1(q)\cdot f_2(q\cdot a_1)\cdots f_{n-1}(q\cdot a_1\cdots a_{n-2})*f_n(q\cdot a_1\cdots a_{n-1})\end{array}$$

From this we can determine that the product $(f_1, a_1)(f_2, a_2)$ in $\Sigma'^{Q} \times \Sigma$ should be defined by:

$$(f_1, a_1)\,(f_2, a_2) = (\lambda q.\, f_1(q)\, f_2(q \cdot a_1),\ a_1a_2)$$

6.15 Wreath product of transformation semigroups Given transformation semigroups $\mathcal{M}' = (X, T)$ and $\mathcal{M} = (Q, S)$, the wreath product is a transformation semigroup $\mathcal{M}' \wr \mathcal{M} = (X \times Q,\ T^Q \times S)$ with the action:

$$(x,q)\cdot(f,a) \ = \ (x\cdot f(q),\ q\cdot a)$$

The semigroup $T^Q \times S$ has the multiplication operation:

$$(f_1, a_1)\,(f_2, a_2) = (\lambda q.\, f_1(q)\, f_2(q \cdot a_1),\ a_1a_2)$$

We verify that it satisfies the associativity law:

$$\begin{array}{rcl}(x,q)\cdot((f_1,a_1)(f_2,a_2)) & = & (x,q)\cdot(\lambda q.\, f_1(q)\, f_2(q\cdot a_1),\ a_1a_2)\\ & = & (x\cdot f_1(q)\cdot f_2(q\cdot a),\ q\cdot a_1a_2)\\ ((x,q)\cdot(f_1,a_1))\cdot(f_2,a_2) & = & (x\cdot f_1(q),\ q\cdot a_1)\cdot(f_2,a_2)\\ & = & (x\cdot f_1(q)\cdot f_2(q\cdot a),\ q\cdot a_1\cdot a_2)\end{array}$$

7 Modules

7.1 Modules If X is a set and $\delta : A \to \mathrm{End}(X)$ is a monoid homomorphism, we obtain the identities:

$$\delta_{ab}(x) = \delta_a(\delta_b(x))$$
$$\delta_{1_A}(x) = x$$

The representation δ can be equivalently viewed as an action $\cdot : A \times X \to X$ with identities:

$$(ab) \cdot x = a \cdot (b \cdot x)$$
$$1_A \cdot x = x$$

X is called a (left) *A-module* or *A-act*. We refer to the elements of A as "scalars," those of X as "vectors" and "$\cdot$" as "scalar multiplication". We refer to the first identity above as the "associativity law." It allows us to omit the "$\cdot$" and write $a \cdot (b \cdot x)$ as simply abx.

A right A-module is defined similarly using $\mathrm{End}(X)^{\mathrm{op}}$ in place of $\mathrm{End}(X)$, i.e., as morphisms $\delta : A \to \mathrm{End}(X)^{\mathrm{op}}$, or equivalently as $\delta : A^{\mathrm{op}} \to \mathrm{End}(X)$. Scalar multiplication is represented as a function $\cdot : X \times A \to A$ which satisfies the "associativity" law $x \cdot (ab) = (x \cdot a) \cdot b$. Hence we often write $(x \cdot a) \cdot b$ as simply xab.

7.2 Bimodules Given monoids A and B, an *(A, B)-bimodule* is a set X with an action of A on the left and an action of B on the right, satisfying:

$$a(xb) = (ax)b$$

An (A, B)-bimodule is equivalent to an $(A \times B^{\mathrm{op}})$-module, or a representation $\delta : A \times B^{\mathrm{op}} \to \mathrm{End}(X)$. The coherence condition above is merely an unraveling of the homomorphism condition of δ for the case $(a, 1_B)(1_A, b) = (a, b) = (1_A, b)(a, 1_B)$. The action of $(a, 1_B)(1_A, b)$ on x is written as $a(xb)$ in the scalar multiplication notation and that of $(1_A, b)(a, 1_B)$ is written as $(ax)b$.

We also use the notation $X : A \rightsquigarrow B$ to represent the situation that X is an (A, B)-bimodule. Note that a left A-module is nothing but a bimodule $X : A \rightsquigarrow \mathbf{0}$ (where $\mathbf{0}$ is the one-element monoid) and a right A-module is nothing but a bimodule $X : \mathbf{0} \rightsquigarrow A$.

If A is a commutative monoid then every left (or right) A-module is automatically an (A, A)-bimodule, by defining $axb = (ab) \cdot x$.

7.3 Monoids as self-modules The monoid A itself is a (left and right) A-module, hence an (A, A)-bimodule. Scalar multiplication is just the restriction of the multiplication in A. Thus we have a type rule:

$$\frac{}{A : A \rightsquigarrow A}$$

To be more precise, we should write $|A| : A \rightsquigarrow A$ because it is the *underlying set* of the monoid A that has the bimodule structure. A product A^n is is also a (left and right) A-module and (A, A)-bimodule. The scalar multiplication works pointwise: $a \cdot (b_1, \ldots, b_n) = (ab_1, \ldots, ab_n)$ and $(b_1, \ldots, b_n) \cdot c = (b_1 c, \ldots, b_n c)$.

If $U \subseteq A$ is a submonoid of A, then any bimodule with an action of A on the left can be specialized to have an action of U instead, by restricting the scalars to U. This construction also works on the right. Thus we have the rules:

$$\frac{X : A \rightsquigarrow B}{X : U \rightsquigarrow B} \text{ if } U \subseteq A \qquad \frac{X : A \rightsquigarrow B}{X : A \rightsquigarrow V} \text{ if } V \subseteq B$$

Hence, whenever $U \subseteq A$ is a submonoid, one can regard the underlying set of the monoid A as a bimodule of all of these types: $A : U \rightsquigarrow A$, $A : A \rightsquigarrow U$ or $A : U \rightsquigarrow U$.

7.4 Zeros in modules An element θ of an A-module X such that $a\theta = \theta$ is a called a *zero* or a *sink* element of the module.

If the monoid A has a zero element 0, then every element of the form $0x$, for $x \in X$, will be a zero of X. If there is a unique such zero, we write it as 0 and call X an *A-module with zero*.

7.5 Linear maps A morphism of A-modules $f : X \to_A Y$ is a function satisfying:

$$f(ax) = a \cdot f(x)$$

We call it an *A-linear map*. (It is also called a *homogeneous* map in the literature, but our terminology shows the connection with ring-modules and vector spaces more clearly.) The category of (left) A-modules is denoted A-**Mod**. Similarly, the category of right A-modules is denoted **Mod**-A and its morphisms are annotated as $X \to^A Y$. The category of (A, B)-bimodules is denoted A-**Mod**-B with its morphisms annotated as $X \to^B_A Y$.

If A is a monoid with zero, the category of A-modules with zero, denoted A-$\mathbf{Mod}_0$, has morphisms that preserve the zero elements of the modules. (Similar notations apply to right-modules and bimodules.)

The collection of linear maps between left A-modules X and Y is denoted $\mathrm{Hom}_A(X, Y)$ and the collection of linear maps from X to itself is denoted $\mathrm{End}_A(X)$. Note that $\mathrm{End}_A(X)$ is a monoid under composition. The notations $\mathrm{Hom}^A(X, Y)$ and $\mathrm{Hom}^B_A(X, Y)$ are used for the hom-sets of right modules and bimodules.

7.6 Structure of linear maps Contrary to expectation, the collection $\mathrm{Hom}_A(X, Y)$ of linear maps between left A-modules does *not* form a left A-module. The attempt to define:

$$(af)(x) = a \cdot f(x)$$

fails to give a linear map: $(af)(bx) = a \cdot f(bx) = a \cdot b \cdot f(x) \neq b \cdot (af)(x)$. In essence, scalar multiplication cannot be inherited in a pointwise manner.

If A is a *commutative* monoid, then $\mathrm{Hom}_A(X, Y)$ is a left A-module (as well as a right A-module). This situation is familiar from linear algebra of vector spaces, but the apparent symmetry there is misleading. The situation can be understood more clearly by stating the affairs in terms of bimodules.

If $X : A \rightsquigarrow B$ and $Y : A \rightsquigarrow C$ are bimodules, then $\mathrm{Hom}_A(X, Y)$ can be given a (B, C)-bimodule structure. For any f in $\mathrm{Hom}_A(X, Y)$ define:

$$(bfc)(x) \;=\; f(xb) \cdot c \tag{7.1}$$

where we use the fact that X has a right action by B and Y has a right action by C. This does give an A-linear map:

$$(bfc)(ax) \;=\; f(axb) \cdot c \;=\; a \cdot f(xb) \cdot c \;=\; a \cdot (bfc)(x)$$

Note that:

$$(b'bfcc')(x) = f(xb'b) \cdot cc' = (bfc)(xb') \cdot c' = (b'(bfc)c')(x)$$

Thus, we have established the type rule:

$$\frac{X : A \rightsquigarrow B \quad Y : A \rightsquigarrow C}{\mathrm{Hom}_A(X, Y) : B \rightsquigarrow C} \tag{7.2}$$

If $X : A \rightsquigarrow C$ and $Y : B \rightsquigarrow C$ are bimodules, then $\mathrm{Hom}^C(X, Y)$ can be given a (B, A)-bimodule structure by defining:

$$(bfa)(x) = b \cdot f(ax) \tag{7.3}$$

We can verify that bfa is C-linear on the right:

$$(bfa)(xc) = b \cdot f(axc) = b \cdot f(ax) \cdot c = (bfa)(x) \cdot c$$

Thus, we have the type rule:

$$\frac{X : A \rightsquigarrow C \quad Y : B \rightsquigarrow C}{\mathrm{Hom}^C(X, Y) : B \rightsquigarrow A} \tag{7.4}$$

Note that the left action is by B rather than A.[13]

The apparent symmetry of *commutative* monoid modules can be understood using bimodules. If A is a commutative monoid then a left A-module X may be regarded as an (A, A)-bimodule. (Y may be regarded as a bimodule as well, but we do not use that fact.) Therefore, we have a derivation:

$$\frac{\dfrac{X : A \rightsquigarrow \mathbf{0}}{X : A \rightsquigarrow A} \quad Y : A \rightsquigarrow \mathbf{0}}{\mathrm{Hom}_A(X, Y) : A \rightsquigarrow \mathbf{0}}$$

showing that $\mathrm{Hom}_A(X, Y)$ has a left A-module structure. However, it is the *right* action of A on X that gives rise to the left action on $\mathrm{Hom}_A(X, Y)$, as can be seen in the verification:

$$(af)(xa') = a \cdot f(xa') = a \cdot f(x) \cdot a' = (af)(x) \cdot a'$$

7.7 Dual module If X is a left A-module, the linear maps $\mathrm{Hom}_A(X, A)$ into the monoid A form a *right* A-module.

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

This is called the *dual module* or *left dual* of X, and denoted X^*.[14] The dual of a right A-module Y is similarly a *left* A-module $Y^* = \mathrm{Hom}^A(A, Y)$.

Note that we have a contravariant functor $(-)^* : A\text{-}\mathbf{Mod} \to \mathbf{Mod}\text{-}A$. If $h : X \to_A Y$ is a linear map then $h^* : t \mapsto t \circ h$ is a linear map $Y^* \to^A X^*$. The double dual $(-)^{**}$ is therefore a covariant functor $A\text{-}\mathbf{Mod} \to A\text{-}\mathbf{Mod}$.

Define a function $\langle -, - \rangle : X \times X^* \to A$ by

$$\langle x, t \rangle = t(x)$$

It is linear in each argument in an appropriate way:

$$\begin{array}{rcccl} \langle kx,\, t \rangle & = & k \cdot t(x) & = & k \cdot \langle x, t \rangle \\ \langle x,\, tk \rangle & = & t(x) \cdot k & = & \langle x, t \rangle \cdot k \end{array}$$

[13] Can this be explained abstractly?

[14] Why is this the "left" dual?

Thus, we have a morphism $\omega_X : X \to_A \mathrm{Hom}^A(X^*, A)$ given by $\omega_X(x)(t) = t(x)$. Since $Hom^A(X^*, A)$ is nothing but X^{**}, we have given a map $\omega_X : X \to_A X^{**}$. This is a *natural transformation*. It is a straightforward verification to check that the following square commutes:

$$\begin{array}{ccc} X & \xrightarrow{\omega_X} & X^{**} \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle f^{**}} \\ Y & \xrightarrow{\omega_Y} & Y^{**} \end{array}$$

7.8 Duals of bimodules Suppose $X : A \rightsquigarrow B$ is bimodule. Since the monoid A has action on itself on both sides, i.e., $A : A \rightsquigarrow A$, we obtain the bimodule structure $X^* = \mathrm{Hom}_A(X, A) : B \rightsquigarrow A$.

$$(bfa)(x) = f(xa) \cdot b$$

X^* is called the *dual* of X. We have established the (derived) rule:

$$\frac{X : A \rightsquigarrow B}{X^* : B \rightsquigarrow A}$$

So, evidently, the dual of a left A-module (a bimodule $A \rightsquigarrow \mathbf{0}$) is a right A-module (a bimodule $\mathbf{0} \rightsquigarrow A$), and, similarly, the dual of a right B-module is a left B-module. If K is a commutative monoid, a K-module X can be regarded as a bimodule $X : K \rightsquigarrow K$. In this case, $X^* : K \rightsquigarrow K$ is another bimodule of the same type.

7.9 Kernel and annihilator If $f : X \to_A Y$ is a linear map, the *kernel* of f is an equivalence relation $\cong_f$ on X such that

$$x \cong_f x' \iff f(x) = f(x')$$

Clearly, $\cong_f$ is a congruence relation: if $x \cong_f x'$ then $ax \cong_f ax'$ because $a \cdot f(x) = a \cdot f(x')$.

Since A itself is a left A-module, we can consider kernels of linear maps $|A| \to_A X$ to a left module X. For a particular $x \in X$, let $\lambda_x : A \to X$ be the linear map

$$\lambda_x(a) = ax$$

The kernel of λ_x, identifies all $a, a' \in A$ such that $ax = a'x$:

$$a \cong_x a' \iff ax = a'x$$

The relation $\cong_x$ is called the *annihilator* of x.

For a subset $S \subseteq X$, the annihilator $\cong_S$ is the intersection of the annihilators of all its elements:

$$a \cong_S a' \iff \forall x \in S.\ ax = a'x$$

7.10 Spans and basis If $U \subseteq X$ is a finite subset of a module, then the *span* of U (also called the *orbit* of U) is the least submodule of X that contains U or, equivalently, it is the intersection of all submodules that contain U. We denote the span of U by $\langle U \rangle$. If $\langle U \rangle = X$ then we say that U *spans* X.

If U spans X and every element of X can be *uniquely* represented as $x = as$ for some $a \in A$ and $s \in U$, then U is called a *basis* for X. By "uniquely represented," we mean that if x can

be represented as $a_1 s_1$ and $a_2 s_2$ then $a_1 = a_2$ and $s_1 = s_2$. If X has a basis U, then we say that it is a *free module* generated by U.

It is easy to verify that any two bases of X are of the same cardinality.[15] So the cardinality of the basis may be referred to as the *rank* of the module.

7.11 Products The categorical product of two A-modules X_1 and X_2 has the cartesian product $X_1 \times X_2$ as the underlying set and pointwise scalar multiplication:

$$a(x_1, x_2) = (ax_1,\ ax_2)$$

The projections $\pi_i : X_1 \times X_2 \to_A X_i$ are linear. The pairing operation $\langle f, g\rangle : Z \to_A X_1 \times X_2$, given by $\langle f, g\rangle(z) = (f(z), g(z))$, is linear because

$$\begin{array}{rcl} \langle f, g\rangle(az) & = & (f(az),\ g(az)) = (a \cdot f(z),\ a \cdot g(z)) \\ & = & a(f(z), g(z)) = a \cdot \langle f, g\rangle(z) \end{array}$$

The terminal object is the singleton module $1 = \{\star\}$ with the action $a(\star) = \star$.

7.12 Coproducts The coproduct of A-modules X_1 and X_2 has the disjoint union $X_1 + X_2 = (\{\mathit{1}\} \times X_1) \uplus (\{\mathit{2}\} \times X_2)$ as its carrier and the evident scalar multiplication:

$$a(\mathit{1}, x) \ = \ (\mathit{1},\ ax) \qquad a(\mathit{2}, x) \ = \ (\mathit{2},\ ax)$$

The injections $\iota_i : X_i \to_A X_1 + X_2$ given by $\iota_i(x) = (i, x)$ are linear. Similarly, $[f_1, f_2] : X_1 + X_2 \to_A Z$ is linear:

$$[f_1,\ f_2](a(i, x)) \quad = \quad [f_1, f_2](i, ax) = f_i(ax) = a \cdot f_i(x) = a \cdot [f_1, f_2](i, x)$$

In the category A-$\mathbf{Mod}_0$, the coproduct $X_1 + X_2$ is the "amalgamated" sum where the two zero elements are identified: $(\mathit{1}, 0) \equiv (\mathit{2}, 0)$.

7.13 Tensor product of modules Let X_A be a right A-module and ${}_A Y$ a left A-module. A function $f : X \times Y \to S$ to a *set* S is called a *balanced map* if it satisfies:

$$f(xa,\ y) \quad = \quad f(x,\ ay)$$

We construct a *set* $X \otimes_A Y$ which gives a universal balanced map, i.e., a balanced map $u : X \times Y \to X \otimes_A Y$ such that every other balanced map from $X \times Y$ uniquely factors through u.

The elements of $X \otimes_A Y$ are equivalence classes of of $|X| \times |Y|$ under the equivalence relation ρ generated by

$$(xa,\ y) \ \equiv_\rho \ (x,\ ay)$$

Let $x \otimes y$ denote the equivalence class $[(x, y)]_\rho$. By taking $u(x, y) = x \otimes y$, we obtain a balanced map $u : X \times Y \to X \otimes_A Y$. Every balanced map $f : X \times Y \to S$ uniquely factors through u with the factor $f' : x \otimes_A Y \to S$ given by $f'(x \otimes y) = f(x, y)$.

[15] This is true for ring-modules, but what about monoid modules?

7.14 Tensor product of bimodules If ${}_AX_B$ and ${}_BY_C$ are bimodules then their tensor product $X \otimes_B Y$ has the structure of an (A, C)-bimodule. Its scalar multiplication is given by:

$$a \cdot (x \otimes y) \cdot c \;=\; ax \otimes yc$$

Thus the tensor product may be viewed as a form of composition, along with identities:

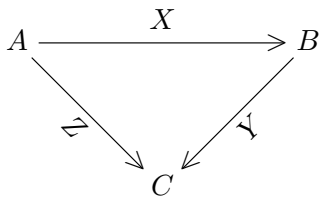
$$\frac{X : A \rightsquigarrow B \quad Y : B \rightsquigarrow C}{X \otimes_B Y : A \rightsquigarrow C} \qquad \frac{}{A : A \rightsquigarrow A}$$

We have canonical isomorphisms:

$$\begin{array}{rcl} (X \otimes_B Y) \otimes_C Z & \cong & X \otimes_B (Y \otimes_C Z) \\ X \otimes_B B & \cong & X \\ A \otimes_A X & \cong & X \end{array}$$

This notion of composition gives rise to a bicategory (cf. §17.6). The 0-cells are monoids, 1-cells are bimodules $X : A \rightsquigarrow B$ and 2-cells are bimodule homomorphisms. The horizontal composition is the tensor product, as indicated above. It is evidently not associative, but only associative up to coherent isomorphism.

7.15 Monoidal closed structure Let $X : A \rightsquigarrow B$, $Y : B \rightsquigarrow C$ and $Z : A \rightsquigarrow C$ be bimodules connected in a triangular pattern:



Then the bimodule homomorphisms $X \otimes_B Y \to^C_A Z$ are the same as balanced maps $X \times Y \to Z$. We have two other ways of representing these maps.

1. The bimodule $\mathrm{Hom}^C(Y, Z) : A \rightsquigarrow B$ is of the same type as $X : A \rightsquigarrow B$ and we can look at (A, B)-bimodule homomorphisms $f : X \to^B_A \mathrm{Hom}^C(Y, Z)$. The B-linearity of the homomorphism gives $f(xb) = f(x) \cdot b$. Since this is an element of $\mathrm{Hom}^C(Y, Z)$, we have $f(xb)(y) = (f(x) \cdot b)(y) = f(x)(by)$ by (7.3). Thus, we notice that these homomorphisms are the same as balanced maps $X \times Y \to Z$.

2. The bimodule $\mathrm{Hom}_A(X, Z) : B \rightsquigarrow C$ is of the same type as $Y : B \rightsquigarrow C$. If $f : Y \to^C_B \mathrm{Hom}_A(X, Z)$ is a bimodule homomorphism, we have $f(by)(x) = (b \cdot f(y))(x) = f(y)(bx)$ by (7.1). Once again, these morphisms represent balanced maps.

All said and done, we have the currying isomorphisms:

$$\mathrm{Hom}^B_A(X,\, \mathrm{Hom}^C(Y, Z)) \;\cong\; \mathrm{Hom}^C_A(X \otimes_B Y,\, Z) \;\cong\; \mathrm{Hom}^C_B(Y,\, \mathrm{Hom}_A(X, Z))$$

The first isomorphism is an adjunction:

$$- \otimes_B Y \;\dashv\; \mathrm{Hom}^C(Y, -) \;:\; A\text{-}\mathbf{Mod}\text{-}C \to A\text{-}\mathbf{Mod}\text{-}B$$

The second isomorphism is also an adjunction:

$$X \otimes_B - \;\dashv\; \mathrm{Hom}_A(X, -) \;:\; A\text{-}\mathbf{Mod}\text{-}C \to B\text{-}\mathbf{Mod}\text{-}C$$

The counits of these adjunctions are application (evaluation) maps:

$$\begin{array}{ll} \mathrm{apply}^Y : \mathrm{Hom}^C(Y,Z) \otimes_B Y \to Z & g \otimes y \mapsto g(y) \\ \mathrm{apply}_X : X \otimes_B \mathrm{Hom}_A(X,Z) \to Z & x \otimes f \mapsto f(x) \end{array}$$

As a special case, consider the situation where $A = C = \mathbf{0}$. In that case X_B is a right B-module, ${}_BY$ is a left B-module and Z is a set. The above isomorphisms reduce to:

$$\mathrm{Hom}^B(X,\, \mathrm{Hom}(Y,Z)) \;\cong\; \mathrm{Hom}(X \otimes_B Y,\, Z) \;\cong\; \mathrm{Hom}_B(Y,\, \mathrm{Hom}(X,Z))$$

representing the adjunctions:

$$\begin{array}{l} - \otimes_B Y \;\dashv\; \mathrm{Hom}(Y,-) \;:\; \mathbf{Set} \to \mathbf{Mod}\text{-}B \\ X \otimes_B - \;\dashv\; \mathrm{Hom}(X,-) \;:\; \mathbf{Set} \to B\text{-}\mathbf{Mod} \end{array}$$

7.16 Lemma *In the category of left A-modules, we have an isomorphism* $\mathrm{Hom}_A(A,X) \cong |X|$.

A linear map $f : A \to_A X$ satisfies $f(a \cdot x) = a \cdot f(x)$. So, in particular $f(a) = f(a \cdot 1_{|!A|}) = a \cdot f(1_A)$. In other words, the linear map f is uniquely determined by the value $f(1_A) \in |X|$.[16]

7.17 Cauchy duals Recall from §7.6 that for a bimodule $X : A \rightsquigarrow B$, the dual bimodule is of the form $X^* : B \rightsquigarrow A$. The module X is called a *Cauchy module* if there is a canonical isomorphism, for every left A-module Y:

$$\begin{array}{l} \rho^X_Y : X^* \otimes_A Y \cong \mathrm{Hom}_A(X,Y) \\ \rho^X_Y(t \otimes y) : x \mapsto t(x) \cdot y \end{array}$$

The terminology is due to [Street, 2007, Ch. 5]. When such an isomorphism exists, the module X^* is called the *left Cauchy dual* of X. (Note that Cauchy dual is a *stronger* concept than dual module.)

This definition covers two special cases of interest. If X is a left A-module, i.e., $X : A \rightsquigarrow \mathbf{0}$ then X^* is a right A-module (bimodule $X^* : \mathbf{0} \rightsquigarrow A$) and we call X a Cauchy module if $X^* \otimes_A Y$ is isomorphic to $\mathrm{Hom}_A(X,Y)$. Secondly, if X is a module of a commutative monoid K, i.e., $X : K \rightsquigarrow K$ then so is the dual module X^* and X is a Cauchy module if $X^* \otimes_K Y$ is isomorphic to $\mathrm{Hom}_K(X,Y)$ as a K-module.

We do not know if there exist any monoid modules with Cauchy duals. However, see §8.9 for applications to semiring and ring- modules.

7.18 Change of base 1: Restriction of scalars If $U \subseteq A$ is a submonoid of A then any A-module is automatically a U-module. We just restrict the scalar multiplication to scalars from U. (Cf. §7.2.) In particular, the larger monoid A itself is a U-module. This is called "change of base by restriction of scalars".[17]

More generally, if $h : U \to A$ is a monoid homomorphism, we obtain a functor

$$\mathbf{Mod}(h) : A\text{-}\mathbf{Mod} \to U\text{-}\mathbf{Mod}$$

(note the reversal of direction!) which sends an A-module $(Y,\cdot)$ to a U-module $(Y,*)$ with the same carrier. The scalar multiplication of the U-module is given by $u * x = h(u) \cdot x$ using

[16] Is there an abstract categorical reason for this fact?

[17] Can we formulate change of base for bimodules?

the scalar multiplication of the original A-module. A linear map $f : Y \to_A Y'$ of A-modules becomes a linear map of U-modules because $f(u * x) = f(h(u) \cdot x) = h(u) \cdot f(x) = u * f(x)$. This data gives a strict indexed category $\mathbf{Mod} : \mathbf{Mon}^{\mathrm{op}} \to \mathbf{CAT}$.

It is also useful to view this as a fibration $G : \mathbf{Mod} \to \mathbf{Mon}$ where $\mathbf{Mod}$ is the total category of all modules, with objects $(A, X, \cdot)$ consisting of a monoid A, a set X and an action $\cdot : A \times X \to X$. A morphism in $\mathbf{Mod}$ is a pair $(f, u) : (A, X, \cdot) \to (B, Y, \cdot)$ consisting of a monoid homomorphism $f : A \to B$ and a function $u : X \to Y$ such that $u(a \cdot x) = f(a) \cdot u(x)$. Note that $G^{-1}(A) = A\text{-}\mathbf{Mod}$ because a vertical morphism $(\mathrm{id}_A, u) : X \to_A Y$ is a morphism of modules: $u(a \cdot x) = a \cdot u(x)$.

Now, the fibration condition is that, given a monoid homomorphism $h : A \to B$ and a module $Y = (B, Y, \cdot)$ over B, there is a module $\hat{h}(Y)$ above A and cartesian morphism $\hat{h}(Y) \to Y$ over h. Define $\hat{h}(Y)$ to have the same carrier Y, with the scalar multiplication $a * y = h(a) \cdot y$. $(h, \mathrm{id}_Y) : \hat{h}(Y) \to Y$ is a cartesian morphism.

7.1 Orbits

7.19 Orbits and quotients The action of a monoid A on an A-module X induces a preorder:

$$x \preceq_A y \iff \exists a \in A.\, a \cdot x = y$$

The span of x (i.e., the span of $\{x\}$) is nothing but the set of elements above x in the preorder $\preceq_A$. These concepts are a generalization of those for multiplication actions by submonoids (§2.2).

The symmetric, transitive closure of the $\preceq_A$ is an equivalence relation:

$$x \sim_A y \Longleftarrow x \preceq_A y \vee x \succeq_A y$$

The equivalence class of x under this equivalence relation is called the *orbit* of x, and denoted ${}_A\lceil x \rceil$. Whenever y is in the orbit of x, x is also in the orbit of y or, more simply, the orbits of x and y are the same. So, the module X gets partitioned into disjoint orbits.

If A is a group acting on X then the span of x is already an equivalence class (because $a \cdot x = y \implies a^{-1} \cdot y = x$) and so an *orbit*. In this case, we write the orbit if x as ${}_A[x]$.

The set of all orbits is called the *quotient* of the action and denoted $A\backslash X$ or $\mathrm{Orbits}_A(X)$. quotient of a right action of A is correspondingly denoted X/A or $\mathrm{Orbits}^A(X)$.

These quotients are again A-modules, but *trivially* so. The action of A on $A\backslash X$ is $a \cdot {}_A[x] = {}_A[a \cdot x] = {}_A[x]$. In other words the elements of $A\backslash X$ are *fixed* under the action of A. (Cf. §7.21.) Equivalently, A is a *stabilizer* of every element of $A\backslash X$. (Cf. §7.23.)

7.20 Action on cosets Recall that every monoid A is a bimodule for itself: $A : A \rightsquigarrow A$. If $S \subseteq A$ is a submonoid, we can also regard A as a bimodule $A : S \rightsquigarrow A$ or a bimodule $A : A \rightsquigarrow S$ by restricting the multiplication (on the left or the right) as needed.

Considering the restriction on the left, we can take the quotient of the left action by S to obtain $S\backslash A : S \rightsquigarrow A$ whose elements are left orbits ${}_S\lceil x \rceil$. We simplify the notation to $S\backslash A$ but it is $S\backslash A$ that is meant.

$$\frac{A : S \rightsquigarrow A}{S\backslash A : S \rightsquigarrow A}$$

- The left action by S is *trivial*: $k \cdot {}_S\lceil x \rceil = {}_S\lceil kx \rceil = {}_S\lceil x \rceil$.

- The right action is inherited from A: ${}_S\lceil x\rceil \cdot a = {}_S\lceil xa\rceil$.

This right action is *transitive*, as in the case of self-actions (§5.2), because we can go from any ${}_S\lceil x\rceil$ to any ${}_S\lceil y\rceil$ by ${}_S\lceil x\rceil \xleftarrow{x}_R {}_S\lceil 1_A\rceil \xrightarrow{y}_R {}_S\lceil y\rceil$. Moreover, every element ${}_S\lceil a\rceil$ can be written as a right scalar multiple of ${}_S\lceil 1_A\rceil$, *viz.*, ${}_S\lceil 1_A\rceil \cdot a$. In other words, ${}_S\lceil 1_A\rceil$ is a *generator* for $S\backslash A$ as a right A-module.

Similarly, the set of right orbits A/S can be given a left action by A, *viz.*, $a \cdot \lceil x\rceil_S = \lceil ax\rceil_S$. This is summarized by the rule:

$$\frac{A : A \rightsquigarrow S}{A/S : A \rightsquigarrow S}$$

Once again, the right action by S is trivial and the left action by A is generated by $\lceil 1_A\rceil_S$.

7.21 Fixed points If A is a monoid acting on X, any element $x \in X$ that is fixed under the action of A, i.e., $A \cdot x = \{x\}$, is called a *fixed point* of the action. The set of all fixed points is denoted $\mathrm{Fix}_A(X)$ or X^A. If $S \subseteq A$ is a submonoid, we also speak about the fixed points of S, i.e., elements $x \in X$ such that $S \cdot x = \{x\}$. This set of fixed points is denoted X^S. They are studied in [Grassmann, 1979].

If A is a group, a fixed point of ${}_AX$ is an element whose orbit is a *singleton*. This observation fails when A is not a group. The orbit of a fixed point is not necessarily a singleton because $A^{-1} \cdot x$ may not be a singleton. For example, the submonoid $\{\mathrm{id}_{\mathbb{N}}, \bar{1}\} \subseteq \mathrm{End}(\mathbb{N})$ acting on $\mathbb{N}$ has a fixed point, *viz.*, 1. However, the orbit of 1 is $\mathbb{N}$.

7.22 Invariants An invariant of a module is a subset $S \subseteq X$ that is closed under the action of A, i.e., $A \cdot S \subseteq S$.

The invariant of an endomorphism $f : X \to X$ is defined similarly, as a submodule $S \subseteq X$ such that $f(S) \subseteq S$. In vector spaces, this leads to the notion of eigenvalues and eigenvectors. If a linear transformation has an eigenvalue then it has a one-dimensional invariant.[18]

7.23 Stabilizer Given an element x of a monoid-module ${}_AX$, the set of monoid elements that keep it fixed:

$$A_x = \mathrm{Stab}(x) = \{\, a \mid a \cdot x = x \,\}$$

is a submonoid of A. (Evidently, $1_A \cdot x = x$; $a \cdot x = x$ and $b \cdot x = x$ implies $ab \cdot x = x$.) If A is a group then A_x is also a subgroup: $a \cdot x = x$ implies $a^{-1} \cdot x = x$. The submonoid (subgroup) A_x is called the *stabilizer* of x. In group theory, it is also called the *isotropy subgroup* of x.

The notion of stabilizer can also be extended to subsets $U \subseteq {}_AX$.

$$\mathrm{Stab}(U) = \{\, a \mid a \cdot U = U \,\}$$

Since $\mathrm{Stab}(U) = \bigcap_{x \in U} A_x$, it is also a submonoid (subgroup) of A.

7.24 Subgroups as stabilizers It is an important fact that *every* subgroup S of A occurs as the stabilizer of some group action of A [Cohn, 1982, Sec. 3.3].

Take left actions of A. We use the *right* quotient $X = A/S$, which again has a left action $\cdot : A \times A/S \to A/S$ given by $k \cdot [a]_S = [k \cdot a]_S$. For an element $[a]_S \in A/S$, we can determine its

[18] This needs elaboration.

stabilizer as follows. If $[a]_S$ is fixed under left multiplication by some $k \in A$, *i.e.*, $k \cdot [a]_S = [a]_S$, that means that $ka \in [a]_S$. By right multiplication by a^{-1}, we obtain $k \in aSa^{-1}$. Conversely, for any $k \in aSa^{-1}$, we have $k \cdot [a]_S = [a]_S$. Thus the coset $[a]_S$ has the stabilizer aSa^{-1}.

In particular, $x = [1_A]_S$ has the stabilizer $A_x = S$.

Dually, considering right actions of A, the stabilizer of ${}_S[a]$ is $a^{-1}Sa$ and that of $x = {}_S[1_A]$ is precisely S.

7.25 The stabilizer quotient Given a group action ${}_AX$ and an element $x \in X$, the stabilizer A_x is a subgroup of A. Consider the right quotient A/A_x with cosets $[a]_{A_x}$ as elements. If a and b are in the same coset, *i.e.*, $[a]_{A_x} = [b]_{A_x}$, then a can be written as $b \cdot u$ for some $u \in A_x$. Then, $a \cdot x = b \cdot u \cdot x = b \cdot x$. The quotient A/A_x can be made into a left A-module by defining: $k \cdot [a]_{A_x} = [ka]_{A_x}$. There is a canonical A-linear map $h : A/A_x \to_A X$ defined by $h([a]_{A_x}) = a \cdot x$, which is well-defined by the above observation.

Note that $h([1]_{A_x}) = x$. This can be taken as an abstract definition of the stabilizer. The stabilizer of x is the largest subgroup of S of A with an A-linear map $h : A/S \to_A X$ that satisfies $h([1]_S) = x$.

Dually, given a right action X_A and an element $x \in X$, the stabilizer A_x gives an A-linear map $h : A_x\backslash A \to^A X$ defined by $h({}_{A_x}[a]) = x \cdot a$. It satisfies $h({}_{A_x}[1]) = x$.

7.26 Orbit-stabilizer theorem *For a group action of A on X and $x \in X$, there is a bijection $A/A_x \cong {}_A[x]$.*

The A-linear map $h : A/A_x \to_A X$ given in the previous paragraph sends each coset ${}_{A_x}[a]$ to an element $a \cdot x$ in the orbit of x. Thus, regarded as a map from A/A_x to ${}_A[x]$, it is surjective. It is also injective by the following argument: $a \cdot x = b \cdot x$ if and only if $b^{-1}a \cdot x = x$, i.e., $b^{-1}a \in A_x$. This is equivalent to saying that the cosets bA_x and aA_x are the same. Thus $aA_x \mapsto a \cdot x$ is a bijection.

If A is a finite group, this result can be put into a counting statement:

The number of elements in the orbit of x is the same as the index of its stabilizer group: ***card*** ${}_A[x] = [A : A_x]$.

As another easy corollary, we have:

If a group A acts transitively on X and $x \in X$, then $A/A_x \cong X$.

7.2 Modules of commutative monoids

The modules of commutative monoids K have a very different feel to those of non-commutative monoids. Since the multiplication is commutative, $kl = lk$, a K-module X can be regarded as a left K-module as well as a right K-module, or, in fact, as a (K, K)-bimodule, $X : K \rightsquigarrow K$. This last view is indeed the most appropriate.

7.27 Structure of K-modules When X and Y are K-modules, the collection morphisms $\mathrm{Hom}_K(X, Y)$ is another K-module. Defining:

$$(kf)(x) = k \cdot f(x)$$

makes kf a linear map: $(kf)(lx) = k \cdot f(lx) = k \cdot l \cdot f(x) = l \cdot k \cdot f(x) = l \cdot (kf)(x)$. This fact, can be seen more directly by viewing X and Y as bimodules $K \rightsquigarrow K$. The first rule of §7.8 shows that $\mathrm{Hom}_K(X, Y) : K \rightsquigarrow K$.

This makes the category K-**Mod** a *closed* category. We write $X \multimap Y$ for the internal hom, i.e., $\mathrm{Hom}_K(X, Y)$ regarded as a K-module.

The dual X^* of a K-module X is again a K-module. This is another fact that is immediate from the view $X : K \rightsquigarrow K$.

7.28 Duals of commutative monoid modules [The following comments doesn't belong here. Needs to be moved.]

The bilinear map $\langle -, - \rangle : X \times X^* \to K$ factors through the universal bilinear map to give natural transformation $\epsilon_X : X \otimes X^* \to K$. Explicitly, the definition is $\epsilon_X(x \otimes t) = t(x)$.

For X^* to be a dual object in the categorical sense, we also need a natural transformation $\eta_X : K \to X^* \otimes X$. This does not exist in general. See §10.1 for retrieving the situation in case of semiring-modules.

7.29 Tensor product of K-modules The tensor product of two K-modules $X \otimes_K Y$ is another K-module, as demonstrated by the first rule of §7.14. The action of K on $X \otimes_K Y$ is given by:

$$k \cdot (x \otimes y) \;=\; (kx) \otimes y \;=\; x \otimes (ky)$$

In addition to the standard notion of balanced maps from $X \times Y$ to sets C, we have a notion of *bilinear maps* $f : X \times Y \to Z$ to K-modules Z. Such maps preserve scalar multiplication in each argument:

$$f(kx,\, y) = k \cdot f(x, y) \qquad f(x,\, ky) = k \cdot f(x, y)$$

Note that bilinear maps are always balanced (when regarded as functions of type $X \times Y \to |Z|$).

The universal balanced map $u : X \times Y \to X \otimes_K Y$ may now be seen to be bilinear:

$$u(kx,\, y) = kx \otimes y = k \cdot (x \otimes y) \qquad u(x,\, ky) = x \otimes ky = k \cdot (x \otimes y)$$

Since all bilinear maps are balanced, we expect that u is also universal among all the bilinear maps from $X \times Y$. Indeed, it is easy to see that the unique factor $f' : X \otimes_K Y \to Z$ of any bilinear $f : X \times Y \to Z$ defined by $f'(x \otimes y) = f(x, y)$ is linear.

7.30 Change of base 2: Extension of scalars If $M \subseteq K$ is an inclusion of commutative monoids then any *free* M-module X can be *extended* to a free K-module. Since X is a free module it is expressible as the span $\langle U \rangle$ of a basis taking coefficients from M. Its extension Y as a K-module is spanned by the same basis U, but with coefficients from K. There is an evident inclusion $X \subseteq Y$, treating the coefficients m of scalar multiples mx as belonging to K. Thus, Y "extends" X and this construction is called "change of base by extension of scalars."

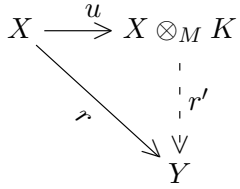
More generally, and without resorting to the choice of a basis, we can formulate the extension as follows. Let $h : M \to K$ be a homomorphism of monoids. We define a pseudo-functor $h_! : \mathbf{Mod}(M) \to \mathbf{Mod}(K)$ that maps M-modules X to K-modules. The monoid K can be regarded as a (K, M)-bimodule, with an action of K on the left (just multiplication) and an action of M on the right via h, given by $k * m = k \cdot h(m)$. To extend an M-module X to a K-module, we take the tensor product $h_!(X) = K \otimes_M X$ and treat it as a K-module by defining scalar multiplication $k' \cdot (k \otimes x) = k'k \otimes x$. The M-linear map

$$u : X \to K \otimes_M X \qquad x \mapsto 1_K \otimes x$$

can be used as the *extension map*. Note that

$$m \cdot u(x) = m \cdot (1_K \otimes x) = (1_K * m) \otimes x = 1_K \otimes mx = u(mx)$$

The extension map u is universal in the sense that any linear map $r : X \to_M Y$ to a K-module Y uniquely factors through u:



Note that the definition of u and the commutativity of the diagram would give:

$$r'(x \otimes k) = r'(k \cdot (x \otimes 1_K)) = k(r'(u(x))) = k(r(x))$$

Hence r', if it exists, is uniquely determined by r.

This gives a functor $h_! : M\text{-}\mathbf{Mod} \to K\text{-}\mathbf{Mod}$. It is left adjoint to $\hat{h} : K\text{-}\mathbf{Mod} \to M\text{-}\mathbf{Mod}$, i.e.,

$$K\text{-}\mathbf{Mod}[h_!(X), Y] \cong_{X,Y} M\text{-}\mathbf{Mod}[X, \hat{h}(Y)]$$

This makes the total category **Mod** *cofibred* over **Mon** (and so *bifibred*).

Further details may be found in [Mac Lane and Birkhoff, 1967, IX.11] and [Lang, 2000, XVI.4].

7.31 Algebras Let K be a commutative monoid. The following definition is based on [Mac Lane and Birkhoff, 1967]:[19]

A *K-algebra* is a K-module A that has an additional binary operation of multiplication making it a monoid and the two forms of multiplication "commute":

$$k(x_1 \cdot x_2) = (kx_1) \cdot x_2 = x_1 \cdot (kx_2)$$

Alternatively, a *K-algebra* is a monoid in the monoidal category $\langle K\text{-}\mathbf{Mod}, \otimes_K, K\rangle$, which means is that it is a K-module equipped with a *bilinear* multiplication operator:

$$\begin{aligned} (kx) \cdot z &= k(x \cdot z) \\ z \cdot (kx) &= k(z \cdot x) \end{aligned}$$

The multiplication operation then determines a formal multiplication morphism $\mu : A \otimes_K A \to A$. The unit morphism $\eta : K \to A$ maps $k \mapsto k1_A$, in particular $1_K \mapsto 1_A$. These definitions satisfy the standard monoid laws in monoidal categories.

A more classical definition of algebras is the following: A *K-algebra* is a monoid A together with a monoid homomorphism $\eta : K \to A$, called the *unit map*, such that $\eta(K)$ is contained in the center of A. The unit map then defines an action of K on A via $k \cdot x = \eta(k)x$.

Example: A prototypical example is the set of complex numbers, which has addition and multiplication defined by:

$$\begin{aligned} (x_1 + iy_1) + (x_2 + iy_2) &= (x_1 + x_2) + i(y_1 + y_2) \\ (x_1 + iy_1) \cdot (x_2 + iy_2) &= (x_1x_2 - y_1y_2) + i(x_1y_2 + x_2y_1) \end{aligned}$$

Moreover, complex numbers $x + iy$ can be viewed as vectors (x, y) in the complex plane, with pointwise addition and scalar multiplication by reals. The complex multiplication commutes

[19] Do we need commutativity to get a notion of algebras? Need to check [Street, 2007].

with scalar multiplication:

$$\begin{aligned} k(x_1 + iy_1) \cdot (x_2 + iy_2) &= (kx_1 + iky_1) \cdot (x_2 + iy_2) \\ &= (kx_1x_2 - ky_1y_2) + i(kx_1y_2 + kx_2y_2) \\ &= k((x_1 + iy_1) \cdot (x_2 + iy_2)) \end{aligned}$$

Thus complex numbers form a $\mathbb{R}$-algebra, where $\mathbb{R}$ is the field of reals.

7.32 Endomorphism algebras If X is a K-module, then the set $\mathrm{End}_K(X)$ of K-module endomorphisms is a K-module as well as a monoid under composition of endomorphisms. This makes it a K-algebra.

For example, the set of $n \times n$ matrices over K is a K-algebra with pointwise scalar multiplication forming the K-module structure and matrix multiplication forming the monoid structure.

7.33 Coalgebras and bialgebras Dually, a *K-coalgebra* is a comonoid in the monoidal category $\langle K\text{-}\mathbf{Mod}, \otimes_K, K\rangle$. So, it is a K-module A equipped with linear map $\delta : A \to_K A \otimes_K A$ and $\epsilon : A \to_K K$ such that the comonoid laws are satisfied.

A *K-bialgebra* is a K-module with both a monoid and comonoid structure $(A, \mu, \eta, \delta, \epsilon)$ such that μ and η are coalgebra maps, or, equivalently, δ and ϵ are algebra maps.

$$\begin{aligned} \delta(xy) &= \delta(x)\delta(y) & \delta(1_A) &= 1_A \otimes 1_A \\ \epsilon(xy) &= \epsilon(x)\epsilon(y) & \epsilon(1_A) &= 1_K \end{aligned}$$

8 Monoids and Modules with addition

8.1 Monoids and modules in symmetric monoidal categories As noted in §1.19, the concept of monoid is defined for any symmetric monoidal category $(\mathcal{C}, \otimes, I)$. Let A be a monoid in $\mathcal{C}$. An A-module is an object X of $\mathcal{C}$ along with a morphism $\bullet : A \otimes X \to X$ that satisfies the usual module laws at the level of morphisms.

If $\mathcal{C}$ is symmetric monoidal *closed*, which is often the case for the categories we deal with, $\mathrm{End}(X)$ is an object of $\mathbf{C}$ and a morphism $\bullet : A \otimes X \to X$ can be equivalently viewed as a morphism $\delta : A \to \mathrm{End}(X)$.

Monoids in $\mathbf{CPO}_\perp$ are *complete ordered monoids*. An A-module in $\mathbf{CPO}_\perp$ is a pointed cpo X along with an action $\bullet : A \otimes X \to X$. Note that $\mathbf{CPO}_\perp$ is a symmetric monoidal closed category. The collection $\mathrm{Hom}(X, Y)$ inherits the structure of $\mathbf{CPO}_\perp$ from Y in a "pointwise" manner. Hence, an A-module X can also be viewed as a strict continuous function $\delta : A \to \mathrm{End}(X)$.

Monoids in **CSL**, the category of complete join-semilattices, are called *quantales*. Since **CSL** is a subcategory of $\mathbf{CPO}_\perp$, this is a special case of complete ordered monoids.

Monoids in **CMon** are *semirings*. An A-module X in **CMon** is a morphism $\bullet : A \otimes X \to X$ or, equivalently, a bilinear morphism $A \times X \to X$. That means that scalar multiplication distributes over addition in both arguments:

$$\begin{array}{rclcrcl} a \bullet (x+y) & = & a \bullet x + a \bullet y & \qquad & a \bullet 0_X & = & 0_X \\ (a_1 + a_2) \bullet x & = & a_1 \bullet x + a_2 \bullet x & & 0_A \bullet x & = & 0_X \end{array}$$

Such an action is often called a *semimodule*. **CMon** is also symmetric monoidal closed. Hence, we can view a module as simply a monoid morphism $\delta : A \to \mathrm{End}(X)$.

In the category **Ab**, the situation is similar to the above. The monoids are *rings* and modules are what are traditionally known as "*modules*" in Algebra.

8.2 Notation In this section, unless otherwise stated, the letters A, B, C will stand for semirings and X, Y, Z for modules of semirings.

8.3 Semiring modules If $\delta : A \to \mathrm{End}(X)$ is a semiring homomorphism, we obtain the identities:

$$\begin{aligned} &\delta_{a+b}(x) = \delta_a(x) + \delta_b(x) \\ &\delta_{ab}(x) = \delta_a(\delta_b(x)) \\ &\delta_{1_A}(x) = x \end{aligned}$$

The representation δ can be equivalently viewed as an action $\cdot : A \times X \to X$ with identities:

$$\begin{aligned} &a(x+y) = ax + ay \\ &(a+b)(x) = ax + bx \\ &(ab)x = a(bx) \\ &1_A x = x \end{aligned}$$

X is called a *left A-module* and the operation "$\cdot$" is called *scalar multiplication*.

Right A-modules and (A, B)-bimodules are defined in an analogous fashion.

Modules of rings are defined similarly. They have additive inverses: $-x$ for the module elements as well as $-a$ for the scalars.

A module is called a *vector space* if the ring is in fact a field. The scalars then have (partial) multiplicative inverses $\frac{1}{a}$ whenever $a \neq 0_A$.

Example: Products of semirings For any semiring A, a finite product $A^n = A \times \cdots \times A$ is an A-module. The addition is pointwise, and the scalar multiplication is defined by $a \cdot (b_1, \ldots, b_n) = (ab_1, \ldots, ab_n)$. This example also generalizes to any power A^X for a set X.

Example: Commutative monoids and groups Any commutative monoid A may be regarded as an $\mathbb{N}$-module, where $n \cdot a$ is interpreted as an n-fold sum $a + \cdots + a$. A commutative group may be similarly regarded as a $\mathbb{Z}$-module.

Example: Sub-semirings If A is a semiring and $R \subseteq A$ is a sub-semiring, then A may be regarded as an R-module.

Example: Ideals If A is a semiring, any left ideal $L \subseteq A$ forms an A-module. Multiplication on the left is all that is required for a module. Similarly, right ideals form right A-modules.

Example: Polynomials For a commutative semiring K, the set of polynomials $K[x]$ in a single variable x (cf. §1.20) is evidently a K-module. Since it is also a semiring, it is an example of a K-algebra.[20]

Polynomials of degree less than n (for a fixed natural number n) also make a K-module. (But they are not closed under the convolution product, and hence do not make K-algebras.) This module is isomorphic to the module K^n.

Example: Monoid semirings More generally, every *monoid semiring* $A[M]$, consisting of functions $M \to A$ is an A-module with pointwise addition and scalar multiplication by A. It is also a semiring with convolution product as the multiplication. Hence it is an A-algebra.

More generally, every "set semiring" $A[X]$, consisting of functions $X \to A$, is an A-module with pointwise addition and scalar multiplication by A. It fails to be a semiring or an A-algebra, because it is not closed under convolution product.

Note that the semiring A itself is a sub-semiring of $A[M]$. It consists of polynomials with a single non-zero coefficient, for the term $x = 1_M$. Hence, this is just an example of the observation above that a semiring is a module for any of its subrings.

Example: Matrices The $m \times n$ matrices over a semiring A form an A-module with component-wise addition and scalar multiplication. Since such a matrix is nothing but a function $\mathbf{m} \times \mathbf{n} \to A$, this is just a special case of the modules $A[X]$.[21]

8.4 Linear transformations A morphism of A-modules $f : X \to_A Y$ is called a linear transformation. It preserves the additive monoid structure and the scalar multiplication:

$$\begin{aligned} f(x + y) &= f(x) + f(y) \\ f(0_X) &= 0_Y \\ f(ax) &= a \cdot f(x) \end{aligned}$$

Alternatively, f preserves all finite "linear combinations:"

$$f(a_1x_1 + \cdots + a_nx_n) \;=\; a_1f(x_1) + \cdots + a_nf(x_n)$$

The collection of linear transformations between A-modules X and Y is denoted $\mathrm{Hom}_A(X, Y)$.

[20] Need to check this.

[21] Does this generalize to $\mathrm{Hom}_A(A^n, A^m)$? Cf. §8.4.

8.5 Structure of linear transformations The pointwise addition of linear transformations $X \to_A Y$:

$$(f+g)(x) \stackrel{\text{def}}{=} f(x) +_Y g(x)$$

has an additive monoid structure (inherited from that of Y). We do not obtain a scalar multiplication operation for linear transformations in this way. (Cf. §7.5.) The attempt:

$$(af)(x) = a \cdot f(x)$$

fails to give a linear transformation because $(af)(bx) = ab \cdot f(x) \neq b \cdot (af)(x)$. Thus the linear transformations $\mathrm{Hom}_A(X,Y)$ form a *commutative monoid*.

However, endomorphisms $\mathrm{End}_A(X) = \mathrm{Hom}_A(X,X)$ have a binary multiplication operation obtained by composition. As a special case of endomorphism semirings (cf. §4.3), this multiplication distributes over addition. Hence, $\mathrm{End}_A(X)$ is always a *semiring*.

8.6 Linear transformations of K-modules If we consider linear transformations $\mathrm{Hom}_K(X,Y)$ for *commutative* semirings K, then the definition

$$(kf)(x) = k \cdot f(x)$$

does give a linear transformation: $(kf)(lx) = kl \cdot f(x) = lk \cdot f(x) = l \cdot (kf)(x)$. So, for commutative semirings K, $\mathrm{Hom}_K(X,Y)$ is again an K-module.

Endomorphims $\mathrm{End}_K(X)$ when K is commutative have both a K-module structure and a semiring structure. These two structures cohere in an important way to form a K-algebra. (Cf. §10.5).

8.7 Dual module As in §7.7, the linear transformations $\mathrm{Hom}_A(X,A)$, into the semiring A, have a scalar multiplication operation on the *right*, forming a right A-module.

$$(f \cdot a)(x) = f(x) \cdot a$$

This is called the *dual module* of X, and denoted X^*. The dual of a right A-module is similarly a *left* A-module.

Note that we have a contravariant functor $(-)^* : A\text{-}\mathbf{Mod} \to \mathbf{Mod}\text{-}A$. If $h : X \to_A Y$ is a linear transformation then $h^* : f \mapsto f \circ h$ is a linear transformation $Y^* \to^A X^*$. The double dual $(-)^{**}$ is therefore a covariant functor $A\text{-}\mathbf{Mod} \to A\text{-}\mathbf{Mod}$.

Define a function $\langle -,- \rangle : X \times X^* \to A$ by

$$\langle x, t \rangle = t(x)$$

It is linear in each argument in an appropriate way:

$$\begin{array}{rcccl} \langle a_1x_1 + a_2x_2,\, t\rangle & = & a_1t(x_1) + a_2t(x_2) & = & a_1\langle x_1,t\rangle + a_2\langle x_2,t\rangle \\ \langle x,\, t_1a_1 + t_2a_2\rangle & = & t_1(x)a_1 + t_2(x)a_2 & = & \langle x_1,t\rangle a_1 + \langle x_2,t\rangle a_2 \end{array}$$

Thus, we have a morphism $\omega_X : X \to_A \mathrm{Hom}^A(X^*, A)$ given by $\omega_X(x)(t) = t(x)$. Since $\mathrm{Hom}^A(X^*,A)$ is nothing but X^{**}, we have given a map $\omega_X : X \to_A X^{**}$. This is a *natural transformation*.

8.8 Tensor product If X is a right A-module and Y is a left A-module, then a map $h : X \times Y \to C$ to a commutative monoid C is called a *bilinear map* if it is a balanced map that preserves addition in each argument:

$$\begin{array}{rclcrcl} h(x_1 + x_2,\, y) & = & h(x_1, y) + h(x_2, y) & \qquad & h(xa,\, y) & = & h(x,\, ay) \\ h(x,\, y_1 + y_2) & = & h(x, y_1) + h(x, y_2) & & & & \end{array}$$

We construct a *commutative monoid* $X \otimes_A Y$ which gives a universal bilinear map, i.e., a bilinear map $u : X \times Y \to X \otimes_A Y$ such that every other bilinear map from $X \times Y$ uniquely factors through u.

To construct $X \otimes_A Y$, we first construct a free commutative monoid F with $X \times Y$ as generators. We will treat commutative monoids as $\mathbb{N}$-modules. Take F to be the set of all those functions $f : X \times Y \to \mathbb{N}$ which have only a finite number of non-zero values. (Equivalently, they are finite multisets over $X \times Y$.) These functions form a commutative monoid under term-wise addition:

$$(f_1 + f_2)(x, y) \quad = \quad f_1(x, y) + f_2(x, y)$$

For $x \in X$ and $y \in Y$, let $[x, y]$ denote the special function in F that is 1 for (x, y) and 0 everywhere else, i.e., it is a singleton multiset. Then every $f \in F$ can be written as a finite linear combination:

$$f = \sum_{x,y} f(x, y) \cdot [x, y]$$

Thus, the elements $[x, y]$ form a possibly infinite basis for F. Define a function $u : X \times Y \to F$ by $u(x, y) = [x, y]$. Every function $h : X \times Y \to C$ can be expressed as $h = u; s$ for a linear transformation $s : F \to C$, given by

$$s(f) = \sum_{x,y} f(x, y) \cdot h(x, y)$$

In particular, $s[x, y] = h(x, y)$.

Consider the congruence relation on F generated by the equivalences:

$$\begin{array}{rclcrcl} [x_1 + x_2,\, y] & \equiv & [x_1, y] + [x_2, y] & \qquad & [xa,\, y] & \equiv & [x,\, ay] \\ [x, y_1 + y_2] & \equiv & [x, y_1] + [x, y_2] & & & & \end{array}$$

Now, $X \otimes_A Y$ is the quotient $F/\equiv$. The equivalence class of $[x, y]$ is written as $x \otimes y$. The map $(x, y) \mapsto x \otimes y$ of type $X \times Y \to X \otimes_A Y$ is evidently bilinear. Every bilinear map $h : X \times Y \to C$ factors through $X \otimes_A Y$ by defining the unique factor $\overline{h} : X \otimes_A Y \to C$ as:

$$\overline{h}(x \otimes y) = h(x, y)$$

Thus we have shown that bilinear maps $X \times Y \to C$ are one-to-one with linear maps $X \otimes_A Y \to C$.

$$\mathrm{Bilin}(X, Y;\, C) \;\cong\; \mathrm{Hom}_A(X \otimes_A Y,\, C)$$

A typical element of $X \otimes_A Y$ is of the form $\sum_{i=1}^{k} x_i \otimes y_i$ where $x_i \in X$ and $y_i \in Y$. These elements satisfy:

$$\begin{array}{rclcrcl} (x + x') \otimes y & = & x \otimes y + x' \otimes y & \qquad & xa \otimes y & = & x \otimes ay \\ x \otimes (y + y') & = & x \otimes y + x \otimes y' & & & & \end{array}$$

directly as a result of the equivalences imposed on the elements of F.

This definition generalizes to *bimodules* in the same way as in §7.14.

8.9 Cauchy duals Recall from §7.17 that a bimodule $X : A \rightsquigarrow B$ has a dual bimodule $X^* : B \rightsquigarrow A$. The module X is called a *Cauchy module* if there is a canonical isomorphism of B-modules:

$$\rho_Y^X : X^* \otimes_A Y \cong_B \mathrm{Hom}_A(X, Y)$$
$$\rho_Y^X(t \otimes y) : x \mapsto t(x) \cdot y$$

for every left A-module Y. In this situation, the module X^* is called the *left Cauchy dual* of X.

The Cauchy dual is equipped with two canonical maps. First, there is always an A-linear map

$$\varepsilon_A : X \otimes_B X^* \to_A A$$

given by $\varepsilon_X(x \otimes t) = t(x)$. This is called the *counit* of the dual pair (X^*, X). Secondly, there is a canonical B-linear map called the *unit*:

$$\eta_X : B \to_B X^* \otimes_A X$$

which is determined as follows. Since ρ_X^X is an isomorphism, there is an element of $X^* \otimes_A X$ that corresponds to $\mathrm{id}_X \in \mathrm{Hom}_A(X, X)$. Write that element as $\sum_i t_i \otimes x_i$ and let it be the value of $\eta_X(1_B)$. Since B is spanned by 1_B, the entire map η_X is determined by this element.

From [Street, 2007, Ch. 5], we obtain the following simpler characterization of the Cauchy dual, which corresponds to the general definition of Cauchy duals in bicategories (cf. §17.9). *If X is a Cauchy module, the following composites are identity morphisms:*

$$X \cong X \otimes_B B \xrightarrow{X \otimes_B \eta_X} X \otimes_B X^* \otimes_A X \xrightarrow{\varepsilon_X \otimes_A X} A \otimes_A X \cong X$$
$$X^* \cong B \otimes_B X^* \xrightarrow{\eta_X \otimes_B X^*} X^* \otimes_A X \otimes_B X^* \xrightarrow{X^* \otimes_A \varepsilon_X} X^* \otimes_A A \cong X^*$$

Proof:

1. Since $\rho_X^X(\sum_i t_i \otimes x_i) = \mathrm{id}_X$, $\rho_X^X(\sum_i t_i \otimes x_i)(y) = \sum_i t_i(y) \cdot x_i = y$ for all $y \in X$.
2. Secondly, for any $u \in X^*$, $u(y) = u(\sum_i t_i(y) \cdot x_i) = \sum_i t_i(y) \cdot u(x_i)$ by linearity of u. Hence, $u = \sum_i t_i \cdot u(x_i)$.

The first composite is now calculated as follows:

$$y = y \otimes 1_B \longmapsto y \otimes (\textstyle\sum_i t_i \otimes x_i) = \sum_i y \otimes t_i \otimes x_i \longmapsto \sum_i t_i(y) \otimes x_i = \sum_i t_i(y) \cdot x_i = y$$

The second composite is calculated as:

$$u = 1_B \otimes u \longmapsto (\textstyle\sum_i t_i \otimes x_i) \otimes u = \sum_i t_i \otimes x_i \otimes u \longmapsto \sum_i t_i \otimes u(x_i) = \sum_i t_i \cdot u(x_i) = u$$

A further simplification is the following: *If X is a Cauchy module, the following composite is the identity morphism:*

$$X \cong X \otimes_B B \xrightarrow{X \otimes_B \eta_X} X \otimes_B X^* \otimes_A X \xrightarrow{\epsilon_X \otimes_A X} A \otimes_A X \cong X$$

8.10 Ring modules: Kernel and annihilator Let $f : X \to_A Y$ be a linear transformation of *ring* modules. The kernel of f is a congruence relation $\cong_f$ on X, as in §7.9. However, since A is a ring, the condition $f(x_1) = f(x_2)$ is equivalent to $f(x_1 - x_2) = 0_Y$. Thus, we can equivalently represent the kernel as the inverse image $f^{-1}(0_Y)$ of 0_Y. This is a *submodule* of X, because it is closed under addition and scalar multiplication.

As in §7.9, we have linear maps $\lambda_X : |A| \to_A X$ given by left multiplication $a \mapsto ax$. The kernel of λ_X is denoted as the *annihilator* $\mathrm{Ann}_A(x)$:

$$\mathrm{Ann}_A(x) = \{\, a \mid ax = 0_X \,\}$$

For a subset $S \subseteq X$, the annihilator $\mathrm{Ann}_A(S)$ is defined as the intersection of the annihilators of all $x \in S$:

$$\mathrm{Ann}_A(S) = \{\, a \mid \forall x \in S.\, ax = 0_X \,\}$$

8.11 Spans If $U \subseteq X$ is a finite subset of a module, then the *span* of U is the least submodule of X that contains U or, equivalently, it is the intersection of all submodules that contain U. It is also called the *linear span* or the *linear hull* of U.

If U is finite, with elements $u_1, \ldots, u_n$, then the span of U consists of the values of all its linear combinations:

$$a_1u_1 + \cdots + a_nu_n$$

It is a submodule of X because it is closed under addition and scalar multiplication:

$$\begin{array}{rcl} (\sum_{i=1}^n a_iu_i) + (\sum_{i=1}^n b_iu_i) & = & \sum_{i=1}^n (a_i + b_i)u_i \\ b \cdot (\sum_{i=1}^n a_iu_i) & = & \sum_{i=1}^n (ba_i)u_i \end{array}$$

and it is easy to see that it is the intersection of all submodules containing U.

A subset U that spans the entire module is called a *spanning set* of the module.

8.12 Linear independence A list of elements $U = \{u_1, \ldots, u_n\}$ is said to be *linearly independent* if, whenever two linear combinations are equal, their scalar coefficients are equal:

$$\textstyle\sum_{i=1}^n a_iu_i = \sum_{i=1}^n b_iu_i \implies a_1 = b_1 \wedge \cdots \wedge a_n = b_n$$

More abstractly, the linear transformation $L_X : A^n \to X$ given by $L_X(a_1, \ldots, a_n) = \sum_{i=1}^n a_iu_i$ is *injective* (a *monomorphism*).

In a *ring-module*, there are additive inverses and, so, the hypothesis of the condition above is equivalent to $\sum_{i=1}^n (a_i - b_i)u_i = 0_X$. Therefore, it is enough to require, for each list of scalars a_i,

$$a_1u_1 + \cdots + a_nu_n = 0_X \implies a_1 = \cdots = a_n = 0_A$$

This is often used as the definition of linear independence, but it only holds for ring-modules (including vector spaces).

8.13 Basis A linearly independent spanning set for a module is called a *basis* for the module. When such a basis exists, the module is a *free module* generated by the basis: $A^n \cong X$. The integer $n = |U|$ is called the *rank* of the module (also called *dimension* in the context of vector spaces).

All finite-dimensional vector spaces are free in this sense. Hence, all vector spaces of the same dimension are isomorphic to A^n and, so, isomorphic to each other. However, *this isomorphism is not natural*, depending on the choice of the basis vectors.

8.14 Free modules A free module X generated by a set $U = \{u_1, \ldots, u_n\}$ consists of linear combinations of the form $a_1u_1 + \cdots + a_nu_n$. Since U is fixed, any such linear combination can be identified with the list of scalars $(a_1, \ldots, a_n)$. The basis vectors $u_1, \ldots, u_n$ then become the unit elements $(1_A, 0_A, \ldots)$, $(0_A, 1_A, 0_A, \ldots)$ etc. Each unit vector spans a submodule of X that is isomorphic to A. Therefore, the free module X generated by n basis vectors is isomorphic to A^n.

A linear transformation $f : X \to Y$ between free modules is uniquely determined by its action on the basis vectors or, equivalently, the unit elements. If $(a_1, \ldots, a_n)$ is an element of X represented as a list of scalars, then

$$f(a_1, \ldots, a_n) = a_1 f(u_1) + \cdots + a_n f(u_n)$$

Since each $f(u_i)$ is an element of Y, it can be represented as list of scalars $(k_{1i}, \ldots, k_{mi}$. Hence, the entire linear transformation is uniquely determined by an $m \times n$ collection of scalars $\{k_{ij}\}$. This fact leads to the matrix representation of linear transformations.

8.15 Matrix representation Consider linear transformations of the form $t : A^n \to A^m$. It is really appropriate to think of the type as $A^m \leftarrow A^n$ because we write the argument of t on the right in the ordinary mathematical notation.

Every element of A^m can be identified with a list of m scalars $\mathbf{a} = (a_1, \ldots, a_m)$, which we write as a "column vector:"

$$\begin{bmatrix} a_1 \\ \vdots \\ a_m \end{bmatrix}$$

The linear transformation t is uniquely determined by its action on the unit elements $u_i = (0_A, \ldots, 1_A, \ldots, 0_A)$ of A^n. The image of a unit element is a column vector of length m. By arranging these column vectors into a rectangular matrix, we obtain:

$$\mathbf{p} = \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \vdots & \vdots \\ p_{m1} & \cdots & p_{mn} \end{bmatrix}$$

where the column vector $(\mathbf{p})_1 = [p_{i1}]_i = [p_{11} \ldots p_{m1}]$ is the image of u_1 and $(\mathbf{p})_n = [p_{in}]_i = [p_{1n} \ldots p_{mn}]$ is the image of u_n. In general, p_{ij} is the i'th component of the image of u_j under t.

To obtain the image $t(\mathbf{a})$ for an arbitrary vector $\mathbf{a} \in A^n$, note that $\mathbf{a} = a_1u_1 + \cdots + a_nu_n$. So, its image is $a_1t(u_1) + \cdots a_nt(u_n)$. Use the notation $\mathbf{w} = (w_1, \ldots, w_m)$ for the image. Since the i'th component of $t(u_j)$ is $p_{i,j}$, we obtain that $w_i = a_1p_{i1} + \cdots + a_np_{in}$, which is also written as the "dot product" $\mathbf{a} \cdot (\mathbf{p}^T)_i$. All said and done, we obtain the equation:

$$\begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \vdots & \vdots \\ p_{m1} & \cdots & p_{mn} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \sum_{j=1,n} a_j p_{1j} \\ \vdots \\ \sum_{j=1,n} a_j p_{mj} \end{bmatrix}$$

Note that the p_{ij} components move to the right in the dot products. This is in the end inevitable. When we work with commutative semirings (such as vector spaces) the order does not matter.

[Mac Lane and Birkhoff, 1967] work with *right* A-modules in order to avoid this change of order. In that case a vector $\mathbf{a} \in A^n$ would be of the form $\mathbf{a} = u_1a_1 + \cdots + u_na_n$ and the

components of the **w** vector would be of the form $\sum_{j=1,n} p_{ij}a_j$. As they explain it, the equation that t preserves right scalar multiplication reads:

$$t(\mathbf{x}a) = t(\mathbf{x})a$$

which appears as an "associativity" law and preserves the order of the letters. See the discussion §7.6 about linear maps preserving scalar multiplication on the right, but not on the left.

8.16 Duals of free modules If $X = A^n$ is a free A-module with basis $\{u_1, \ldots, u_n\}$ then the dual module $X^* = \mathrm{Hom}_A(X, A)$ is also a free A-module of rank n. Its basis elements are $\{u_1^*, \ldots, u_n^*\}$, defined by

$$u_i^*(u_j) = \delta_{ij} = \begin{cases} 1_A, & \text{if } i = j \\ 0_A, & \text{if } i \neq j \end{cases}$$

(The function δ_{ij} is called the "Kronecker delta function.")

Using the matrix representation of linear transformations, note that a transformation $f : A^n \to A$, or better $f : A \leftarrow A^n$, corresponds to a matrix with a single row:

$$\begin{bmatrix} p_1 & \cdots & p_n \end{bmatrix}$$

and its effect on an element $\mathbf{a} \in A^n$ is:

$$\begin{bmatrix} p_1 & \cdots & p_n \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = a_1p_1 + \cdots + a_np_n$$

Thus the basis vector u_i^* of the dual module is nothing but a row vector $[0_A \ldots 1_A \ldots 0_A]$ with a 1_A in the i'th position. Thus, we have shown that the dual X^* is again a free A-module of rank n. However, it is a *right* A-module. We construct a basis for the dual module in terms of unit rows $u_i = (0_A, \ldots, 1_A, \ldots, 0_A)$.

Recall that every free module A^n has an action of A on the left as well as the right. Hence, regarding $X = A^n$ as a right A-module, we have an isomorphism $X^* \cong X$ of right A-modules. All said and done, we have shown that the dual of a left A-module A^n is the right A-module A^n and *vice versa*.

This isomorphism $X^* \cong X$ is dependent on the choice of a basis for X and, so, *not natural*. On the other hand, there is a natural isomorphism $X^{**} \cong X$. Recall the morphism $\omega_A : X \to X^{**}$ is given by $\omega_A(x)(y) = y(x)$.

8.17 Infinite dimensions A (possibly infinite) subset $U \subseteq X$ is a basis if,

1. if every finite subset of U is linearly independent, and
2.

8.18 Biproducts The biproduct (categorical product as well as coproduct) of two A-modules X_1 and X_2 has the cartesian product $X_1 \times X_2$ as the underlying set. It can be endowed with an additive group structure and scalar multiplication pointwise:

$$(x_1, x_2) + (y_1, y_2) = (x_1 + y_1,\ x_2 + y_2)$$
$$a(x_1, x_2) = (ax_1,\ ax_2)$$

The projections $\pi_i : X_1 \times X_2 \to_A X_i$ are linear transformations; so are the injections $\iota_i : X_i \to_A X_1 \times X_2$ given by $\iota_1(x) = (x, 0_{X_2})$ and $\iota_2(x) = (0_{X_1}, x)$. The pairing operation $\langle f, g\rangle : Z \to_A X_1 \times X_2$, given by $\langle f, g\rangle(z) = (f(z), g(z))$, is a linear transformation because

$$\begin{aligned} \langle f,g\rangle(z_1+z_2) &= (f(z_1+z_2),\, g(z_1+z_2)) = (f(z_1)+f(z_2),\, g(z_1)+g(z_2)) \\ &= (f(z_1), g(z_1)) + (f(z_2), g(z_2)) = \langle f,g\rangle(z_1) + \langle f,g\rangle(z_2) \\ \langle f,g\rangle(az) &= (f(az),\, g(az)) = (a\cdot f(z),\, a\cdot g(z)) \\ &= a(f(z), g(z)) = a\cdot\langle f,g\rangle(z) \end{aligned}$$

Similarly, $[f, g] : X_1 \times X_2 \to Z$ given by $[f, g](x_1, x_2) = f(x_1) + g(x_2)$ is a linear transformation:

$$\begin{aligned} [f,g]((x_1,x_2)+(y_1,y_2)) &= [f,g](x_1+y_1,\, x_2+y_2) = f(x_1+y_1)+g(x_2+y_2) \\ &= f(x_1)+g(x_2)+f(y_1)+g(y_2) = [f,g](x_1,x_2)+[f,g](y_1,y_2) \\ [f,g](a(x_1,x_2)) &= [f,g](ax_1,\, ax_2) = f(ax_1)+g(ax_2) \\ &= a\cdot f(x_1)+a\cdot g(x_2) = a\cdot[f,g](x_1,x_2) \end{aligned}$$

8.19 Projective modules In any category, a *projective object* is an object X such that the functor $\mathrm{Hom}(X, -)$ preserves epimorphisms. In more detail, every morphism $f : X \to Z$ to an object Z factors through every epimorphism $e : Z' \to Z$. For example, in **Set** this is always the case. If e is a surjection, then every element $z \in \mathrm{Im}\, f$ has some $z' \in Z'$ such that $e(z') = z$. So, we can define a factor $f' : X \to Z'$ by setting $f'(x)$ to be some z' such that $e(z') = f(x)$.

For A-modules, this is not necessarily the case.

8.20 Ideals of rings An ideal I of a ring A is an additive subgroup of A that is closed under multiplication by the elements of A. In other words, it is an (A, A)-bimodule $I : A \rightsquigarrow A$. If A is a commutative ring, we can regard an ideal as simply an A-module.

A *principal ideal* is a bimodule generated by a single element. We use the notation (a) to denote the principal idea of an element a.

In the ring of integers, every ideal is a principal ideal. Such a ring is called a *principal ideal domain*. In other rings, ideals are not in general principal. However, all ideals can be added and, in commutative rings, they can also be multiplied.

$$\begin{aligned} I + J &= \{\, a+b \mid a \in I, b \in J \,\} \\ IJ &= \{\, ab \mid a \in I, b \in J \,\} \end{aligned}$$

Thus ideals are thought of as "ideal elements" of rings. It originates from the concept of "ideal numbers" defined by Kummer or algebraic number fields and generalized by Dedekind to rings in general.

8.21 Ideals and ring congruences *Ideals of A are the same as the ring congruence relations on A.*

Recall from §2.14ff. that, for an additive group A, a normal subgroup $I \subseteq A$ is one that satisfies $I + a = a + I$ for all $a \in A$. Since A is a commutative group, *any* subgroup of A is normal. We then obtain a group congruence relation $\sim_I$ on A given by $a \sim_I b \iff \exists i \in I.\ a + i = b$, which can also be written more simply as $b \in a + I$ or $a \in b + I$.

For $\sim_I$ to be a ring congruence, it also needs to be compatible with multiplication:

$$a \sim_I a' \wedge b \sim_I b' \implies ab \sim_I a'b'$$

Since $a' = a + i$ and $b' = b + j$ for some $i, j \in I$, we have $a'b' = (a+i)(b+j) = ab + ib + aj + ij$. By requiring that I be an (A, A)-bimodule, we have the sum $(ib + aj + ij)$ as an element of I. It then follows that $a'b' \in ab + I$, i.e., $ab \sim_I a'b'$. Thus ideals determine ring congruence relations.

Conversely, given a congruence relation $\equiv$ on A, the congruence class $I = [0]_\equiv$ forms an ideal. All other congruence classes of $\equiv$ are cosets of the form $a + I$. Note:

$$a \equiv b \iff 0 \equiv b - a \iff 0 \sim_I b - a \implies a \sim_I a + b - a \iff b \in a + I$$

Thus ring congruences are representable by ideals.

8.22 Ideals and semiring congruences An ideal I of a semiring A is similarly a commutative submonoid of A that is closed under multiplication by the elements of A. Thus an ideal I is an (A, A)-bimodule $I : A \rightsquigarrow A$.

Every ideal of a semiring A determines a semiring congruence on A.

The congruence relation $\sim_I$ of an ideal I is given by

$$a \sim_I b \iff \exists i, j \in I.\ a + i = b + j$$

This is called the *Bourne relation* of I [Golan, 1999, Ch. 6].

As noted in §2.15, a normal submonoid $I \subseteq A$ determines a monoid congruence relation $\sim_I = (\preceq_I \cup \succeq_I)^*$ on A. So $a \sim_I b$ if there is a sequence of elements

$$a = a_0 \sim_I a_1 \sim_I \cdots \sim_I a_n = b$$

where the successive steps alternate between $\preceq_I$ and $\succeq_I$. Since A is a commutative monoid, this is equivalent to saying that there exist $i, j \in I$ such that $a + i = b + j$. We prove this by induction on n:

- If $n = 0$, we take $i = j = 0$.
- If $n > 0$ then, by inductive hypothesis, there exist $i', j' \in I$ such that $a_0 + i' = a_{n-1} + j'$.

  If the last step is an instance of $\preceq_I$, then $a_{n-1} + k = a_n$ for some $k \in I$. We then have $a_0 + i' + k = a_{n-1} + j' + k = a_n + j'$. So, we can take $i = i' + k$ and $j = j'$.

  If the last step is an instance of $\succeq$, then $a_{n-1} = a_n + k$ for some $k \in I$. When then have $a_0 + i' = a_{n-1} + j' = a_n + k + j'$. So, we can take $i = i'$ and $j = j' + k$.

As in the previous paragraph, requiring that I be an (A, A)-bimodule is enough to ensure that $\sim_I$ is a semiring congruence.

8.23 On ideals The discussion of ideals in semigroups (Sec. 2.1) generalizes to rings and semirings because these are nothing but monoids internal to **Ab** and **CMon**.

The *quotient* A/I of a ring or semiring A by an ideal I consists of congruence classes $[a]_I$ of elements $a \in A$.

A ring or semiring A has two *improper* ideals, *viz.*, $\{0\}$ and A itself. All others are *proper* ideals.

A ring or semiring that has no proper ideal is said to be *simple*.

9 Commutative rings and Fields

In this section, we restrict attention to commutative rings, and use the letters $K, L, \ldots$ to denote them

9.1 Trivial rings The null object (initial as well as final object) in the category **CRng** is the ring $\{0\}$. We say that such a ring is *trivial* and all other rings are *nontrivial*.

A ring is trivial if and only if $1 = 0$. If $1 = 0$ then any element a of the ring can be written as $a = a \cdot 1 = a \cdot 0 = 0$ and it follows that 0 is the only element of the ring.

9.2 Division Recall that, in any semigroup, we have preorders $\preceq^L$ and $\preceq^R$ induced by the multiplication operation of the semigroup. In a commutative ring K, viewed as a commutative semigroup internal to **Ab**, the two preorders are the same. We write the common preorder $a \preceq b$ as $a \mid b$ and read it as "*a divides b*."

$$a \mid b \iff \exists k \in K.\, ka = b$$

If the factor k is unique whenever it exists, we call it the *quotient* of b divided by a, and denote it by $\frac{b}{a}$ or b/a. Note that there can be no quotient when $a = 0$ unless the ring is trivial, because $k \cdot 0 = 0$ for all $k \in K$.

The unit 1 is a minimal element in the "divides" preorder because $1 \mid a$ for all $a \in K$. Correspondingly, 0 is maximal element in the preorder because $a \mid 0$. If $a \mid b$ and $b \mid a$, then we say that a and b are *associates* of each other. The relation of being an associate is an equivalence relation.

If an element $a \neq 0$ has a multiplicative inverse then it is said to be *invertible*.[22] Note that the the inverse is necessarily unique (§1.7) and it is the same as the quotient $\frac{1}{a}$. The commutative ring has all quotients by invertible elements because we can take b/a to be $b \cdot a^{-1}$. Note: $(b/a) \cdot a = b \cdot a^{-1} \cdot a = b$. All invertible elements are associates of 1 because $a \mid 1$ and $1 \mid a$.

9.3 Integral domains and fields A nontrivial commutative ring is called a *field* if every non-zero element a has a multiplicative inverse. It is called an *integral domain* if it has a partial operation of division. Clearly, every field is an integral domain, but not conversely. For example, the set of integers forms an integral domain, but not a field.

Having a partial operation of division is equivalent to saying that the non-zero elements of the ring form a *cancellative monoid* under multiplication:

$$ka = k'a \wedge a \neq 0 \implies k = k' \tag{9.1}$$

The unique k such that $ka = b$ is then the quotient $\frac{b}{a}$. The condition is also equivalent to saying that there are no "zero divisors," i.e., no elements a that divide 0:

$$ab = 0 \implies a = 0 \vee b = 0 \tag{9.2}$$

or that the product of non-zero elements is non-zero:

$$a \neq 0 \wedge b \neq 0 \implies ab \neq 0 \tag{9.3}$$

[22] The invertible elements are also called "units," but we avoid this terminology.

To see the equivalence of (9.1) and (9.2), suppose K is a nontrivial commutative ring that satisfies the cancellation law. By setting $k' := 0$, we obtain $ka = 0 \wedge a \neq 0 \implies k' = 0$, which is equivalent to (9.2). Conversely, suppose K is a nontrivial commutative ring with no zero divisors and $a \neq 0$. We can write $ka = k'a$ as $(k - k') \cdot a = 0$. By (9.2), this implies $k - k' = 0$, i.e., $k = k'$.

Integral domains form a generalization of integers, and provide a natural setting for studying divisibility.

As a non-example, the product ring $\mathbb{Z} \times \mathbb{Z}$ is not an integral domain, because, for instance, the non-zero elements $(1, 0)$ and $(0, 1)$ multiply to zero.

9.4 Examples

- For a fixed positive integer $n > 1$, the real numbers of the form $a + b\sqrt{n}$ for integers a and b form an integral domain. Note that $a + b\sqrt{n} = 0$ if and only if $a = 0$ and $b = 0$. If $(a + b\sqrt{n})(a' + b'\sqrt{n}) = 0$ then $aa' + (ab' + a'b)\sqrt{n} + bb'n = 0$. This is possible only if both $a + b\sqrt{n}$ and $a' + b'\sqrt{n}$ are 0.
- Complex numbers of the form $a + bi$ for integers a and b form an integral domain. These are called *Gaussian integers*.
- For a fixed positive integer $n > 1$, the real numbers of the form $a + b\sqrt{n}$ for rational numbers a and b form a field. This field is denoted $\mathbb{Q}(\sqrt{n})$.
- The collection of polynomials $K[x]$ for an integral domain K is in turn an integral domain. If two polynomials have leading terms ax^n and bx^m, then the leading term of their product is abx^{n+m}. Since K is an integral domain, this is 0 only if both a and b are 0, which would violate the assumption that ax^n and bx^m are the leading terms of their respective polynomials.

9.5 Divisibility An element b of an integral domain is divisible by all its associates and all invertible elements (the associates of 1). together, these are called the *improper divisors* of b. If $b \neq 0$ has no proper divisors and not invertible, then it is said to be *irreducible* (equivalent to being *prime*).

In a field F, every non-zero element is invertible and there are no irreducibles.

In the ring $\mathbb{Z}$, the only invertible are ± 1. Two integers m and n are associates iff $m = \pm n$. An irreducible is a non-zero element p, other than ± 1, whose only divisors are ± 1 and $\pm p$ (the improper divisors).

In the ring $K[x]$ of polynomials over an integral domain K, a product fg can be only 1 only if both the polynomials are constants and invertible in K. Two polynomials f and g are associates iff $g = cf$ for an invertible constant c.

In the ring $F[x]$ of a polynomials over a field F, the invertible elements are non-zero constant polynomials. Two polynomials f and g are associates iff $g = cf$ for a non-zero constant c. Every linear polynomial is irreducible (prime).

9.6 Ideals and divisibility The *principal ideal* generated by a, denoted (a), consists of all the multiples of a. If $a \mid b$ then the multiples of a include all the multiples of b, and, so, $(a) \supseteq (b)$. The converse also holds: $a \mid b \iff (a) \supseteq (b)$. If a and b are associates then $(a) = (b)$, and conversely.

The ideal generated by a set of elements $a_1, \ldots, a_n$, denoted $(a_1, \ldots, a_n)$, is the set of "linear combinations" $k_1a_1 + \ldots + k_na_n$ with coefficients $k_i \in K$.

9.7 Prime elements In any commutative ring K, a *prime element* is a non-zero, non-invertible element p with the property:

$$p \mid ab \implies p \mid a \ \vee \ p \mid b$$

The definition is motivated by the *prime factorization property*, which requires that every element in a ring can be written uniquely as the product of primes that divide it. The rings that satisfy the property are called *unique factorization domains*.

In an integral domain, the prime elements are precisely the irreducible elements.

9.8 Prime ideals An immediate generalization of the notion of primes to ideals gives the notion of *prime ideals*. An ideal $P \subseteq K$ is prime iff:

$$ab \in P \implies a \in P \ \vee \ b \in P$$

If $P = (p)$ is a principal ideal, this says precisely that p is a prime element. But, the generalization works for non-principal ideals as well.

9.9 Maximal ideals In any commutative ring K, an ideal $I \subseteq K$ is said to be *maximal* if there are no ideals properly between I and K, *i.e.*, $I \subseteq J \subseteq K \implies J = I \vee J = K$.

9.10 Theorem *For any ideal I in a nontrivial commutative ring K,*

- *I is maximal $\iff$ K/I is a field,*
- *I is prime $\iff$ K/I is an integral domain.*

By Theorem 2.37, a nontrivial commutative ring is a field iff it has no proper ideals. But the ideals in the quotient ring K/I correspond to ideals in K between I and K. If I is maximal, there are no proper ideals between I and K and, so, K/I must be a field.

For the second statement, note that the elements of K/I are cosets of the form $I + a$ for elements $a \in K$. The condition (9.2) for integral domains can be expressed as

$$(I + a)(I + b) = I \implies I + a = I \ \vee \ I + b = I$$

$I + a = I$ is equivalent to $a \in I$. The term $(I + a)(I + b)$ expands to $I^2 + aI + Ib + ab$. Since I is closed under multiplication by elements of K, $I^2 = I$, $aI \subseteq I$, and $Ib \subseteq I$. So, the left hand side amounts to $ab \in I$. Thus, the above implication is equivalent to the prime-ness of I:

$$ab \in I \implies a \in I \ \vee \ b \in I$$

An easy corollary: *every maximal ideal in a commutative ring is prime.*

This follows from the fact fields are integral domains.

9.11 Ideals and homomorphisms of fields *A field has no proper ideals.* Suppose $S \subseteq F$ is an ideal with a nonzero element $x \in S$. Any other element $y \in F$ can be written as a multiple of x: $y = \frac{y}{x}x$. Hence, unless $S = \{0\}$, $S = F$.

Conversely, a nontrivial commutative ring K with no proper ideals is necessarily a field. Suppose $a \neq 0$ is an element of K. The ideal Ka of all multiples of a must be the whole of K. In particular, $1 \in Ka$. So, K contains the inverse of every nonzero a, making it a field.

Every homomorphism of fields is a monomorphism. If $h : F \to F'$ is a homomorphism of fields then the kernel of h, as an ideal in F, must be improper. Note that $h(0) = 0'$ and $h(1) = 1'$ and $0' \neq 1'$ because F' must be nontrivial for it to be a field. If $h^{-1}(0') = F$ then it would mean $h(1) = 0'$, forcing $0' = 1'$. Hence the kernel of h must be $\{0\}$, implying that h is a monomorphism (an injective homomorphism).

9.12 Logical relations of fields A *logical relation* of fields $R : F \leftrightarrow F'$ is a relation that preserves all the operations of fields. We treat inverse as a partial operation of type $F \rightharpoonup F$ with the relation action $R \rightharpoonup R$ given by

$$f \left[R \rightharpoonup R\right] f' \iff \forall x, x'.\ x \left[R\right] x' \implies (f(x) = \emptyset \wedge f'(x') = \emptyset) \vee (f(x) \left[R\right] f'(x'))$$

Since x^{-1} is undefined just for 0, this means that a logical relation of fields can relate 0 to *only* 0, *i.e.*,

$$x \left[R\right] x' \implies (x = 0 \iff x' = 0)$$

Consequently, R is necessarily *single-valued*. If $x \left[R\right] y_1$ and $x \left[R\right] y_2$ then $0 \left[R\right] y_1 - y_2$, which implies $y_1 = y_2$. Similar argument shows that R is single-valued in the reverse direction. Thus a logical relation of fields is always a *partial isomorphism*, i.e., there are subfields $F_0 \subseteq F$ and $F'_0 \subseteq F'$ such that R is an isomorphism between F_0 and F'_0.

Since the graphs of homomorphisms are logical relations, this fact also gives an abstract reason for why homomorphisms of fields are monomorphisms. If $R = \langle h \rangle$ is the graph of a homomorphism, then it is single-valued and total and, hence, injective.

Note that this argument does not apply to vector spaces. Even though a logical relation of vector spaces must necessarily relate 0 to 0, it can also relate non-zero vectors to 0.

9.13 Order and characteristic The *order* of an element a in a commutative ring is its order in the additive group, i.e., the least integer n such that $n \cdot a = 0$ if such n exists and ∞ otherwise. The *characteristic* of the ring is the order of the unit 1. So, if the characteristic is a finite integer n then $n \cdot 1 = 0$ and similarly $kn \cdot 1 = 0$ for any positive integer k. If the characteristic is ∞ then all integer multiples of 1 are non-zero.

A finite characteristic n of an integral domain is necessarily a prime number. If it were not prime, then it would be expressible as an integer product $n = m_1 m_2$. If $(m_1 m_2) \cdot 1 = 0$

9.14 Field of fractions If K is an integral domain, we can construct a field from the "fractions" $\frac{a}{b}$ of elements a and non-zero elements b of K. Formally, let K^* denote the set of non-zero elements of K. Then a "fraction" is an equivalence class of pairs in $K \times K^*$ under the equivalence relation:

$$(a_1, b_1) \sim (a_2, b_2) \iff a_1 b_2 = a_2 b_1$$

We write $\frac{a}{b}$ for the equivalence class of $(a, b) \in K \times K^*$. These fractions have the familiar operations of addition and multiplication, forming a field $Q(K)$. Note that $\mathbb{Q} = Q(\mathbb{Z})$.

If F is a field, the field of fractions of the polynomial ring $F[x_1, \ldots, x_n]$ is denoted $F(x_1, \ldots, x_n)$. Its elements are called the *rational expressions* over F and, regarded as functions of $x_1, \ldots, x_n$, the *rational functions* over F. If $F = Q(K)$ is the field of fractions over an integral domain, the rational expressions of $F(x_1, \ldots, x_n)$ can be written as quotients of polynomials with coefficients from K rather than F [Escofier, 2001].

Field extensions

9.15 Field extensions Whenever $F \subseteq L$ is a subfield of a larger field L, we call L an *extension* of F, and use the notation $L \,|\, F$ to talk about the *extension* as a concept. For example, $\mathbb{R}$ is an extension of $\mathbb{Q}$.

If $\alpha_1, \ldots, \alpha_n \in L$ are elements of the larger field, we can construct a field $F(\alpha_1, \ldots, \alpha_n)_L$ which is the field *generated* by F and $\alpha_1, \ldots, \alpha_n$, i.e., it is the intersection of all the subfields of L that have F as a subfield and $\alpha_1, \ldots, \alpha_n$ among their elements. For example $\mathbb{Q}(\sqrt{2})$ is a subfield of $\mathbb{R}$ whose elements can be written as $a + b\sqrt{2}$ for rational numbers a and b.[23]

For the same data, $F[\alpha_1, \ldots, \alpha_n]_L$ denotes the *commutative ring* generated by F and $\alpha_1, \ldots, \alpha_n$, i.e., it is the intersection of all the subrings of L that have F as a subring and have $\alpha_1, \ldots, \alpha_n$ among their elements. The difference between $F(\alpha_1, \ldots, \alpha_n)_L$ and $F[\alpha_1, \ldots, \alpha_n]_L$ is that the latter is only closed under (addition and) multiplication whereas the former is also closed under division.

A field extension $M \,|\, F$ is said to be *simple* if $M = F(\alpha)_L$ for single element $\alpha \in L$. The extensions $\mathbb{Q}(\sqrt{2})_{\mathbb{R}}$ and $\mathbb{R}(i)_{\mathbb{C}}$ are examples of simple extensions. The extension $\mathbb{Q}(\sqrt{2}, \sqrt{3})_{\mathbb{R}}$ is also simple because it can be reduced to $\mathbb{Q}(\sqrt{2} + \sqrt{3})_{\mathbb{R}}$.

9.16 Homomorphisms and automorphisms of field extensions A *homomorphism* of field extensions $h : (L_1 \,|\, F) \to (L_2 \,|\, F)$ is a field homomorphism $h : L_1 \to L_2$ that is identity on the common subfield F. We refer to it as an *F-homomorphism* or a homomorphism that "keeps F *fixed*." Similarly, *F-isomorphisms* are invertible F-homomorphisms and *F-automorphisms* are F-isomorphisms from a field extension to itself.

Since homomorphisms of fields are always injective (§9.11), F-homomorphisms are also injective. Moreover, F-homomorphisms preserve the roots of polynomials in $F[x]$. Consider a polynomial $f(x) = a_0x^n + \cdots + a_n$ in $F[x]$. Note that the same formula for $f(x)$ gives polynomials in the extensions L_1 and L_2. Denote these copies of the polynomial $f(x)$ as $f_1(x)$ and $f_2(x)$ respectively. If $h : (L_1 \,|\, F) \to (L_2 \,|\, F)$ is an F-homomorphism, then $h(f_1(x)) = f_2(h(x))$. If f_1 has a root $\alpha \in L_1$ then $f_1(\alpha) = 0$ and $f_2(h(\alpha)) = h(f_1(\alpha)) = h(0) = 0$. Thus, $h(\alpha)$ is a root of f_2 in L_2.

The collection of F-automorphisms of L is denoted $\mathrm{Aut}(L \,|\, F)$. Note that it forms a *group*. By the argument of the previous paragraph, all automorphisms of $\mathrm{Aut}(L \,|\, F)$ *carry the roots of a polynomial to other roots of the same polynomial*. In other words, they represent permutations of the roots of polynomials. For this reason, the automorphism group $\mathrm{Aut}(L \,|\, F)$ plays an important role in Galois theory.

9.17 Degree of field extensions If $L \,|\, F$ is a field extension then L can be treated as (left or right) module of F (i.e., an F-vector space). The *degree* of the field extension, denoted

[23] How does $F(\alpha_1, \ldots, \alpha_n)_L$ relate to $F(\alpha_1, \ldots, \alpha_n)$?

$[L:F]$, is the dimension of the vector space ${}_FL$. For example, the field of complex numbers $\mathbb{C} = \mathbb{R}(i)$ is of degree 2.

If $M \mid L \mid F$ are successive field extensions, then

$$[M:F] = [M:L] \cdot [L:F]$$

To see this, let $\{u_1, \ldots, u_n\}$ be a basis of ${}_FL$, and $\{v_1, \ldots, v_m\}$ be a basis of ${}_LM$. Then each element $c \in {}_LM$ is a linear combination $\sum_j b_j v_j$ with coefficients $b_j \in L$. Each b_j is a linear combination $\sum_i a_{ji} u_i$ with coefficients a_{ji} in F. So, c is a linear combination $\sum_{i,j} a_{ji} u_i v_j$. In other words, $\{u_i v_j\}_{i,j}$ is a basis for ${}_FM$. The dimension of the vector space ${}_FM$ is thus mn.

9.18 Algebraic elements and minimal polynomials An element $\alpha \in L$ is said to be *algebraic* over F when there is a nonzero polynomial $f \in F[x]$ with α is its root, i.e., $f(\alpha) = 0$. Examples over $\mathbb{Q}$ include all the roots of rational numbers. An nth root $\sqrt[n]{u}$ is the root of polynomial $x^n - u$. The complex number i is the root of $x^2 + 1$. A complex cube root ω of 1 is the root of $x^2 + x + 1$.

Elements $\alpha \in L$ that are not algebraic are said to be *transcendental*. They satisfy $f(\alpha) = 0 \implies \alpha = 0$ for all polynomials $f \in F[x]$. Examples include e and π over $\mathbb{Q}$.

Among all the polynomials that have α as their root, there exists a *minimal polynomial* m_α. It can be described directly as the monic[24] polynomial h of *least* degree for which $h(\alpha) = 0$. To see that such a polynomial exists, note that the set D of polynomials $f \in F[x]$ with root α is an ideal in $F[x]$. It is indeed a prime ideal because $f(\alpha)g(\alpha) = 0$ implies $f(\alpha) = 0$ or $g(\alpha) = 0$. Like any ideal in $F[x]$, D is a *principal ideal*.[25] The polynomial h that generates $D = (h)$ is the minimal polynomial m_α of α over F.

For a more abstract view of this phenomenon, consider the evaluation map $F[x] \to L$ that sends each polynomial f to $f(\alpha)$. Since L is a field, $F[x]$ is a principal ideal domain and the kernel of this map is of the form (h) for a unique polynomial h, which is then the *minimal polynomial* of α [Robalo, 2009].

The *degree* of an algebraic element α is defined to be the degree of the minimal polynomial m_α. For instance, $\sqrt{3}$ is of degree 2. The transcendental elements have no degree.

9.19 Algebraic extensions A field extension $L \mid F$ is *algebraic* if every element of L is algebraic over F. Otherwise, it is a *transcendental extension*.

Every finite-dimensional field extension is algebraic. If $[L:F] = n$ is finite then at most n elements of L can be linearly independent. So, for any $\alpha \in L$, the set of $\{1, \alpha, \ldots, \alpha^n\}$ of $n+1$ elements is linearly dependent and there exist linear combinations $\sum_{i=0}^n a_i \alpha^i$ that evaluate to 0. That means that α is the root of the polynomial $\sum_{i=0}^n a_i x^i$.

If $\alpha \in L$ is an algebraic element over F, then the extension $F(\alpha)_L \mid F$ is called a *simple algebraic extension*.

9.20 Theorem *If α is an algebraic element over F, then the minimal polynomial m_α of α over F is irreducible in $F[x]$ and divides every polynomial in $F[x]$ with α is its root.*

[24] "Monic" polynomial in this context means one with its leading coefficient 1. "Normalized" would be a better term for it.

[25] Needs elaboration.

9.21 Theorem *If $F(\alpha)_L \,|\, F$ is a simple algebraic extension then every element $u \in F(\alpha)_L$ can be expressed as $u = Q(\alpha)$ for a unique polynomial $Q \in F[x]$. Further, the degree of Q is less than the degree of α.*

9.22 Theorem *The degree of a simple algebraic extension $[F(\alpha)_L : F]$ is precisely the degree of α, i.e., the degree of the minimal polynomial m_α. If $F(\alpha)_L \,|\, F$ is a transcendental extension then the degree of the extension is ∞.*

Proof: Suppose the minimal polynomial m_α is of degree n. Consider the n elements $1, \alpha, \ldots, \alpha^{n-1}$. By the previous theorem, every element $u \in F(\alpha)_L$ can be expressed as $Q(\alpha)$ for a unique polynomial Q of degree less than n, i.e., as a linear combination of $1, \ldots, \alpha^{n-1}$. Hence, $\{1, \ldots, \alpha^{n-1}\}$ is a basis for $F(\alpha)_L$.

If α is transcendental over F then the elements $1, \alpha, \alpha^2, \ldots$ are linearly independent over F. So, the corresponding vector space is infinite-dimensional. □

Thus, every simple algebraic extension has a finite degree. For example, consider the field $\mathbb{Q}(\sqrt{2}+\sqrt{3})_{\mathbb{R}}$, which is a simple algebraic extension of $\mathbb{Q}$. The minimal polynomial of $\sqrt{2}+\sqrt{3}$ is

$$x^4 - 10x^2 + 1 = (x - (\sqrt{2} + \sqrt{3}))(x - (-\sqrt{2} + \sqrt{3}))(x - (\sqrt{2} - \sqrt{3}))(x - (-\sqrt{2} - \sqrt{3}))$$

which is of degree 4. The field $\mathbb{Q}(\sqrt{2} + \sqrt{3})_{\mathbb{R}}$, regarded as a $\mathbb{Q}$-vector space, has the basis $\{1, \sqrt{2}, \sqrt{3}, \sqrt{6}\}$ and, so, has degree 4 as well.

Conversely, *if $L \,|\, F$ has finite degree then L can be expressed as $L = F(\alpha_1, \ldots, \alpha_n)_L$ for finitely many algebraic elements $\alpha_1, \ldots, \alpha_n \in L$.* Thus, all the elements of L are algebraic (over F).

9.23 Automorphism groups of algebraic extensions Recall, from §9.16, that if $L \,|\, F$ is a field extension then the collection of all F-automorphisms forms a group denoted $\mathrm{Aut}(L \,|\, F)$.

If $L \,|\, F$ is a finite algebraic extension, then the automorphism group $\mathrm{Aut}(L \,|\, F)$ is finite. Moreover, it is bounded by [Cox, 2004, Cor. 6.1.5]:

$$|\mathrm{Aut}(L \,|\, F)| \leq \deg(\alpha_i) \cdots \deg(\alpha_i)$$

Suppose $L = F(\alpha_1, \ldots, \alpha_n)$ is a finite algebraic extension of F. Then any $u \in L$ can be written as $u = h(\alpha_1, \ldots, \alpha_n)$ for some polynomial $h \in F[x_1, \ldots, x_n]$. An automorphism $\sigma \in \mathrm{Aut}(L \,|\, F)$ must preserve addition and multiplication, and, therefore, the values of all polynomials. So, we have $\sigma(u) = \sigma(h(\alpha_1, \ldots, \alpha_n)) = h(\sigma(\alpha_1), \ldots, \sigma(\alpha_n))$. Since this is the case for all $u \in L$, *σ is uniquely determined by its action on $\alpha_1, \ldots, \alpha_n$.*

Each α_i is algebraic over F, *i.e.*, a root of the minimal polynomial $m_{\alpha_i} \in F[x]$. So, there are at most $\deg(\alpha_i)$ possibilities for $\sigma(\alpha_i)$. Hence, the possibilities for σ are bounded: $|\mathrm{Aut}(L \,|\, F)| \leq \deg(\alpha_i) \cdots \deg(\alpha_i)$.

9.24 Galois extension and Galois group A field extension $L \,|\, F$ is called a *Galois extension* if the fixed points of $\mathrm{Aut}(L \,|\, F)$ are precisely the elements of F:

$$\mathrm{Fix}(\mathrm{Aut}(L \,|\, F)) = F$$

In other words, every element of L not in F is moved by some automorphism in $\mathrm{Aut}(L \,|\, F)$. When $L \,|\, F$ is a Galois extension, $\mathrm{Aut}(L \,|\, F)$ is also called the *Galois group* of the extension and denoted $\mathrm{Gal}(L \,|\, F)$. Alternative notations are $\mathrm{Gal}(L/F)$, $\Gamma(L/F)$ and $\Gamma(L : F)$.

9.25 Example A simple example of a Galois group is $\mathrm{Aut}(\mathbb{C}\,|\,\mathbb{R})$. Note that $\mathbb{C} = \mathbb{R}(i)$. If σ is an $\mathbb{R}$-automorphism of $\mathbb{C}$ then $\sigma(i)$ must be some complex number j. Since σ preserves multiplication, we obtain $\sigma(i^2) = j^2$. But, $i^2 = -1 \in \mathbb{R}$ and σ keeps the elements of $\mathbb{R}$ fixed. Therefore, $\sigma(-1) = -1 = j^2$, *i.e.*, $j \in \{i, -i\}$. Thus, σ is either the identity transformation $\sigma : x + iy \mapsto x + iy$ or the complex conjugation $\sigma : x + iy \mapsto x - iy$. The automorphism group $\mathrm{Aut}(\mathbb{C}\,|\,\mathbb{R})$ is the cyclic group of order 2, a remarkably simple group given the structure of the fields that it is dealing with.

The extension $\mathbb{C}\,|\,\mathbb{R}$ is Galois because the complex conjugation automorphism leaves fixed precisely the elements of $\mathbb{R}$. Note also that the order of the Galois group is the same as the order of the field extension $[\mathbb{C}:\mathbb{R}]$. This is an instance of a general phenomenon. See Lemma 9.32.

9.26 Counterexample An example of a field extension that is not Galois is $L\,|\,\mathbb{Q}$ where $L = \mathbb{Q}(\sqrt[3]{2})$ [Cox, 2004, Examp. 6.1.6]. The minimal polynomial of $\sqrt[3]{2}$ is $x^3 - 2$, which has roots $\sqrt[3]{2}$, $\omega\sqrt[3]{2}$ and $\omega^2\sqrt[3]{2}$. Any automorphism of $L\,|\,\mathbb{Q}$ can only map $\sqrt[3]{2}$ to one of the other two, but they do not lie in L. Hence, the only $\mathbb{Q}$-automorphism is the identity, *i.e.*, $\mathrm{Aut}(L\,|\,\mathbb{Q}) = \{1\}$, and $\mathrm{Fix}(\mathrm{Aut}(L\,|\,F)) = L \neq F$. Thus, the extension is not Galois.

Characterization of Galois extensions

Galois groups and Galois extensions play a key role in determining which polynomials are solvable. We first identify a more elementary characterization of Galois extensions in the following paragraphs.

9.27 Separable extension A polynomial $f \in F[x]$ is *separable* if all its roots are distinct (in some algebraic extension of F). For example $x^2 + 1 = (x + i)(x - i)$ is separable, while $x^2 + 2x + 1 = (x + 1)^2$ is not.

An element α of an algebraic extension L of F is *separable* if its minimal polynomial over F is separable. The extension itself is said to be *separable* if all its elements are separable.

Separability is automatic in fields of characteristic ∞, because a well-known criterion implies that an irreducible polynomial has multiple roots if and only if its derivative is zero. However, the derivative can be zero for irreducible polynomials in prime fields, e.g., for $x^p - a$.

9.28 Normal extension An extension $L\,|\,F$ is called a *normal extension* if every irreducible polynomial $f \in F[x]$ that has a root in L splits into linear factors in L. In other words, an irreducible polynomial $f \in F[x]$ either *has no roots in L or has all roots in L.*

If f has a root in L, say α, then the minimal polynomial m_α divides f. So, an extension $L\,|\,F$ is a normal extension if and only if the minimal polynomials of all $\alpha \in L$ split in L.

9.29 Lemma *A Galois extension $L\,|\,F$ is separable and normal (i.e., the minimal polynomial m_α splits into linear factors in L).*

Let $\alpha \in L$ be an element. Consider the stabilizer of α in $\Gamma = \mathrm{Aut}(L\,|\,F)$. For each σ in the stabilizer Γ_α, we have $\sigma(\alpha) = \alpha$. Consider the polynomial $f = \prod_\sigma (x - \sigma(\alpha))$ where σ ranges over a system of coset representatives of the stabilizer Γ_α. This product is *finite* because every $\sigma(\alpha)$ must be a root of the minimal polynomial m_α, which is of finite degree.

In fact, f must be the same as m_α by the following argument. Both f and m_α are in $F[x]$ and have α as their root. Hence each $\sigma(\alpha)$ must be a root of both. Thus f divides m_α. However, m_α is irreducible. So, $f = m_\alpha$.

Finally, f has no multiple roots by construction. Hence α is separable over F. ■

9.30 Splitting field

9.31 Lemma *A finite extension $L \mid F$ is Galois if and only if it is the splitting field of an irreducible separable polynomial $f \in F[x]$.*

9.32 Lemma *The order of the Galois group of a Galois extension is the same as the degree of the extension: $|\mathrm{Aut}(L \mid F)| = [L : F]$. Conversely, if $|\mathrm{Aut}(L \mid F)| = [L : F]$ then $L \mid F$ is a Galois extension.*

If $L \mid F$ is a Galois extension, it is the splitting field of an irreducible separable polynomial $f \in F[x]$ by Lemma 9.31. If f is of degree n then it has roots $\alpha_1, \ldots, \alpha_n$ and $L = F(\alpha_1, \ldots, \alpha_n)$ and each α_i is separable over F.[26]

By the theorem of the primitive element **??**, we can find $\beta \in L$, separable over F, such that $L = F(\beta)$. The degree $[L : F]$ is then nothing but the degree of β, say k.

Since L is the splitting field of m_β, it is a normal extension of F. So, all of the roots of m_β must lie in L (as β is already in L). Let the roots be $\beta_1, \ldots, \beta_k$. Then

$$m_\beta(x) = (x - \beta_1) \cdots (x - \beta_k)$$

$\mathrm{Aut}(L \mid F)$ has an automorphism sending β to each of the roots $\beta_1, \ldots, \beta_k$. ...

9.33 Intermediate fields If $L \mid F$ is a field extension then a field M such that $F \subseteq M \subseteq L$ is called an *intermediate field* of the extension $L \mid F$ and we write it as a "tower" of extensions $L \mid M \mid F$. A group $M^{*L} = \mathrm{Gal}(L \mid M)$ of all M-automorphisms of L can be associated with each intermediate field M. These are the automorphisms of L that keep M fixed. Since M-automorphisms of L are also F-automorphisms, we may note that $M^{*L} \subseteq F^{*L}$. Using this notation, F^{*L} is the entire Galois group of the field extension and $L^{*L} = \mathbf{0}$ with just the identity automorphism. So, we obtain a subgroup hierarchy $F^{*L} \supseteq M^{*L} \supseteq L^{*L}$.

Since the subgroups of finite groups can be easily enumerated, this gives a powerful device for reasoning about intermediate fields. For example, consider the field extension $\mathbb{C} \mid \mathbb{R}$. Since its Galois group is the cyclic group of order 2, which does not have any nontrivial subgroups, we immediately obtain the fact that there are no nontrivial intermediate fields $\mathbb{R} \subseteq M \subseteq \mathbb{C}$.

If M is a subfield of another intermediate field N, i.e., $F \subseteq M \subseteq N \subseteq L$, then the group $N^{*L} = \mathrm{Gal}(L \mid N)$ is a subgroup of M^{*L}, i.e., $N^{*L} \subseteq M^{*L}$. Thus, the intermediate fields of $L \mid F$ give rise to subgroups of $\mathrm{Gal}(L \mid F)$ with the partial order inverted.

Conversely, each subgroup $S \subseteq \mathrm{Gal}(L \mid F)$ has an associated set $S^\dagger = \mathrm{Fix}(S) \subseteq L$ of elements that are kept fixed by the automorphisms in S. This set can be shown to be a subfield of L containing F, i.e., $F \subseteq S^\dagger \subseteq L$. We call $S^\dagger$ the *fixed field* of S. If $R \subseteq S$ then $R^\dagger \supseteq S^\dagger$, because any element fixed under the automorphisms of S is also fixed under the automorphisms of R.

[26] Why?

(To see that $S^\dagger$ is a field, note that each $\sigma \in S$ is a field automorphism of L. So, if $\sigma(x) = x$ and $\sigma(y) = y$ then $\sigma(x + y) = x + y$, $\sigma(0) = 0$, $\sigma(xy) = xy$ and $\sigma(1) = 1$, showing that $S^\dagger$ is closed under all field operations.)

Under certain extra hypotheses discussed below, there is a one-to-one correspondence between the subgroups of $\mathrm{Gal}(L \,|\, F)$ and the intermediate fields of the extension $L \,|\, F$. This is the "fundamental theorem" of Galois theory.

9.34 Galois extensions and intermediate fields If $L \,|\, F$ is a Galois extension and M is an intermediate field $F \subseteq M \subseteq L$ then:

- The extension $L \,|\, M$ is also a Galois extension.
- The extension $M \,|\, F$ is a Galois extension if and only if $\mathrm{Gal}(L \,|\, M)$ is a *normal* subgroup of $\mathrm{Gal}(L \,|\, F)$. In that case, $\mathrm{Gal}(M \,|\, F) = \mathrm{Gal}(L \,|\, F)/\mathrm{Gal}(L \,|\, M)$ is the quotient group.

Note that $M \subseteq \mathrm{Fix}(\mathrm{Gal}(L \,|\, M))$. It is in fact an equality because $L \,|\, M$ is a Galois extension. For subgroups $S \subseteq \mathrm{Gal}(L \,|\, M)$, $S^\dagger = \mathrm{Fix}(S)$ is an intermediate field $F \subseteq S^\dagger \subseteq L$ and we have $S \subseteq \mathrm{Gal}(L \,|\, S^\dagger)$. Thus, we have a contravariant Galois connection (adjunction)

$$\{M \mid F \subseteq M \subseteq L\}^{\mathrm{op}} \dashv \{S \mid S \subseteq \mathrm{Gal}(L \,|\, M)\}$$

Polynomials and Galois theory

This section is based on [Stewart, 2004] and [Edwards, 1984].

All rings in this section will be commutative, with letter K standing for such. Fields are denoted F. Other letters $L, M, \ldots$ stand for both.

9.35 The idea behind Galois theory The Galois groups of field extensions give us insight into how to find roots of polynomials. Consider a polynomial:

$$f(x) = x^4 - 4x^2 - 5$$

which factorizes as

$$f(x) = (x^2 + 1)(x^2 - 5)$$

and has four roots: $\alpha = i$, $\beta = -i$, $\gamma = \sqrt{5}$ and $\delta = -\sqrt{5}$. It is not possible to distinguish α from β or γ from δ by "algebraic means." That is, given any polynomial equation with rational coefficients from the four roots, with examples such as

$$\alpha^2 + 1 = 0 \quad \alpha + \beta = 0 \quad \delta^2 - 5 = 0 \quad \gamma + \delta = 0 \quad \alpha\gamma - \beta\delta = 0$$

exchanging α with β or γ with δ leaves the *validity* of the equation unchanged, i.e., a valid equation remains valid and invalid equation remains invalid. The permutations that have this property form a *group*. They include the identity permutation and are closed under composition and inverses. This is called the *Galois group* of the polynomial. (This is closely related to Galois groups of field extensions, as discussed below.) It is a particular subgroup of the symmetric group S_4 of all permutations of the four roots.

Let R stand for the permutation exchanging α and β, and S for the permutation exchanging γ and δ. The Galois group A consists of the permutations $\{I, R, S, RS, SR\}$. Consider also the subgroup $B = \{I, R\} \subseteq A$, which only permutes α and β while keeping the other roots fixed.

All the polynomial expressions in $\alpha, \beta, \gamma, \delta$ that are *symmetric* in α and β are fixed by the permutation group B. All such symmetric expressions can be expressed as polynomials in $\alpha + \beta$, $\alpha\beta$, γ and δ. For example, $\alpha^2 + \beta^2 - 5\gamma\delta^2$ can be expressed as $(\alpha + \beta)^2 - 2\alpha\beta - 5\gamma\delta^2$.

Consider the three fields related to $\alpha, \beta, \gamma, \delta$, namely:

$$\mathbb{Q} \subseteq \mathbb{Q}(\gamma, \delta) \subseteq \mathbb{Q}(\alpha, \beta, \gamma, \delta)$$

where $\mathbb{Q}(\gamma, \delta)$ is the field of rational expressions in γ and δ and similarly for $\mathbb{Q}(\alpha, \beta, \gamma, \delta)$. (These three fields can also be regarded as subfields of $\mathbb{C}$ but that is another matter.) Two observations follow: The rational expressions of $\mathbb{Q}(\alpha, \beta, \gamma, \delta)$ fixed by the Galois group A are precisely those of $\mathbb{Q}$. The rational expressions fixed by the subgroup B are precisely those of $\mathbb{Q}(\gamma, \delta)$. Thus, the above hierarchy of subfields of $\mathbb{C}$ is related an inverted hierarchy of subgroups of the Galois group:

$$A \quad \supseteq \quad B \quad \supseteq \quad \mathbf{0}$$

We can use this fact to work out how to solve the quartic equation $f(x) = 0$ as follows.

The expressions $\alpha + \beta$ and $\alpha\beta$ are obviously both fixed by B and they lie in $\mathbb{Q}(\gamma, \delta)$. Note that α and β satisfy a quadratic equation whose coefficients are in $\mathbb{Q}(\gamma, \delta)$:

$$(x - \alpha)(x - \beta) = x^2 - (\alpha + \beta)x + \alpha\beta$$

Thus, we can use the formula for solving a quadratic equation to express α and β as radical expressions of γ and δ.

We can repeat this trick to find γ and δ. The expressions $\gamma + \delta$ and $\gamma\delta$ are fixed by the whole of A. Therefore, they satisfy a quadratic equation over $\mathbb{Q}$ and so can be expressed by radical expressions in rational numbers.

Thus, we found a method for solving the polynomial $f(x)$ using facts about symmetries. The fact that the Galois group A has a normal subgroup B allows us to decompose the problem of the quartic equation into two quadratic equations. If the Galois group does not have normal subgroups then such a decomposition is impossible.

9.36 Symmetric polynomials Let $K[r_1, \ldots, r_n]$ by the polynomial ring with n indeterminates. If $f : K \to L$ is a ring homomorphism and $h : \{1, \ldots, n\} \to |L|$ is a selection of elements of L then, by the universal property of polynomials, there is a unique homomorphism $\phi : K[r_1, \ldots, r_n] \to L$ such that

$$\phi(a) = f(a) \qquad \phi(r_i) = h(i)$$

for all $a \in K$ and for all $i = 1, n$.

If $\sigma \in S_n$ is a permutation, we can use the injection $K \hookrightarrow K[r_1, \ldots, r_n]$ and the assignment $i \mapsto r_{\sigma(i)}$ to obtain a unique homomorphism $\phi_\sigma : K[r_1, \ldots, r_n] \to K[r_1, \ldots, r_n]$, which has the property $\phi_\sigma(r_i) = r_{\sigma(i)}$. In fact, for any polynomial $P(r_1, \ldots, r_n)$, we can determine that $\phi_\sigma(P(r_1, \ldots, r_n) = P(r_{\sigma(1)}, \ldots, r_{\sigma(n)})$.

If K is an integral domain with a field of fractions F, then ϕ_σ also extends to the field $F(r_1, \ldots, r_n)$ of rational expressions in $r_1, \ldots, r_n$. Recall that an element of this field can be represented as a quotient of two polynomials P/Q, with both P and Q in $K[r_1, \ldots, r_n]$. Therefore, $\phi_\sigma(P/Q) = \phi_\sigma(P)/\phi_\sigma(Q)$.

A polynomial P in $K[r_1, \ldots, r_n]$ is said to be *symmetric* if it is fixed by ϕ_σ for every $\sigma \in S_n$, i.e., $\phi_\sigma(P) = P$. If K is an integral domain with a field of fractions F then a rational function P/Q is said to be *symmetric* if it is fixed by ϕ_σ, i.e., $\phi_\sigma(P/Q) = P/Q$. For example, the following polynomials are symmetric in $K[r_1, r_2, r_3]$:

$$\begin{array}{l} r_1 + r_2 + r_3 \\ r_1 r_2 r_3 \\ r_1^3 r_2 + r_2^3 r_3 + r_3^3 r_1 + r_2^3 r_1 + r_3^3 r_2 + r_1^3 r_3 \end{array}$$

9.37 Elementary symmetric polynomials Given a field F, consider the equation for the so-called *generic polynomial* of degree 3:

$$x^3 + bx^2 + cx + d = (x - r)(x - s)(x - t) \tag{9.4}$$

By multiplying the right hand side and equating the terms with like powers we obtain the equations:

$$\begin{array}{rcl} r + s + t & = & -b \\ rs + st + tr & = & c \\ rst & = & -d \end{array}$$

Note that all the polynomials on the left hand side are symmetric. They are called the *elementary symmetric polynomials* in 3 variables. More generally, for an equation for the generic polynomial in nth degree:

$$x^n + b_1 x^{n-1} + b_2 x^{n-2} + \cdots + b_n = (x - r_1) \cdots (x - r_n) \tag{9.5}$$

the coefficients $b_1, \ldots, b_n$ equate to the elementary symmetric polynomials in n variables. Write these polynomials as $\psi_1, \ldots, \psi_n$, each in the n indeterminates $r_1, \ldots, r_n$. Then all other symmetric polynomials in the n indeterminates can be expressed in terms of the elementary symmetric polynomials. For example, referring to (9.4), we have

$$\begin{aligned} r^3 + s^3 + t^3 &= b^3 - 3bc + 3d \\ &= (r+s+t)^3 - 3(r+s+t)(rs+st+tr) + 3rst \end{aligned}$$

Thus, given a polynomial with coefficients $b_1, \ldots, b_n$, we are able to evaluate all symmetric polynomials of its roots *immediately*, without ever finding the roots themselves! Moreover, if all the b's are rational numbers, so will be the values of the symmetric polynomials, even if the roots themselves might be complex.

This technique dates back to Newton.

9.38 Fundamental theorem of symmetric polynomials *Every symmetric polynomial in $r_1, \ldots, r_n$ can be expressed as a polynomial in the elementary symmetric polynomials $\psi_1, \ldots, \psi_n$. Moreover, a symmetric polynomial with integer coefficients can be expressed as a polynomial in $\psi_1, \ldots, \psi_n$ with integer coefficients.*

The proof is by induction on the number of variables n.

In particular, if $P(x) \in F[x]$ is a polynomial with distinct roots $\alpha_1, \ldots, \alpha_n$ in some larger field L, then every symmetric polynomial $S(r_1, \ldots, r_n) \in F[r_1, \ldots, r_n]$ satisfies

$$S(\alpha_1, \ldots, \alpha_n) \in F$$

Thus, even if $\alpha_1, \ldots, \alpha_n$ are in the larger field L, the symmetric polynomials in them still have values within F. The reason is that S is expressible as a polynomial in terms of the elementary polynomials $\psi_1, \ldots, \psi_n$. The values of the elementary polynomials, when evaluated at the roots $\alpha_1, \ldots, \alpha_n$, are given by the coefficients of $P(x)$ which are all within the field F.

9.39 Unique representation of symmetric polynomials *A symmetric polynomial in $F[r_1, \ldots, r_n]$ can be expressed in terms of elementary symmetric polynomials $\psi_1, \ldots, \psi_n$ uniquely.*

Let $u_1, \ldots, u_n$ be new variables. Consider a ring homomorphism $\phi : F[u_1, \ldots, u_n] \to F[r_1, \ldots, r_n]$ that sends each u_i to the corresponding ψ_i. We denote the image of ϕ as $F[\psi_1, \ldots, \psi_n]$, which is now a subring of $F[r_1, \ldots, r_n]$. So, we can regard ϕ as homomorphism of type $F[u_1, \ldots, u_n] \to F[\psi_1, \ldots, \psi_n]$. To show that the map is injective, we show that its kernel is $\{0\}$.

9.40 Discriminant The discriminant in n indeterminates is the product $\prod_{i,j}(r_i - r_j)^2$ where i and j run over all pairs of distinct integers $1 \leq i < j \leq n$. Two indeterminates are equal if and only if the discriminant is 0.

Evidently, the discriminant is a symmetric polynomial in the indeterminates $r_1, \ldots, r_n$ and, so, can be expressed in terms of the elementary symmetric polynomials in these indeterminates.

For example, the discriminant in two indeterminates is $(r-s)^2 = r^2 - 2rs + s^2$. For a quadratic, we have $x^2 + bx + c = (x-r)(x-s) = x^2 - (r+s)x + rs$, giving the elementary symmetric polynomials:

$$\begin{aligned} r + s &= -b \\ rs &= c \end{aligned}$$

Thus the discriminant expressed in terms of coefficients is $b^2 - 4c$, which will be 0 if and only if the two roots are equal.

9.41 Solving the quadratic equation Given a quadratic equation $x^2 + bx + c = 0$, we proceed by naming the two roots r and s. Then we can derive:

$$\begin{aligned} r &= \tfrac{1}{2}((r+s)+(r-s)) \\ &= \tfrac{1}{2}((r+s)+\sqrt{(r-s)^2}) \end{aligned}$$

Now, $r+s$ and $(r-s)^2$ are symmetric polynomials whose values are known: $r+s=-b$ and $(r-s)^2 = b^2-4c$. Thus, we obtain $r = \frac{1}{2}(-b+\sqrt{b^2-4c})$. The other root s is obtained by using $-$ in place of $+$.

$$\begin{aligned} s &= \tfrac{1}{2}((r+s)-(r-s)) \\ &= \tfrac{1}{2}((r+s)-\sqrt{(r-s)^2}) \end{aligned}$$

Note that we do not know in advance which root is r and which is s.

9.42 Solving the cubic equation Given the cubic equation $x^3 + bx^2 + cx + d = 0$, we proceed by naming the three roots r, s and t.

(Recall that a number u has three cube roots: $\sqrt[3]{u}$, $\omega\sqrt[3]{u}$ and $\omega^2\sqrt[3]{u}$, where ω is the complex cube root of 1, viz., $\omega = (-1+\sqrt{-3})/2$.)

$$\begin{aligned} r &= \tfrac{1}{3}((r+s+t)+(r+\omega s+\omega^2 t)+(r+\omega^2 s+\omega t)) \\ &= \tfrac{1}{3}((r+s+t)+\sqrt[3]{(r+\omega s+\omega^2 t)^3}+\sqrt[3]{(r+\omega^2 s+\omega t)^3}) \end{aligned}$$

The polynomials $P = (r+\omega s+\omega^2 t)^3$ and $Q = (r+\omega^2 s+\omega t)^3$ are not symmetric. However, note that:

- The expressions PQ and $P+Q$ are invariant under the interchange of s and t.
- Each of P and Q is invariant under the cyclic permutation $r \to t \to s \to r$. The cyclic permutation carries P to $(t+\omega r+\omega^2 s)^3 = \omega(\omega^2 t + r + \omega s)^3 = P$ and, similarly, Q to Q.

Since any permutation of r, s, t can be obtained as a composition of interchange of s and t and the cyclic permutations, it follows that PQ and $P+Q$ are symmetric polynomials in r, s, t and their values are known. Then we have the "normal form problem" (quadratic equation) to find the values of P and Q. Once this is done, we can plug these values into the above definition to find r. The other solutions s and t are obtained by multiplying the two cube roots by ω and ω^2. (Note that there are three choices for each cube root in the above formula, giving a total of 9 combinations. However, only 3 choices among them give roots. The others are to be discarded!)

9.43 Lagrange resolvent For the cubic equation, the *Lagrange resolvent* is the quantity $T = r+\omega s+\omega^2 t$ where r, s, t are the solutions of the cubic. Lagrange observes that T has six values, depending on the order in which the roots are taken. The six values are the solutions of a 6th degree equation:

$$f(x) = (x-t_1)(x-t_2)(x-t_3)(x-t_4)(x-t_5)(x-t_6) = 0$$

which is called the resolvent equation. Its coefficients, being symmetric in the six values of T, are symmetric in r, s, t, and therefore known quantities expressible in the coefficients of the cubic. Although $f(x)$ is of a higher degree than the original polynomial, it is solvable because it is in fact a quadratic in x^3 and can be solved by solving a quadratic equation and then taking a cube root.

Lagrange resolvents can be constructed for quartic and quintic equations in a similar way. These resolvents have three crucial properties:

1. They are rationally expressible in terms of the roots of the equation and known quantities (including all rational numbers, the coefficients of the given equation and the roots of unity).

2. Conversely, each root of the equation can be expressed rationally in terms of the resolvent and known quantities.

3. The resolvent is the solution of a solvable equation.

However, the degree of the resolvent equation grows rapidly. For $n = 3$, the resolvent equation has degree $3! = 6$, but luckily it is expressible in terms of x^3 as a polynomial of degree $2! = 2$. So, it is solvable. For $n = 4$, the resolvent equation has degree $4! = 24$ but actually degree $3! = 6$ in x^4. With ingenuity, this particular equation of degree 6 can be solved. Alternatively, the resolvent $T = x_1 + \alpha x^2 + \alpha^2 x_3 + \alpha^3 x_4$ can be used with the non-primitive 4th root of unity, $\alpha = -1$, and the solution of the resolvent equation $f(x) = 0$ reduces to $f(x) = g(x)^4$, where $g(x)$ has degree 6. This equation turns out to be solvable because it is actually a cubic in x^2.

For $n = 5$, these tricks fail. The resolvent equation is of degree $5! = 120$. It can be expressed as $f(x) = g(x)^5$ in terms of a polynomial $g(x)$ of degree 24. There is no general way of solving the equation for $g(x)$. There are no non-primitive 5th roots of unity other than 1 itself.

9.44 Galois resolvent *This paragraph is taken from [Swallow, 2004].*

Abstract Galois theory

This section is based on [Borceux and Janelidze, 2001, Szamuely, 2009] .

9.45 Algebraic closure A field is said to be *algebraically closed* if it has no algebraic extensions other than itself. That means that it contains the roots of all polynomials over itself. The *algebraic closure* of a field F is an algebraic extension $\bar{F}$ that is algebraically closed.

The existence of an algebraic closure can only be proved by means of Zorn's lemma or some other equivalent form of the axiom of choice.

Every field F has an algebraic closure $\bar{F}$, and it is unique up to (non-unique) isomorphism.

If $L \,|\, F$ is an algebraic extension then there is an F-monomorphism $L \hookrightarrow \bar{F}$.

The embedding $L \hookrightarrow \bar{F}$ can be extended to an F-isomorphism $\bar{L} \to \bar{F}$.

For proof see [Lang, 2000, Chapter V] or [van der Waerden, 1949, §71].

9.46 Lemma *If $L \,|\, F$ is a finite extension of degree n, then L has at most n distinct F-algebra homomorphisms to $\bar{F}$. All homomorphisms are equal if and only if $L \,|\, F$ is separable.*

Grothendieck Galois theory

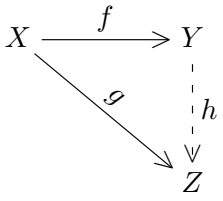
This section is based on [Dubuc and de la Vega, 2000] and [Dubuc, 2003]. See also [Tonini, 2009].

9.47 Strict epimorphism If $f : X \to Y$ is an arrow in a category $\mathcal{C}$, define the *kernel* of f (Ker_f) as the category of pairs of arrows $(x : C \to X, y : C \to X)$ such that $f \circ x = f \circ y$. The morphisms $h : (x, y) \to (x', y')$ are arrows $h : C \to C'$ such that $h; x' = x$ and $h; y' = y$.

- An arrow $g : X \to Z$ is said to be *compatible* with f if $\mathrm{Ker}_f \subseteq \mathrm{Ker}_g$, i.e.,

$$\forall x, y \in \mathrm{Hom}(C, X).\ f \circ x = f \circ y \implies g \circ x = g \circ y$$

- The arrow f is called a *strict epimorphism* if any compatible arrow $g : X \to Z$ uniquely factors through f, *i.e.*, there exists a unique $h : Y \to Z$ such that $g = h \circ f$.



(The definition is reportedly from Grothendieck's SGA4, Expose I, 10.2–3.) In essence, a strict epimorphism is the joint coequalizer of all parallel pairs of morphisms that it coequalizes. If Ker_f has a terminal object (called *the kernel pair* of f) then coequalizing the kernel pair is all that is required for a strict epimorphism. In that case, it is also called an *effective epimorphism.*

A strict epimorphism is always an epimorphism, because the latter only requires that for any $g : X \to Z$ there is at most one factor $h : Y \to Z$ (through $f : X \to Y$). If there is a factor h then $g = f; h$ is compatible with f and the defining condition of strict epimorphism then implies that this factor is unique.

If a morphism is a strict epimorphism and a monomorphism then it is an isomorphism. If f is a monomorphism then $f \circ x = f \circ y$ implies $x = y$. That means that id_X is compatible with f, and its unique factor h satisfies $f; h = \mathrm{id}_X$. To see that $h; f = \mathrm{id}_Y$, of id_X through f gives $f^{-1} : Y \to X$.

9.48 Categorical quotient Let $\mathcal{C}$ be a category and X an object of $\mathcal{C}$. Consider the group of automorphisms $\mathrm{Aut}(X)^{\mathrm{op}}$ under sequential composition. Given a group A, a group homomorphism $\alpha : A \to \mathrm{Aut}(X)^{\mathrm{op}}$ constitutes a left action on X: $\alpha(1_A) = \mathrm{id}_X$ and $\alpha(a_1 a_2) = \alpha(a_1); \alpha(a_2)$. The *categorical quotient* $q : X \to X/A$ is defined by the following conditions:

1. It is constant on the orbits of A in X, i.e., $X \xrightarrow{\alpha(a)} X \xrightarrow{q} X/A \ = \ q$ for all $a \in A$.
2. It is universal with respect to this property: given any arrow $f : X \to Z$ such that $\alpha(a); f = f$ for all $a \in A$, there exists a unique $h : X/A \to Z$ such that $q; h = f$.

The morphism q is a strict epimorphism.

9.49 Lemma If $A \subseteq \mathrm{Aut}(X)^{\mathrm{op}}$ is a subgroup, the action of A on X extends to a left action the hom-sets $\mathrm{Hom}(X, Z)$ of $\mathcal{C}$, given by $a \cdot f = f \circ \alpha(a)$. Note that:

- $1_A \cdot f = f \circ \mathrm{id}_X = f$, and
- $(a_1 a_2) \cdot f = f \circ \alpha(a_1 a_2) = f \circ \alpha(a_2) \circ \alpha(a_1) = a_1 \cdot (f \circ \alpha(a_2)) = a_1 \cdot a_2 \cdot f$.

10 Commutative algebra

10.1 Duals of K-modules If K is a commutative semiring, then the dual of a (left) K-module X is again a (left) K-module. (Cf. §7.27.)

The bilinear map $\langle -, - \rangle : X \times X^* \to K$ factors through the universal bilinear map to give natural transformation $\epsilon_X : X \otimes X^* \to K$. Explicitly, the definition is $\epsilon_X(x \otimes y) = y(x)$.

For X^* to be a dual object in the categorical sense, we also need a natural transformation $\eta_X : K \to X^* \otimes X$. It is not clear if this exists...

10.2 Bilinear functions For K-modules X, Y and Z, a function $h : X \times Y \to Z$ is said be *bilinear* if it is linear in each argument:

$$\begin{array}{rclcrcl} h(x_1 + x_2,\, y) & = & h(x_1, y) + h(x_2, y) & \qquad & h(kx,\, y) & = & k \cdot h(x, y) \\ h(x,\, y_1 + y_2) & = & h(x, y_1) + h(x, y_2) & & h(x,\, ky) & = & k \cdot h(x, y) \end{array}$$

Note that a bilinear function is *not* a linear function because

$$h(x_1 + x_2, y_1 + y_2) = \sum_{i=1,2} \sum_{j=1,2} h(x_i, y_j)$$

and this is not the same as $h(x_1, y_1) + h(x_2, y_2)$. It is possible to define a tensor product $X \otimes Y$ such that linear functions $X \otimes Y \to_K Z$ are the same as bilinear functions $X \times Y \to Z$. This we do, in the next paragraph.

10.3 Tensor product The tensor product of K-modules X and Y is another K-module $X \otimes_K Y$ (or $X \otimes Y$, for short) along with a bilinear map $\mu : X \times Y \to X \otimes Y$ such that every bilinear map $X \times Y \to Z$ factors through μ.

To construct $X \otimes Y$, we first construct a free module F with $X \times Y$ as generators. Take F to be the set of all those functions $f : X \times Y \to K$ which have only a finite number of non-zero values. These functions form an K-module under term-wise addition and scalar multiples:

$$\begin{array}{rcl} (f_1 + f_2)(x, y) & = & f_1(x, y) + f_2(x, y) \\ (kf)(x, y) & = & k \cdot f(x, y) \end{array}$$

For $x \in X$ and $y \in Y$, let $[x, y]$ denote the special function in F that is 1_K for (x, y) and 0_K everywhere else. Then every $f \in F$ can be written as a finite linear combination:

$$f = \sum_{x,y} [x, y] \cdot f(x, y)$$

Thus, the elements $[x, y]$ form a possibly infinite basis for F. Define a function $u : X \times Y \to F$ by $u(x, y) = [x, y]$. Every function $h : X \times Y \to Z$ can be expressed as $h = u; s$ for a linear transformation $s : F \to Z$, given by

$$s(f) = \sum_{x,y} h(x, y) \cdot f(x, y)$$

In particular, $s[x, y] = h(x, y)$.

Consider the congruence relation on F generated by the equivalences:

$$\begin{array}{rclcrcl} [x_1 + x_2,\, y] & \equiv & [x_1, y] + [x_2, y] & \qquad & [kx,\, y] & \equiv & k[x, y] \\ [x,\, y_1 + y_2] & \equiv & [x, y_1] + [x, y_2] & & [x,\, ky] & \equiv & k[x, y] \end{array}$$

Now, $X \otimes Y$ is the quotient $F/\equiv$. The equivalence class of $[x, y]$ is written as $x \otimes y$. The map $(x, y) \mapsto x \otimes y$ of type $X \times Y \to X \otimes Y$ is evidently bilinear. Every bilinear map $h : X \times Y \to Z$ factors through $X \otimes Y$ by defining the unique factor $\overline{h} : X \otimes Y \to Z$ as:

$$\overline{h}(x \otimes y) = h(x, y)$$

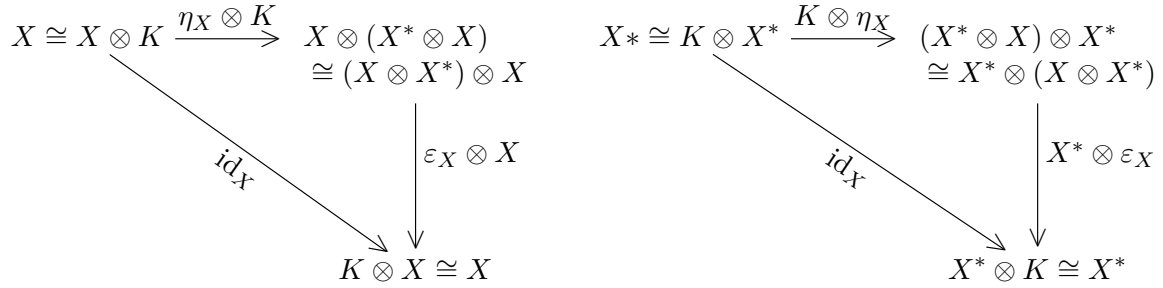
Thus we have shown that bilinear maps $X \times Y \to Z$ are one-to-one with linear maps $X \otimes_K Y \to Z$.

$$\mathrm{Bilin}(X, Y;\, Z) \;\cong\; \mathrm{Hom}_K(X \otimes_K Y,\, Z)$$

10.4 Dualizable objects The evaluation map $\langle -, - \rangle : X \times X^* \to K$ mentioned in §8.7 is bilinear. So, it can be regarded as a linear map $\varepsilon_X : X \otimes X^* \to K$. Explicitly:

$$\varepsilon_X(x \otimes y) = y(x)$$

If, in addition, there is a map $\eta_X : K \to X^* \otimes X$ such that the following diagrams commute, then the module X is said to be *dualizable*:



10.5 Algebras Let K be a commutative semiring. The following definition is based on [Mac Lane and Birkhoff, 1967]:

A *K-algebra* is a K-module A that has an additional binary operation of multiplication making it a semiring. The addition operation is shared between the module and semiring structures and the two forms of multiplication "commute":

$$k(x_1 \cdot x_2) = (kx_1) \cdot x_2 = x_1 \cdot (kx_2)$$

Alternatively, a *K-algebra* is a monoid in the monoidal category $\langle K\text{-}\mathbf{Mod},\, \otimes_K,\, K \rangle$, which means is that it is a K-module equipped with a *bilinear* multiplication operator:

$$\begin{array}{rclcrcl} (x+y)\cdot z & = & xz + yz & \qquad & (kx)\cdot z & = & k(x \cdot z) \\ z \cdot (x+y) & = & zx + zy & & z \cdot (kx) & = & k(z \cdot x) \end{array}$$

The equations on the left amount to distributivity making it a semiring. The multiplication operation then determines a formal multiplication morphism $\mu : A \otimes_K A \to A$. The unit morphism $\eta : K \to A$ maps $k \mapsto k1_A$, in particular $1_K \mapsto 1_A$. These definitions satisfy the standard monoid laws in monoidal categories.

A more classical definition of algebras is the following: A *K-algebra* is a semiring A together with a semiring homomorphism $\eta : K \to A$, called the *unit map*, such that $\eta(K)$ is contained in the center of A. The unit map then defines an action of K on A via $k \cdot x = \eta(k)x$.

Example: A prototypical example is the set of complex numbers, which has addition and multiplication defined by:

$$\begin{array}{rcl} (x_1 + iy_1) + (x_2 + iy_2) & = & (x_1 + x_2) + i(y_1 + y_2) \\ (x_1 + iy_1) \cdot (x_2 + iy_2) & = & (x_1x_2 - y_1y_2) + i(x_1y_2 + x_2y_1) \end{array}$$

Moreover, complex numbers $x + iy$ can be viewed as vectors (x, y) in the complex plane, with pointwise addition and scalar multiplication by reals. The complex multiplication commutes with scalar multiplication:

$$\begin{aligned} k(x_1 + iy_1) \cdot (x_2 + iy_2) &= (kx_1 + iky_1) \cdot (x_2 + iy_2) \\ &= (kx_1x_2 - ky_1y_2) + i(kx_1y_2 + kx_2y_2) \\ &= k((x_1 + iy_1) \cdot (x_2 + iy_2)) \end{aligned}$$

Thus complex numbers form a $\mathbb{R}$-algebra, where $\mathbb{R}$ is the field of reals.

10.6 Endomorphism algebras If X is a K-module, then the set $\mathrm{End}_K(X)$ of K-module endomorphisms is a K-module as well as a semiring under composition of endomorphisms. (Cf. §8.5.) This makes it a K-algebra. As a special case, the set of $n \times n$ matrices over K is a K-algebra with pointwise addition and scalar multiplication forming the K-module structure and matrix multiplication forming the semiring structure.

10.7 Coalgebras and bialgebras Dually, a *K-coalgebra* is a comonoid in the monoidal category $\langle K\text{-}\mathbf{Mod}, \otimes_K, K\rangle$. So, it is a K-module A equipped with linear transformations $\delta : A \to_K A \otimes_K A$ and $\epsilon : A \to_K K$ such that the comonoid laws are satisfied.

A *K-bialgebra* is a K-module with both a monoid and comonoid structure $(A, \mu, \eta, \delta, \epsilon)$ such that μ and η are coalgebra maps, or, equivalently, δ and ϵ are algebra maps.

$$\begin{array}{rclcrcl} \delta(xy) &=& \delta(x)\delta(y) & \qquad & \delta(1_A) &=& 1_A \otimes 1_A \\ \epsilon(xy) &=& \epsilon(x)\epsilon(y) & & \epsilon(1_A) &=& 1_K \end{array}$$

10.8 Inner product spaces For a real vector space X, an *inner product* operation is a bilinear form $\langle -, - \rangle : X \times X \to \mathbb{R}$ that is symmetric and positive definite. Explicitly,

$$\begin{array}{l} \langle x, y\rangle = \langle y, x\rangle \\ \langle 0,\ y\rangle = 0 \\ \langle x_1 + x_2,\ y\rangle = \langle x_1, y\rangle + \langle x_2, y\rangle \\ \langle kx,\ y\rangle = k \cdot \langle x, y\rangle \\ \langle x, x\rangle \geq 0 \\ \langle x, x\rangle = 0 \implies x = 0 \end{array}$$

(The last two properties can be combined into $x \neq 0 \implies \langle x, x\rangle > 0$.) The inner product operation is also often written as $x \cdot y$.

For a complex vector space X, the inner product operation is defined analogously except that the first property is replaced by:

$$\langle x, y\rangle = \overline{\langle y, x\rangle}$$

where $\overline{k}$ is the complex conjugate of k. Note that $\langle x, x\rangle$ is still real-valued.

A standard example of an inner product operation is the *dot product* operation of $\mathbb{R}^n$:

$$x \cdot y \;=\; \sum_{i=1}^{n} x_i y_i$$

All the axioms of inner product are easily verified.

The only dependence on real numbers in the definition of inner products is the *positivity axiom*: $\langle x, x\rangle \geq 0$. If we drop this axiom, we would obtain a general concept that is applicable to all semiring modules.

10.9 Norm and distance The *norm* of a vector (or "length") in an inner product space is defined as $\|x\| = \sqrt{\langle x, x\rangle}$. The following properties of norm can be verified:

$$\begin{aligned}&\|x\| \geq 0\\ &\|x\| = 0 \implies x = 0\\ &\|kx\| = k\|x\|\\ &\|x + y\| \leq \|x\| + \|y\|\end{aligned}$$

If a vector space has *any* norm function $\| - \| : X \to \mathbb{R}$ satisfying the above properties then it is called a *normed vector space*.

The *distance* between two vectors in an inner product space is defined by $d(x, y) = \|x - y\|$. The following properties can be easily verified:

- $d(x, y) = d(y, x)$.
- $d(x, y) \geq 0$.
- $d(x, y) = 0 \implies x = y$.
- $d(x, z) \leq d(x, y) + d(y, z)$.

Other example of norms on $\mathbb{R}^n$ include:

- the *Manhattan norm* or L^1*-norm* (also called the *taxicab norm*), given by:

  $$\|\vec{x}\| = |x_1| + \cdots + |x_n|$$

  with the corresponding distance function:

  $$d(\vec{x}, \vec{y}) = |x_1 - y_1| + \cdots + |x_n - y_n|$$

- the *Chebyshev norm* or *supremum norm*, given by:

  $$\|(x_1, \ldots, x_n)\| = \max(|x_1|, \ldots, |x_n|)$$

  with the corresponding distance function:

  $$d(\vec{x}, \vec{y}) = \max(|x_1 - y_1|, \ldots, |x_n - y_n|)$$

- the L^p*-nom*, given by

  $$\|\vec{x}\| = (x_1^p + \cdots + x_n^p)^{\frac{1}{p}}$$

Note that the Manhattan norm is the special case of L^p-norm for $p = 1$. The standard Euclidean norm is the case of $p = 2$. The Chebyshev norm is the limit of L^p-norm as $p \to \infty$.

It is amusing to note that the distance seen by a rook on a chess board is the Manhattan distance. That seen by a king or queen is the Chebyshev distance.

10.10 Angle and orthogonality The inner product norm also satisfies the following *Cauchy-Schwartz inequality.*

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$$

For real vector spaces, the Cauchy-Schwartz inequality implies that the ratio $\frac{\langle x,y\rangle}{\|x\|\cdot\|y\|}$ is in the interval $[-1, 1]$. So, it is equal to $\cos\theta$ for exactly one angle θ, which we define to be the *angle* between x and y. In terms of this angle, we can rewrite the Cauchy-Scwartz inequality as an equation:

$$\langle x, y\rangle = \|x\| \cdot \|y\| \cdot \cos\theta$$

The triangle inequality can be derived from the Cauchy-Scwartz inequality: ...

Two vectors are said to be *orthogonal* (or *perpendicular*) if $\langle x, y\rangle = 0$. We denote such a fact by $x \perp y$. For any fixed vector x, the bilinear form $\langle x, y\rangle$ is linear in y, i.e., $\langle x, \sum_{i=1}^n k_i y_i\rangle = \sum_{i=1}^n k_i\langle x, y_i\rangle$. So, $y_1, \ldots, y_n$ are orthogonal to x, then any linear combination of them is also orthogonal to x. Consequently, the set of all vectors orthogonal to x is a subspace of X, called the *orthogonal complement* of x.

A vector x is said to be *orthogonal* to a subspace $S \subseteq X$ if it is orthogonal to every vector in S. By linearity again, it suffices for x to be orthogonal to every basis vector of S. The set of all vectors orthogonal to a subspace S is another subspace, denoted $S^\perp$, and called the *orthogonal complement* of S.

10.11 Metric spaces A *metric space* is a set X together with a function $d : X \times X \to \mathbb{R}$, presumed to define the "distance" between two elements, satisfying the following properties:

$$\begin{aligned}
&d(x, y) = d(y, x)\\
&d(x, y) > 0 \text{ if } x \neq y\\
&d(x, x) = 0\\
&d(x, z) \leq d(x, y) + d(y, z)
\end{aligned}$$

Metric spaces can be regarded as topological spaces using open balls around points.

A *convergent sequence* (or *Cauchy sequence*) in a metric space is a sequence $x_1, x_2, \ldots$ such that for every real number $\epsilon > 0$ there is a positive integer N satisfying $m, n > N \implies d(x_m, x_n) < \epsilon$. The metric space is said to be *complete* if every convergent sequence has a limit. Every finite-dimensional inner product space is automatically complete. Completeness is an issue only for infinite-dimensional spaces.

Example: Normed vector spaces Evidently every normed vector space has a metric $d(x, y) = \|x - y\|$ and forms a metric space.

Example: Discrete metric The discrete metric on X is given by $d(x, x) = 0$, and $d(x, y) = 1$ whenever $x \neq y$.

10.12 Metric continuity If X and Y are metric spaces, a function $f : X \to Y$ is said to be *continuous* at $x_0 \in X$ if, for every $\epsilon > 0$, there exists $\delta > 0$ such that $d(x, x_0) < \delta \implies d(f(x), f(x_0)) < \epsilon$. The function f is *continuous* if it is continuous at every $x_0 \in X$.

10.13 Euclidean and Hilbert spaces A *Hilbert space* is an inner product space which is complete with the standard metric $d(x, y) = \|x - y\|$. Every finite-dimensional inner product space is automatically complete, and hence a Hilbert space. Completeness is an issue only for infinite-dimensional spaces.

10.14 Metric topology If $x \in X$ is a point in a metric space, an *open ball* around x of radius $\epsilon > 0$ is the set $B(x, \epsilon) = \{\, y \mid d(x, y) < \epsilon \,\}$. A subset $u \subseteq X$ is said to be *open* if, for every $x \in u$, there is an open ball $B(x, \epsilon_x)$ included in u. Note that open balls themselves are open sets. If $y \in B(x, \epsilon)$ in any point in an open ball then $B(y, \epsilon - d(x, y))$ is included in it, showing that it is an open set.

A subset $u \subseteq X$ is *closed* if its complement $X - u$ is open. Evidently, a set u is closed iff, whenever every open ball $B(x, \epsilon)$ around x meets u, $x \in u$ [Willard, 1970, Def. 2.5]. Note that:

$$B(x, \epsilon) \subseteq X - u \iff B(x, \epsilon) \cap u = \emptyset$$

$X - u$ being an open set means that, if $x \in X - u$, there is $B(x, \epsilon)$ included in $X - u$. The contrapositive of the statement is that, if, for all ϵ, $B(x, \epsilon) \not\subseteq X - u$, then $x \notin X - u$, i.e., $x \in u$.

The open sets induced by the metric form a topology. That means:

- If $u_1, \ldots, u_n$ are open sets, then the intersection $u = \bigcap_{i=1,n} u_i$ is an open set. Given a point $x \in u$, note that x is a member of each u_i. So, for each $i = 1, n$, there is a radius $\epsilon_i > 0$ such that $B(x, \epsilon_i) \subseteq u_i$. If $\epsilon \leq \min_i \epsilon_i$ is any positive real number then $B(x, \epsilon) \subseteq u$. (In particular, if $n = 0$, then u is the entire space X, and ϵ can be any positive real number.)
- If $\{u_i\}_{i \in I}$ is a family of open sets and $u = \bigcup_{i \in I} u_i$, then a point $x \in u$ is a point in some u_i. So, we can take the radius ϵ_i of the open ball around x in u_i as the radius for the open ball around x in u. (If I is empty then $u = \emptyset$, which is vacuously open.)

We call the topology induced by the metric the *metric topology* on the space. (This is also often called the "standard" topology). The open balls around points form a *basis* for the metric topology, because every open set u can be written as $\bigcup_{x \in u} B(x, \epsilon_x)$.[27]

10.15 Theorem *A function $f : X \to Y$ between metric spaces is continuous at x_0 iff, for each open set $v \subseteq Y$, there is an open set $u \in X$ such that $f[u] \subseteq v$.* [Willard, 1970, Theorem 2.8].

The forward direction is quite immediate. If v is an open set containing $f(x)$, then it has an open ball $B(f(x), \epsilon) \subseteq v$ for some $\epsilon > 0$. By continuity of f, there is $\delta > 0$ such that $f[B(x, \delta)] \subseteq B(f(x), \epsilon)$. Take $u = B(x, \delta)$.

In the backward direction, suppose that for each open set v containing $f(x)$ there is an open set containing x such that $f[u] \subseteq v$. Given $\epsilon > 0$, consider the open set $v = B(f(x), \epsilon)$. There is a corresponding open set u containing x such that $f[u] \subseteq v$. But, since $x \in u$ and u is an open set, there is an open ball $B(x, \delta)$ contained in u for some $\delta > 0$. Since $f[B(x, \delta)] \subseteq v$, we have found a δ corresponding to ϵ, showing that f is continuous at x_0.

10.16 Bilinear forms Inner products represent an instance of a more general notion.

Any commutative semiring K is a module under its own action. A bilinear transformation $X \times Y \to K$, *i.e.*, a linear transformation of type $X \otimes_K Y \to K$, is called *bilinear form*. The inner product operation is evidently a bilinear form $X \times X \to K$.

A bilinear form $\beta : X \times X \to K$ is said to be

- *symmetric* if $\beta(x, y) = \beta(y, x)$,

[27] This is not entirely clear.

- *skew-symmetric* if $\beta(x,y) = -\beta(y,x)$,
- *alternating* if $\beta(x,x) = 0$, and
- *reflexive* if $\beta(x,y) = 0 \iff \beta(y,x) = 0$.

A reflexive bilinear form is necessarily symmetric or alternating. The inner product is both symmetric and alternating, hence reflexive.

Any reflexive bilinear form gives rise to a notion of orthogonality: $x \perp_\beta y \iff \beta(x,y) = 0$.

10.17 Quadratic forms If K is a field of characteristic other than 2, then a *quadratic form* is a unary function $q : X \to K$ such that $q(-x) = q(x)$ and the function $h : X \times X \to K$ defined by

$$2 \cdot h(x,y) = q(x+y) - q(x) - q(y)$$

is a bilinear form. We say that the bilinear form h is obtained by "polarizing" the quadratic form q.

A simple example of quadratic forms is the function $q : k \mapsto k^2$ of K. Evidently $q(-k) = q(k)$ and $h(k,l) = kl$ is bilinear. (Note that $2kl = (k+l)^2 - k^2 - l^2$.)

10.18 Affine space Intuitively, an affine space is a vector space that has "forgotten its origin."

Formally, an *affine space* is a non-empty set X (of "points") together with a vector space ${}_{\mathbb{R}}V$, and a faithful and transitive group action by the underlying additive group of V. We use the additive notation for the action: $(v,x) \mapsto v + x$. The fact that it is a group action implies:

$$0 + x = x \qquad (v+w) + x = v + (w + x) \qquad (-v) + x = y \iff v + y = x$$

Since the action is transitive, there is a single orbit, and faithfulness implies $v + x = v' + x$ iff $v = v'$. In other words, for any $x \in X$, $v \mapsto v + x$ is a bijection $|V| \cong X$.

Define a subtraction operation for the points of X by

$$y - x \;=\; \text{the unique vector } v \text{ such that } v + x = y$$

Note that the difference of two points is a *vector*. It is easy to verify:

$$(y-x) + x = y \qquad (z-y) + (y-x) = z - x \qquad y - x = 0 \iff y = x$$

For any vector $v \in V$, the translation $\tau_v : x \mapsto v + x$ is an isomorphism $X \cong X$. The translations can be made into a vector space by defining $\tau_v + \tau_w = \tau_{v+w}$ and $k \cdot \tau_v = \tau_{kv}$. This vector space, which is isomorphic to V, is called the *space of translations* of X and denoted $X^\sharp$.

11 Lattices and Boolean algebras

11.1 Lattices Recall from §1.17 that a (unital) join-semilattice $(A, \vee, \bot)$ is a commutative semigroup where the operation is idempotent. In this section, we assume that semilattices are always unital. A join-semilattice has an induced partial order $a \leq b \iff a \vee b = b$. Dually, a meet-semilattice $(A, \wedge, \top)$ is again an idempotent commutative monoid whose induced partial order is defined by $a \leq b \iff a \wedge b = a$. (We feel free to omit $\wedge$ in writing $a \wedge b$ as ab.) A set with both the semilattice structures where the induced partial order is the same in both cases is called a *lattice*. Insisting that the induced partial should be the same in both cases is equivalent to requiring the absorption laws:

$$a \wedge (a \vee b) = a \qquad a \vee (a \wedge b) = a \tag{11.1}$$

A *homomorphism* of lattices preserves all the four components of the structure $(\vee, \bot, \wedge, \top)$. It is necessarily monotone. A *logical relation* of lattices is compatible with all the four components of the structure. The reflexive graph category of lattices is denoted **Lat**.

11.2 Lattices from order Alternatively, the structure of lattices can be defined starting with posets. If $(A, \leq)$ is a poset, $a \vee b$ is the least upper bound of a and b, $a \wedge b$ is the greatest lower bound, $\bot$ is the least element and $\top$ is the greatest element. If a poset has all four components of the structure, it is a lattice.

Note that we can treat the poset A as a category, with unique morphisms $a \to b$ whenever $a \leq b$. Then $a \vee b$ is the categorical coproduct, $a \wedge b$ is the categorical product, $\bot$ is the initial object and $\top$ is the terminal object. For example, saying that $a \wedge b$ is the categorical product of a and b means:

$$x \leq a \wedge b \iff (x \leq a) \wedge (x \leq b)$$

The other operations are characterized similarly. The dual of A is the poset $A^{\mathrm{op}} = (A, \geq)$ with the order inverted. The products in A become the coproducts in A^{op} and so on.

11.3 Distributive lattices A *distributive lattice* is a lattice that satisfies either one of the following laws:

$$\begin{aligned} a \wedge (b \vee c) &= (a \wedge b) \vee (a \wedge c) \\ a \vee (b \wedge c) &= (a \vee b) \wedge (a \vee c) \end{aligned}$$

The two laws are *equivalent* in a lattice. Once we have a distributive lattice, the two absorption laws (11.1) are equivalent to each other.

The reflexive graph subcategory of distributive lattices is denoted **DLat**.

11.4 Complements In a lattice, the *complement* of an element a is an element x satisfying $x \wedge a = \bot$ and $x \vee a = \top$. Such a complement need not be unique.

However, in a *distributive* lattice, the complement of an element a is necessarily unique. Suppose x and x' are both complements of a. Then,

$$\begin{aligned} x &= x \wedge (x \vee a) && \text{by absorption} \\ &= x \wedge \top \\ &= x \wedge (x' \vee a) \\ &= (x \wedge x') \vee (x \wedge a) && \text{by distributivity} \\ &= (x \wedge x') \vee \bot \\ &= x \wedge x' \end{aligned}$$

This implies $x \leq x'$. By a symmetric argument, we conclude $x' \leq x$ and, hence, $x = x'$.

We denote the unique complement of a as $\neg a$.

11.5 Boolean algebras A *Boolean algebra* is a distributive lattice with the additional operation of complement $\neg$. Homomorphisms and logical relations of boolean algebras are taken to be those preserving complements, giving rise to a reflexive graph category **BA**.

A lattice homomorphism $h : A \to B$ between Boolean algebras A and B is automatically a Boolean algebra homomorphism.

To see that h preserves complements, suppose $h(x) = y$ and $h(\neg x) = y'$. Since $x \vee \neg x = \top_A$, we obtain $y \vee y' = \top_B$. Since $x \wedge \neg x = \bot_A$, we obtain $y \wedge y' = \bot_B$. Hence $y' = \neg y$.

Thus the category **BA** is a *full subcategory* of **Lat**. This situation is similar to that of **Grp** being a full subcategory of **SGrp** (cf. §1.7).

11.6 Heyting algebras A *Heyting algebra* is a lattice with an operation $a \to b$, for every pair of elements a and b, satisfying the equivalence:

$$x \leq (a \to b) \iff x \wedge a \leq b$$

When the lattice is viewed as a category, $a \to b$ is the exponential b^a. Homomorphisms and logical relations of Heyting algebras are defined to preserve the exponential. This data gives a reflexive graph category **HA**. (Unlike **BA**, **HA** is not a full subcategory of **Lat**.)

An equivalent definition of $a \to b$ is that it is the *greatest element x such that $x \wedge a \leq b$*. In other words, $a \to b$ satisfies $(a \to b) \wedge a \leq b$ and, if any element x satisfies $x \wedge a \leq b$ then $x \leq a \to b$.

Given that $\neg a$ in a Boolean algebra is the greatest element x such that $x \wedge a \leq \bot$, we can think of $a \to b$ as a "relative complement" of a relative to b. It can also be thought of as the "residuation" of a with respect to b.

A Heyting algebra is necessarily a distributive lattice.

Cf. [Johnstone, 1982, I.1.9]. Since we always have $ab \vee ac \leq a(b \vee c)$ from the monotonicity of $\wedge$, we need to prove the reverse inequality. Monotonicity of $a \to (-)$ implies $a \to ab \leq a \to (ab \vee ac)$ and, similarly, $a \to ac \leq a \to (ab \vee ac)$. Hence

$$(a \to ab) \vee (a \to ac) \leq a \to (ab \vee ac)$$

Since $b \leq (a \to ab)$ and $c \leq (a \to ac)$, we have

$$b \vee c \leq a \to (ab \vee ac)$$

That implies $a(b \vee c) \leq (ab \vee ac)$.

11.7 Pseudocomplements In a Heyting algebra, the *pseudocomplement* of a is defined by $\neg a = a \to \bot$, *i.e.*, the greatest element x such that $x \wedge a \leq \bot$. Unlike in a Boolean algebra, $\neg a \vee a = \top$ may not hold.

A Heyting algebra A is a Boolean algebra if and only if $\neg\neg a = a$ for all $a \in A$.

Elements of the form $\neg\neg a$ in a Heyting algebra are called the *regular elements* of A. It can be shown that all the regular elements of A form a Boolean algebra $A_{\neg\neg}$. The partial order and meets of $A_{\neg\neg}$ are the same as those of A, but the joins may be different.

11.8 Ideals An *ideal* (also called *order ideal*) I in a join-semilattice A is a lower set that is closed under joins (and, hence, a join-subsemilattice).

$$\begin{aligned} &a \in I \implies \forall x \leq a.\, x \in I \\ &\bot \in I \\ &a, b \in I \implies a \vee b \in I \end{aligned}$$

If A is a *distributive lattice*, then it is also a semiring with meets distributing over joins. An ideal of A in the sense of semiring ideals (§8.22) is the same as an order ideal because a subset being closed under meets by A is the same as being a lower set:

$$\exists c \in A.\, x = a \wedge c \iff x \leq a$$

Thus order ideal is a generalization of the semiring ideal for semilattices.

The congruence relation induced by an order ideal I is the Bourne relation (§8.22):

$$a \sim_I b \iff \exists i, j \in I.\, a \vee i = b \vee j$$

This is compatible with joins because $a \vee i = b \vee j$ and $a' \vee i' = b' \vee j'$ imply ... The semilattice $A/I \stackrel{\text{def}}{=} A/\!\sim_I$ is the evident quotient with equivalence classes $[a]_I$ as elements. If A is a distributive lattice then A/I is a distributive lattice.[28]

Old text that needs to be adapted:

If $f : A \to B$ is a semilattice homomorphism then $f^{-1}(\bot)$, called the *kernel* of f, is an ideal of A.

Conversely, for every ideal I of A, there exists a semilattice homomorphism $f : A \to B$ whose kernel is I. The semilattice B is obtained as the quotient of A under an equivalence relation $\sim_I$ given by:

$$a \sim_I b \iff \exists i, j \in I.\, a \vee i = b \vee j$$

The semilattice $B = A/\!\sim_I$ is the evident quotient with equivalence classes $[a]_I$ as elements. The homomorphism f is the projection $f(a) = [a]_I$, whose kernel $f^{-1}([\bot]_I)$ is precisely I.

These observations generalize to *distributive* lattices. If A is a distributive lattice with an ideal I, then $A/\!\sim_I$ is in turn a distributive lattice. To see that the meet operation is well-defined, note that $a \sim_I b$ implies $a \wedge c \sim_I b \wedge c$. (Suppose $a \vee i = b \vee j$ for some $i, j \in I$. Then, $(a \vee i)c = (b \vee j)c$, which is to say $ac \vee ic = bc \vee jc$. Note that ic and jc are in I because I is a lower set. Thus $ac \sim_I bc$.)

Distributivity of the lattice is necessary for this generalization. For example, consider the lattice $\{\bot\} \leq \{a, b, c\} \leq \{\top\}$. It is not distributive: $a \wedge (b \vee c) \neq (a \wedge b) \vee (a \wedge c)$. Let I be the ideal $\{\bot, a\}$. We find that $b \sim_I c$. However, $b \wedge b \not\sim_I c \wedge b$.

11.9 Filters The concept of *filter* is dual to that of ideal. Explicitly, a filter F in a meet-semilattice A is an upper set that is closed under meets.

$$\begin{aligned} &a \in F \implies \forall x \geq a.\, x \in F \\ &\top \in F \\ &a, b \in F \implies a \wedge b \in F \end{aligned}$$

If $f : A \to B$ is a meet-semilattice homomorphism then $f^{-1}(\top)$ is a filter of A. The quotient of A under a filter F is the set of equivalence classes under the equivalence relation $\sim_F$ given by:

$$a \sim_F b \iff \exists i, j \in F.\, a \wedge i = b \wedge j$$

[28] Need to check Bourne relations for commutative monoids, instead of semirings.

A subset F_0 of A that is a directed set in the downward direction is called a *filter base*.

$$a, b \in F_0 \implies \exists c \in F_0.\, c \leq ab$$

The upper set of a filter base $\uparrow F_0$ is a filter. (Even though ab may not be included in F_0, it will be included in $\uparrow F_0$.)

Examples [Dixmier, 1984, Sec. 2.1]: If X is a set, we say "filter on X" for a filter in the lattice $\mathcal{P}X$, but not containing $\emptyset$.

- If X is a topological space, the set v of neighbourhoods of a point x is a filter on X. If v_0 is a fundamental system of neighbourhoods of x_0 then v_0 is a filter base on X.
- Let $x \in \mathbb{R}$. The set of open intervals $(x-\epsilon,\, x+\epsilon)$, for all $\epsilon > 0$, is a filter base on $\mathbb{R}$. More such examples include the set of intervals $[x, x+\epsilon)$, $(x, x+\epsilon)$, and $(x-\epsilon, x) \cup (x, x+\epsilon)$.
- In the set of natural numbers $\mathbb{N}$, the set of subsets of the form $\{n, n+1, \ldots\}$ is a filter base on $\mathbb{N}$.
- If X is a topological space, $Y \subseteq X$, and $x \in X - Y$, then the set of subsets of the form $u \cap Y$, where u is a neighbourhood of x in X, is a filter on Y.

11.10 Prime ideals An ideal I in a lattice A is said to be *prime* if $\top \notin I$ and $a \wedge b \in I$ implies $a \in I$ or $b \in I$. Dually, a filter F is a *prime filter* if $\bot \notin F$ and $a \vee b \in F$ implies $a \in F$ or $b \in F$.

The set complement of a prime ideal is a prime filter.

We first define a lattice homomorphism $f : A \to \mathbf{2}$, whose kernel is the given prime ideal I. The definition is: $f(x) = \bot$ for all $x \in I$ and $f(x') = \top$ for all $x' \notin I$. We verify that f preserves meets. Consider $f(x \wedge y)$. If both $x, y \in I$ then $x \wedge y \in I$ and $f(x \wedge y) = \bot$. If $x, y \notin I$ then $x \wedge y \notin I$ by the prime-ness of the ideal. So, $f(x \wedge y) = \top$. If $x \in I$ and $y \notin I$ then since I is a lower set, $x \wedge y \in I$ and $f(x \wedge y) = \bot = f(x) \wedge f(y)$. Similarly, it is easy to see that f preserves joins, $\bot$ and $\top$.

Evidently, $f^{-1}(\bot)$ and $f^{-1}(\top)$ are complements of each other and the latter is a filter.

Conversely, if the set complement of an ideal is a filter then the ideal is prime.

Let $F = A \backslash I$ be the filter. Since $\top \in F$ by definition, $\top \notin I$. Suppose $a \wedge b \in I$ for elements $a, b \notin I$. We have that $a, b \in F$ and, therefore, $a \wedge b \in F$, a contradiction. Hence, one of a and b must be in I, and that means that I is a prime ideal.

11.11 Theorem *Let F be a filter in a distributive lattice A and I an ideal which is maximal among those disjoint from F. Then I is prime [Johnstone, 1982, I.2.4].*

12 Posets and domains

12.1 Posets A *poset* is a set with a partial order $A = (A, \sqsubseteq_A)$. The *join* or *supremum* of a subset $u \subseteq A$ is written as $\bigsqcup u$. The *meet* or *infimum* is written as $\bigsqcap u$. The binary versions of these operations are written as $\sqcup$ and $\sqcap$ respectively. The least and greatest elements of A, which are regarded as the supremum and infimum of the empty subset, are written as $\bot_A$ and $\top_A$ respectively. Note that none of these operations may actually exist in a given poset.

A function $f : (A, \sqsubseteq_A) \to (B, \sqsubseteq_B)$ is said to be *monotone* if it preserves the order: $x \sqsubseteq_A y \implies f(x) \sqsubseteq_B f(y)$. Posets and monotone functions between them form a category **Poset**. The category is cartesian closed with products and exponentials given pointwise.

The *dual* of a poset $(A, \sqsubseteq_A)$ is the poset $(A, \sqsupseteq_A)$.

12.2 Directed sets Directed sets capture the idea of "tending towards a limit," where "limit" means the supremum of a subset. Formally, a subset $d \subseteq A$ is a *directed set* if every finite subset of d has an upper bound in d. Since the empty set is a subset of d, the requirement to include an upper bound for it implies that d is necessarily *nonempty*. Alternatively, we can also define a directed set as a nonempty subset $d \subseteq A$ that has an upper bound for every pair of elements in d.

Note that nonempty totally ordered subsets (called *chains*) are directed sets. Similarly, increasing sequences $x_1 \sqsubseteq x_1 \sqsubseteq \cdots$ are also directed sets. For an example that is not a chain or a sequence, consider the set of all finite subsets of a set X. If $a, a' \subseteq_{\text{fin}} X$, the set $a \cup a'$ is also a finite subset of X. Thus $\mathcal{P}_{\text{fin}}X$ is a directed set (in $\mathcal{P}\mathcal{P}X$).

The directed sets in A are preorderd by the relation $d \preceq d' \iff \forall x \in d.\, \exists x' \in d'.\, x \sqsubseteq x'$. Note that $d \preceq d'$ implies $\bigsqcup d \sqsubseteq \bigsqcup d'$. If $d \preceq d'$ and $d' \preceq d$, we write $d \approx d'$.

12.3 Directed-complete partial orders A poset in which every directed subset has a supremum is called a *directed complete partial order* or *dcpo*. Similarly, one has the notions of *chain complete partial orders* and *ω-complete partial orders* (the latter when the poset has all sups of increasing sequences).

The following result gives insight into the meaning of directed-completeness:

A semilattice that is also directed-complete is a complete semilattice.

To see this, consider an arbitrary subset $u \subseteq A$ of the semilattice. The set of all finite subsets of u is a directed set in $\mathcal{P}A$. Since A is a semilattice each of these finite sets has a supremum, and the set of the suprema $d = \{ \bigsqcup a \mid a \subseteq_{\text{fin}} u \}$ forms a directed set in A. It is easy to see that $\bigsqcup u = \bigsqcup d$ and $\bigsqcup d$ exists by directed-completeness.

A *continuous function* $f : A \to B$ between dcpo's is a monotone function that preserves directed sup's: $f(\bigsqcup d) = \bigsqcup f[d]$.

12.4 Order ideals and filters A directed set that is also down-closed (or a *lower set*) is called an *order ideal* (or just *ideal*, when it is clear from the context).

We write $\downarrow a$ for the down-closure of a subset $a \subseteq A$. Note that $d \approx \downarrow d$ for any directed set d, which has the consequence that $\bigsqcup d = \bigsqcup(\downarrow d)$. A lower set of the form $\downarrow\{x\}$ is always an ideal, called the *principal ideal* generated by x.

A *prime ideal* is an ideal d with the property that $x \sqcap y \in d$ implies $x \in d$ or $y \in d$. If a principal ideal $\downarrow\{x\}$ is a prime ideal, the element x is called a *prime element*.

A *filter* in a poset A is an ideal in A^{op}, *i.e.*, it is an upper set that is directed with respect to the $\sqsupseteq_A$ order.

An ideal d is prime iff its complement $A\backslash d$ is a filter.

12.5 Compact elements An element $x \in A$ is called a *compact element* if, whenever $x \sqsubseteq \bigsqcup d$ for a directed set d, there exists an element $y \in d$ such that $x \sqsubseteq y$. In words, to go "beyond" a compact element x, a directed set must actually "pass" it. The set of compact elements of A is denoted $K(A)$. The non-compact elements are referred to as "limit elements."

The compact elements of A carry the intuition of being "finite elements." For example, in the complete lattice $\mathcal{P}X$, the compact elements are precisely the finite sets.

If A has a least element $\bot$ then it is compact (trivially because every directed set is nonempty). If A has a join $k_1 \sqcup k_2$ for compact elements k_1 and k_2, then the join is a compact element. For if a directed set d is to go beyond $k_1 \sqcup k_2$, then it must pass each k_i. Since k_i are compact elements, d must pass them, via some elements $y_1, y_2 \in d$. But, being a directed set, it needs to have an upper bound y of y_1 and y_2. We then obtain that $k_1 \sqcup k_2 \sqsubseteq y$ and thus the directed set must pass $k_1 \sqcup k_2$ as well. Notice, however, that these facts hold only if $\bot$ and $k_1 \sqcup k_2$ actually *exist* in A.

If $a \subseteq A$, we use the notation $\downarrow^0 a$ for $\downarrow a \cap K(A)$, and refer to it as the *compact lower set of a*. Note that, in any dcpo, we have $x = \bigsqcup(\downarrow\{x\})$. If $x = \bigsqcup(\downarrow^0\{x\})$, then we call the dcpo *algebraic* [Gierz et al., 2003, Def. I-4.2].

12.6 Algebraic dcpo's In more detail, a dcpo is said to be *algebraic* if every element can be expressed as a directed supremum of compact elements. (Alternatively, the compact elements "generate" the dcpo via the operation of directed sup's.) For $\downarrow^0\{x\}$ to be a directed set, it must have an upper bound for every finite subset. If $k_1, k_2 \in \downarrow^0\{x\}$ then $k_1 \sqcup k_2$ is a compact element and belongs to $\downarrow^0\{x\}$. So, nonempty subsets of $\downarrow^0\{x\}$ have upper bounds in it. However, we also need that the empty subset to have an upper bound in $\downarrow^0\{x\}$, which is equivalent to requiring that $\downarrow^0\{x\}$ should be nonempty.

If A is a *pointed* dcpo, *i.e.*, with a least element $\bot$, then $\bot$ is a compact element, and always belongs to $\downarrow^0\{x\}$. A pointed dcpo that is algebraic is called a *domain*. An unpointed dcpo that is algebraic is called a *predomain*.

The terminology of "algebraic" owes to the fact that the lattices of subalgebras of an algebra always have this property.

12.7 Way below relation An element k is said to be *way below* x, denoted $k \ll x$ if, whenever $x \sqsubseteq \bigsqcup d$ for a directed set d, there exists $y \in d$ such that $k \sqsubseteq y$. Clearly, k is a compact element iff $k \ll k$. So, the way-below relation is a form of relative compactness. It is also regarded as an *approximation* relation [Abramsky and Jung, 1994].

The following properties hold for the way-below relation:

- $k \ll x$ implies $k \sqsubseteq x$.
- $k' \sqsubseteq k \ll x \ll x'$ implies $k' \ll x'$.
- $\bot \ll x$ whenever A has a least element $\bot$.

- $k \ll x$ and $k' \ll x$ imply $k \sqcup k' \ll x$ whenever $k \sqcup k'$ exists in A.

The set $\{ k \in A \mid k \ll x \}$ is denoted $\downarrow\!\!\downarrow x$

13 Topological spaces

13.1 Topological spaces A *topological space* is a pair $X = (|X|, \Omega_X)$ where $|X|$ is a set and $\Omega_X \subseteq \mathcal{P}X$ is a family of subsets of $|X|$ (called the *open sets* of X) that is closed under finite intersections and arbitrary unions. The family Ω_X is referred to as a *topology* on $|X|$.

A *continuous function* $f : X \to Y$ between topological spaces is a function on the underlying sets such that the inverse image $f^{-1}[v]$ of an open set $v \in \Omega_Y$ is an open set in X. In other words, the inverse image function f^{-1} is a function of type $\Omega_Y \to \Omega_X$.

Example: Metric spaces A metric space $X = (|X|, d)$ has a standard topology called the *metric topology*

13.2 Closed sets A subset S of a space X is said to be *closed* if its complement $X \backslash S$ is open.

A subset that is both closed and open is called a *clopen* set.

13.3 Discrete topology A topological space is said to be *discrete* if all subsets of X are regarded as open sets.

13.4 Topological subspaces If X is a topological space, a *subspace* $S \subseteq X$ is a subset $|S|$ of the underlying set $|X|$ with the topology $\Omega_S = \{\, u \cap |S| \mid u \in \Omega_X \,\}$. To verify that Ω_S consitutes a topology on $|S|$:

- Let $s_1, \ldots, s_n \in \Omega_S$. Each s_i is of the form $u_i \cap |S|$ for some $u_i \in \Omega_X$. The intersection $\bigcap_{i=1.n} s_i$ is $\bigcap_{i=1,n}(u_i \cap |S|) = (\bigcap_{i=1,n} u_i) \cap |S|$ which is clearly in Ω_S.
- Let $\{s_i\}_{i \in I} \in \Omega_S$, with each s_i being of the form $u_i \cap |S|$ for some $u_i \in \Omega_X$. Their union $\bigcup_{i \in I} s_i$ is $\bigcup_{i \in I}(u_i \cap |S|) = (\bigcup_{i \in I} u_i) \cap |S|$ by the fact that intersection distributes over union. Thus the union is in Ω_S.

14 Matroid theory

Matroid theory originated from studying the abstract linear independence in collections of vectors [Whitney, 1935]. We start with the axiomatization of linear dependence by van der Waerden [van der Waerden, 1949].

14.1 Axioms for linear dependence Let X be a set. We are interested in a relation $U \longrightarrow v$ where $U \subseteq_{\text{fin}} X$ and $v \in X$. This is called an *(abstract) dependence relation*. An example is linear dependence, i.e., v is expressible as a linear combination of vectors in U. We extend the notation to finite sets of elements on both sides of the arrow by saying $U \longrightarrow \{v_1, \ldots, v_n\}$ if $U \longrightarrow v_i$ for each $i = 1, n$.

Linear dependence is then abstractly characterized by the following axioms:

- (Reflexive) If $u \in U$ then $U \longrightarrow u$.
- (Transitive) If $U \longrightarrow V$ and $V \longrightarrow w$ then $U \longrightarrow w$.
- (Exchange) If $U \uplus \{v\} \longrightarrow w$ but not $U \longrightarrow w$ then $U \uplus \{w\} \longrightarrow v$.

The axiomatization can be extended to *infinite sets* by adding:

- (Finitary) If $U \longrightarrow v$ then there is finite $U_0 \subseteq U$ such that $U_0 \longrightarrow v$.

The first two axioms say that dependence is a *closure* operator. Denote the set of elements dependent on U by $\text{cl}(U)$. Then the axioms can be restated as

- (Reflexive) $U \subseteq \text{cl}(U)$.
- (Transitive) $\text{cl}(U) \subseteq \text{cl}(\text{cl}(U))$.
- (Exchange) If $w \in \text{cl}(U \uplus \{v\}) - \text{cl}(U)$ then $v \in \text{cl}(U \uplus \{w\})$.
- (Finitary) If $v \in \text{cl}(U)$ then there is finite $U_0 \subseteq U$ such that $v \in \text{cl}(U_0)$.

The third axiom is called the *Steinitz exchange principle*, used crucially in algorithms such as Gaussian elimination. The extension to infinite sets amounts to saying $\text{cl}(U) = \bigcup \{ \text{cl}(U_0) \mid U_0 \subseteq_{\text{fin}} U \}$.

A finite set U is said to be *independent* if none of its elements is dependent on the others.

Two finite sets U and V are said to be *equivalent* if $U \longrightarrow V$ and $V \longrightarrow U$. Write $U \approx V$ in this situation.

14.2 Basis If $X = \text{cl}(U)$ then U is called a *spanning set* for X. An independent spanning set is called a *basis*. The following statements are equivalent:

- U is a maximal independent subset of X.
- U is a minimal spanning set for X.
- U is a basis for X.

14.3 Steinitz Replacement theorem If U and V are finite sets of size m and n respectively such that U is independent and $U \longrightarrow V$, then there exists $V' \subseteq V$ of size exactly m such that $U \approx V'$. In particular, $m \leq n$.

This is proved by induction on m using the exchange principle.

It follows that any two bases of a given set X are of the same size.

14.4 Matroids Whitney reformulated abstract dependence using the notion of independence instead. Consider a set X and a set of subsets $I \subseteq \mathcal{P}_{\text{fin}}(X)$, whose elements are called "independent subsets." The pair (X, I) is called an independence system. It is a *matroid* if it satisfies the following axioms:

- I is non-empty.
- I is closed under subsets.
- If U and V are in I and $|V| = |U| + 1$, then there is an element in $v \in V - U$ such that $U \uplus \{v\}$ is in I.

The last axiom is, once again, the Steinitz exchange principle.

Every dependence relation determines a matroid and *vice versa*.[29]

If (X, I) is a matroid then a submatroid (X', I') is a matroid such that $X' \subseteq X$ and $I' = I \upharpoonright X'$ is the restriction of I to X'.

14.5 Vector matroids Given an $m \times n$ matrix A over a field F, the *vector matroid* $M[A]$ is obtained by using the labels of the columns $\mathbf{n}$ as ground set E and the collection of linearly independent column sets as I. This is evidently a matroid.

A matroid M that is isomorphic to $M[A]$ for some matrix A over field F is called F-*representable*. A matroid representable by a Galois Field $\mathbb{Z}_p$ is said to have the *arity* p. In particular, a matroid representable by $\mathbb{Z}_2$ is a *binary* matroid.

14.6 Graphic matroids Consider an undirected graph $G = (V, E)$. Let I denote the collection of all paths in the graph that do not contain a cycle. Then (E, I) is called the *cyclic matroid* of G, denoted $M(G)$. A matroid representable as the cyclic matroid of a graph is called a *graphic matroid*.

14.7 Graph-thereotic matroids A matroid (X, I) can be viewed as an undirected hypergraph. The nodes are the elements of X and the subsets in I are the hyperedges.

A useful special case is the restriction to (binary) graphs. Here the edges are just the two-element subsets in I. A *circuit* in the graph is a path from a vertex to itself without passing through any vertex twice. It represents a maximal independent subset.

[29] Nees to be proved.

15 Relations

15.1 Green's relations The relation $\mathcal{L} : A \to A$ on a semigroup A is defined by $a \left[\mathcal{L}\right] b$ iff a and b generate the same principal left ideal. Since the principal left ideal of a is $A^{\cdot}a$ this means that $A^{\cdot}a = A^{\cdot}b$. Dually, the relation $\mathcal{R} : A \leftrightarrow A$ is defined as having the same principal right ideal.

Suppose $a \left[\mathcal{L}\right] b$, i.e., $A^{\cdot}a = A^{\cdot}b$. Since $a \in A^{\cdot}a$, we must have $a \in A^{\cdot}b$, i.e., there must be $y \in A^{\cdot}$ such that $a = yb$. Similarly, there must be $x \in A^{\cdot}$ such that $b = xa$. These two properties completely capture $\mathcal{L}$, i.e., $a \left[\mathcal{L}\right] b$ *iff there are* $x, y \in A^{\cdot}$ *such that* $b = xa$ *and* $a = yb$. To see the reverse implication, assume the hypotheses hold.

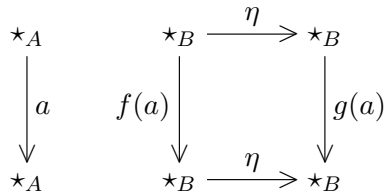
16 Categorification

The term "categorification" refers to the restatement of mathematics by replacing sets by categories. In the process, concepts that are identical become isomorphic.

16.1 Monoid as a category A monoid A may be viewed as a category with a unique object, which may be denoted $\star$. Arrows $a : \star \to \star$ are the elements of the monoid, which are closed under composition and have a unit (the identity arrow). We write A_1 for the set of monoid elements. Note that the dual of A (A^{op}) in the sense of a monoid is the same as its dual in the sense of a category.

A monoid homomorphism $f : A \to B$ is precisely a functor, because it preserves the composition and the identity arrow. (We write the unique objects in the categorified monoids as $\star_A$, $\star_B$, etc.)

A natural transformation $\eta : f \dot{\to} g : A \to B$ is an arrow in B (element of the monoid B_1) such that, for all elements a of A_1,



The naturality condition is $g(a) \cdot \eta = \eta \cdot f(a)$ in the monoid B. Does this correspond to anything in traditional semigroup theory? Probably not. They are mentioned in [Street, 2007, Ch. 15] as 2-cells "for the first time" in a "text at this level."

As an example, consider the monoid $T(X)$, the set of endomorphisms on a set X. We have two homomorphisms $i, j : T(X) \to T(X \times X)$ given by $i(a) = a \times 1_X$ and $j(a) = 1_X \times a$. A natural transformation $\eta : i \dot{\to} j$ is an element $\eta \in T(X \times X)$ satisfying $j(a) \cdot \eta = \eta \cdot i(a)$ for all $a \in T(X)$, i.e., $(1_X \times a) \cdot \eta = \eta \cdot (a \times 1_X)$. The only such natural transformation is the permutation $\sigma(x, y) \mapsto (y, x)$. We have $(1_X \times a) \cdot \sigma = \sigma \cdot (a \times 1_X)$.

16.2 Semiring as a semiadditive category A *semiadditive category* is a category enriched in **CMon** (which is a monoidal category). That means that $\mathrm{Hom}(A, B)$ has the structure of a commutative monoid and the composition $\mathrm{Hom}(A, B) \times \mathrm{Hom}(B, C) \to \mathrm{Hom}(A, C)$ is a bihomomorphism.

A semiring R may be viewed as a semiadditive category with a unique object $\star$. Arrows $a : \star \to \star$ are the elements of the semiring along with their commutative monoid structure. Composition of arrows is the multiplication of the semiring $a \circ b = ab$.

A left R-module is then semiadditive functor $h : R \to \mathbf{CMon}$. A right R-module is a contravariant semiadditive functor.

Rings are similarly viewed as one object *preadditive categories*, which are categories enriched in **Ab**. (An *additive category* is a preadditive category that has all finite biproducts $A_1 \oplus \cdots \oplus A_n$.)

16.3 Presheaves of monoids are modules A functor $h : A \to \mathbf{Set}$ picks out a set $X = h(\star_A)$ and maps each monoid element a of A_1 to an endomorphism $h_a = h(a) : X \to X$. Hence it gives a representation of A in $\mathrm{End}(X)$, which may also be viewed as a left-action

of A on X. We write these left-actions as $(X,\, h : A \to \mathrm{End}(X))$. Natural transformations $\eta : (X,h) \stackrel{\cdot}{\to} (Y,k)$ are morphisms of modules because, for every element a of A_1, η must satisfy the commutative square:

$$\begin{array}{ccc} \star & X & \xrightarrow{\eta} Y \\ \downarrow a & \downarrow h_a & \downarrow k_a \\ \star & X & \xrightarrow{\eta} Y \end{array}$$

which is nothing but a homomorphism of left A-modules.

Dually a functor $h : A^{\mathrm{op}} \to \mathbf{Set}$ gives an anti-representation of A in $\mathrm{End}(X)^{\mathrm{op}}$ or a right-action of A on X.

Both representations and anti-representations are now *presheaf categories* and, hence, form cartesian closed categories and toposes! They can interpret simply typed lambda calculus and intuitionistic logic [Goldblatt, 1984, Lambek and Scott, 1986, Wells, 1976, Wraith, 1975]. (Beware that the exponent defined in [Goldblatt, 1984] seems buggy.)

The product of two A-modules $(X_1,h_1) \times (X_2,h_2)$ is given pointwise. So it uses the set product $(h_1 \times h_2)(\star) = X_1 \times X_2$ as the underlying set and $(h_1 \times h_2)(a) = h_1(a) \times h_2(a)$ as the action of a. This is the same as the direct product of modules.

The exponent of two representations $(X,h) \Rightarrow (Y,k)$ is "futuristic," where an arrow $m : \star \leftarrow \star$ denotes a transition to the "future." The underlying set for $(X,h) \Rightarrow (Y,k)$ is $(h \Rightarrow k)(\star) = \int_{m:\star\leftarrow\star}[X \to Y]$, which is the set of families $\{\, f_m : X \to Y \,\}_m$ that are natural in the sense that

$$\begin{array}{ccc} \star & X & \xrightarrow{f_m} Y \\ \downarrow a & \downarrow h_a & \downarrow k_a \\ \star & X & \xrightarrow{f_{am}} Y \end{array}$$

In textual notation, the naturality condition is $f_{am}(h_a(x)) = k_a(f_m(x))$. The action on monoid elements is $(h \Rightarrow k)_a : \{f_m\}_m \mapsto \{f_{ma}\}_m$.

We put this in the standard notation of modules. (Cf. [Lambek and Scott, 1986, Example 9.6].) If $\langle X,\cdot\rangle$ and $\langle Y,\cdot\rangle$ are left A-modules, then their exponent is an A-module $\langle X \Rightarrow Y,\, \bullet\rangle$ where $X \Rightarrow Y$ consists of *A-linear maps* $f : A \times X \to Y$ where $A \times X$ is the direct product of left A-modules and A is regarded as a left A-module by multiplication. Note that the naturality condition, $f(am,\, a \cdot x) = a \cdot f(m,x)$, amounts to just A-linearity: $f(a \cdot (m,x)) = a \cdot f(m,x)$. The action $\bullet$ is defined by $a \bullet f : (m,x) \mapsto f(ma,x)$.

If A is a group, a further simplification is possible [Johnstone, 2002, Example 2.1.4]. Since $f(a,x) = f(a \cdot 1_A,\, a \cdot a^{-1} \cdot x) = f(a \cdot (1_A,\, a^{-1} \cdot x)) = a \cdot f(1_A,\, a^{-1} \cdot x)$, and 1_A is fixed, we can take $X \Rightarrow Y$ to be just the set of all functions $g : X \to Y$, with the interpretation that $g(x) = f(1_A,\, x)$ or, equivalently, $f(a,x) = a \cdot g(a^{-1} \cdot x)$. The action $\bullet$ is defined by $a \bullet g : x \mapsto a \cdot g(a^{-1} \cdot x)$.

The left regular representation of A as an action on itself can be represented as the functor $\lambda : A \to \mathbf{Set}$ given by $\lambda(\star) = A(\star,\star)$ and $\lambda(a) = A(\star,a) = A_1$. So, λ picks out the set of monoid elements of A_1 as the carrier, and maps a monoid element a of A_1 to the left multiplication function $x \mapsto ax$.

16.4 Modules in general More generally, a functor $h : A \to \mathcal{C}$ to any category $\mathcal{C}$ gives a representation of the monoid A in $\mathrm{End}(X)$, where $X = h(\star)$. Or, it is a left action on X.

If A and $\mathcal{C}$ are semiadditive categories, with A having a single object, then A is a *semiring* and h represents a *semiring-module* (semimodule). If they are preadditive categories, with A having a single object, then A is a *ring* and h represents a *ring-module*. See [Mitchell, 1965] for a detailed coverage of rings and modules using this formalism.

The category of left A-modules in $\mathcal{C}$ is the functor category $[A, \mathcal{C}]$, which we also write as ${}^A\mathcal{C}$ instead of the old notation $\mathbf{Mod}_{\mathcal{C}}(A)$. An object of ${}^A\mathcal{C}$ is a functor $h : A \to \mathcal{C}$, viewed as a module $(X,\ h : A \to \mathrm{End}(X))$, and a morphism is a natural transformation $\eta : h \dot{\to} k$ between such functors, viewed as an A-linear map $\eta : (X, h) \to_A (Y, k)$.

Consider now the category ${}^B({}^A\mathcal{C}) = [B,\ [A, \mathcal{C}]] = [B,\ {}^A\mathcal{C}]$. A functor k of this type picks out a left A-module $k(\star_B) = (X,\ h\colon A \to \mathrm{End}(X))$ and maps every monoid element b of B_1 to an A-linear endomorphism $k_b : (X, h) \to_A (X, h)$. The A-linearity is nothing but naturality in A:

$$\begin{array}{ccc} \star_A & (X,h) & \xrightarrow{\quad k_b \quad} (X,h) \\ \Big\downarrow a & \Big\downarrow h_a & \qquad\qquad \Big\downarrow h_a \\ \star_A & (X,h) & \xrightarrow{\quad k_b \quad} (X,h) \end{array}$$

which amounts to the equation $x : X \vdash k_b(h_a(x)) = h_a(k_b(x))$, or, in the more traditional notation, $b(ax) = a(bx)$.

Since **Cat** is a cartesian closed category, the functor category $[B,\ [A, \mathcal{C}]]$ is the same as $[B \times A,\ \mathcal{C}] \cong [A \times B,\ \mathcal{C}] \cong [A,\ [B, \mathcal{C}]]$. In our notation, this means ${}^B({}^A\mathcal{C}) \cong {}^{A\times B}\mathcal{C} \cong {}^A({}^B\mathcal{C})$.

The category of right A-modules in $\mathcal{C}$ is the functor category $[A^{\mathrm{op}},\ \mathcal{C}]$, which we also write as $\mathcal{C}^B$.

An (A, B)-*bimodule* can be represented similarly as ${}^A\mathcal{C}^B = ({}^A\mathcal{C})^B = {}^A(\mathcal{C}^B)$. The homogeneity/naturality condition is the same as above, $x : X \vdash k_b(h_a(x)) = h_a(k_b(x))$. But, in the traditional notation, it is written as $(ax)b = a(xb)$. Cf. §7.1.

16.5 Group as Groupoid A *groupoid category* (or *groupoid* for short) is a category where every morphism is invertible, i.e., for every $m : x \to y$ there is an inverse $m^{-1} : y \to x$.

Every group A can be regarded as a groupoid with a single object $\star$ and the elements $a \in A$ regarded as morphisms $a : \star \to \star$. The multiplication becomes composition, the unit the identity arrow on $\star$, and inverses the inverses.

A group homomorphism $f : A \to B$ is precisely a functor, because it preserves the composition and the identity arrow. It then follows that it preserves the inverses.

A natural transformation $b : f \dot{\to} g : A \to B$ is a morphism in B (element of the group B_1) such that, for all elements a of A_1,

$$\begin{array}{ccc} \star_A & \star_B & \xrightarrow{\quad b \quad} \star_B \\ \Big\downarrow a & \Big\downarrow f(a) & \qquad\qquad \Big\downarrow g(a) \\ \star_A & \star_B & \xrightarrow{\quad b \quad} \star_B \end{array}$$

The naturality condition is $g(a)b = bf(a)$, which is equivalent to $g(a) = bf(a)b^{-1}$ or $g = C_b \circ f$,

where C_b is the conjugation automorphism. Thus, a natural transformation $f \dot{\to} g$ is a single conjugation isomorphism that uniformly maps the image of f to the image of g.

16.6 Action groupoids Given a group action $A \times X \to X$ we define its *action groupoid* $X//A$ as follows [Armstrong, 2007]. The objects are the elements of X. The morphisms are pairs $(a, x) : x \to ax$ where $a \in A$ and $x \in X$. The identity morphisms are pairs of the form $(1_A, x) : x \to x$. The inverse of (a, x) is $(a^{-1}, ax) : ax \to x$.

In effect, the "states" in the G-set X have become the objects of a category and the transformations have become morphisms.

The action groupoid is related to the quotient of the action (Cf. §7.19). The quotient X/A is nothing but the skeleton of $X//A$, where all isomorphic configurations are identified as a single "orbit". For this reason $X//A$ is also called a "weak quotient."

There is a functor $F : X//A \to A$ where A is regarded as one-object groupoid, which sends all objects of $X//A$ to the unique object $\star$. Note that F is a faithful functor because distinct morphisms in $X//A$ are still distinct in A.

In fact, any groupoid H with a faithful functor $H \to A$ is equivalent to the action groupoid $X//A$. (Needs proof.)

16.7 Semidirect product and the Grothendieck construction The semidirect product $A \rtimes_\phi B$ is defined using a monoid homomorphism $\phi : B \to \mathrm{End}(A)$. Treating the monoid B as a one-object category, ϕ may be seen as a functor $\phi : B \to \mathbf{Mon}$ with $\phi\star_B$ as the chosen monoid A and $\phi(b : \star_B \to \star_B)$ as an endomorphism on A.

The semidirect product $A \rtimes_\phi B$ is then a monoid with a canonical object $(\star_{\phi\star_B}, \star_B)$, morphisms $(a, b) \in A \times B$, where $b : \star_B \to \star_B$ and $a : \star_{\phi\star_B} \to \star_{\phi\star_B}$. Composition is defined by:

$$(a_2, b_2) \circ (a_1, b_1) = (a_2 \circ \phi b_2(a_1),\ b_2 \circ b_1)$$

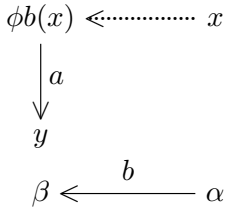
We might write $A \rtimes_\phi B$ as $\phi \times B$ since $A = \phi\star_B$ is automatically determined by ϕ.

We can generalize the construction to categories by taking ϕ to be a functor $\phi : \mathcal{B} \to \mathbf{Cat}$, i.e., a strict indexed category, indexed by a category $\mathcal{B}$. For every object α of $\mathcal{B}$, $\phi\alpha$ is a category and, for every morphism $b : \alpha \to \beta$ in $\mathcal{B}$, $\phi b : \phi\alpha \to \phi\beta$ is a functor. We think of the category $\phi\alpha$ as the "fibre" over α, with its morphisms regarded as "vertical morphisms." The functor ϕb then extends the morphism $b : \alpha \to \beta$ to a functor from the fibre $\phi\alpha$ to the fibre $\phi\beta$, giving an "image" of $\phi\alpha$ inside the fibre $\phi\beta$. Thus an object x in the fibre $\phi\alpha$ has an image $\phi b(x)$ in $\phi\beta$, and a vertical morphism $f : x \to y$ has an image $\phi b(f) : \phi b(x) \to \phi b(y)$, a vertical morphism in the fibre $\phi\beta$.

The equivalent of semidirect product is the *total category* $\int^{\mathcal{B}} \phi$ under *Grothendieck construction*, also called a *crossed product* $\phi \times \mathcal{B}$ by Ehresmann. It is given by the data:

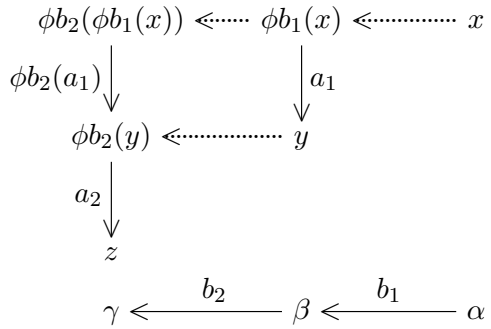
- objects, pairs (x, α) where $\alpha \in \mathrm{Ob}\,\mathcal{B}$ and $x \in \mathrm{Ob}(\phi\alpha)$, and
- morphisms, pairs $(a, b) : (x, \alpha) \to (y, \beta)$ where $b : \alpha \to \beta$ is a morphism of $\mathcal{B}$ and $a : \phi b(x) \to y$ is a vertical morphism of $\phi\beta$. (Note that $\phi b(x)$ is an object in the fibre $\phi\beta$ and so is y.)

Diagrammatically:



Composition of $(a_1, b_1) : (x, \alpha) \to (y, \beta)$ and $(a_2, b_2) : (y, \beta) \to (z, \gamma)$ is defined by

$$(a_2, b_2) \circ (a_1, b_1) = (a_2 \circ \phi b_2(a_1),\ b_2 \circ b_1)$$



The identity arrows of $\phi \times \mathcal{B}$ are $(1_{\phi 1_\alpha(x)},\ 1_\alpha) = (1_x,\ 1_\alpha) : (x, \alpha) \to (x, \alpha)$.

In comparing this with the semidirect product $A \rtimes_\phi B$, we should note that there is no "A" in the categorical generalization because each object of $\mathcal{B}$ has a corresponding "A," *viz.*, the fibre $\phi\alpha$.

There is a forgetful functor $\phi \times \mathcal{B} \to \mathcal{B}$ sending objects (x, α) to α and morphisms (a, b) to b. This functor is an *opfibration*, in fact, a split opfibration. If $b : \alpha \to \beta$ in $\mathcal{B}$ then an opcartesian arrow is $(1_{\phi b(x)}, b) : (x, \alpha) \to (\phi b(x), \beta)$. Notice that $1_{\phi b(x)}$ is a vertical morphism in $\phi\beta$, as required.

17 Monoidal categories

17.1 Strict monoidal category A monoid in **Cat** is called a *strict monoidal category*. It is a category $\mathcal{C}$ equipped with a bifunctor $\otimes : \mathcal{C} \times \mathcal{C} \to \mathcal{C}$ and an object I such that

$$\begin{aligned} A \otimes (B \otimes C) &= (A \otimes B) \otimes C \\ A \otimes I &= A = I \otimes A \end{aligned} \tag{17.1}$$

A *monoidal functor* $F : (\mathcal{C}, \otimes, I) \to (\mathcal{D}, \otimes', I')$ is a functor $F : \mathcal{C} \to \mathcal{D}$ along with a natural transformation $\eta_{A,B} : FA \otimes' FB \to F(A \otimes B)$ and an arrow $\eta_I : I' \to FI$ satisfying certain axioms. We use a linear term calculus notation to express morphisms composed of $\eta_{A,B}$ and η_I:

$$\frac{\Gamma \vdash \mathbf{t} : FA \quad \Delta \vdash \mathbf{u} : FB}{\Gamma, \Delta \vdash \mathbf{t} \cdot \mathbf{u} : F(A \otimes B)} \qquad \eta_{A,B} : FA \otimes' FB \to F(A \otimes B)$$

$$\frac{}{\vdash 1 : FI} \qquad \eta_I : I' \to FI$$

(Note that all terms denote morphisms in $\mathcal{D}$.) The axioms for monoidal functors are:

$$\begin{aligned} x : FA,\, y : FB,\, z : FC \;&\vdash\; x \cdot (y \cdot z) = (x \cdot y) \cdot z : F(A \otimes B \otimes C) \\ x : FA \;&\vdash\; x \cdot 1 = x : FA \\ y : FB \;&\vdash\; 1 \cdot y = y : FB \end{aligned}$$

If $F = (F, \cdot, 1_F)$ and $G = (G, \bullet, 1_G)$ are two such monoidal functors, a *monoidal natural transformation* $\varphi : F \Rightarrow G$ is a natural transformation $\varphi : F \dot{\to} G$ satisfying:

$$\begin{aligned} x : FA,\, y : FB \;&\vdash\; \varphi_A(x) \bullet \varphi_B(y) = \varphi_{A \otimes B}(x \cdot y) \;:\; G(A \otimes B) \\ &\vdash\; \varphi_I(1_F) = 1_G \;:\; GI \end{aligned}$$

A monoidal functor is said to be *strong* if $\eta_{A,B}$ and η_I are isomorphisms. It is said to be *strict* if they are identities, i.e., if $FA \otimes' FB = F(A \otimes B)$ and $I' = FI$.

17.2 Monoids as discrete monoidal categories A *discrete* category is one whose only morphisms are the identity morphisms. Since the morphisms add no value, we can ignore them and regard such a category as just a (large) set. Thus **Set** is a subcategory of **Cat**.

Inverting the generalization in the previous paragraph, we note that a strict monoidal category internal to **Set** is precisely a monoid. In other words, a monoid is nothing but a *discrete monoidal category*. (We can drop the requirement of "strict" in this context.) In the absence of nontrivial morphisms, the objects of the category can be regarded as just elements $a, b, c, \ldots$ The bifunctor $\otimes$ becomes just a binary operation and the tensor unit I is a unit element. The equations (17.1) are then precisely the equational axioms of monoids.

A monoidal functor $f : (\mathcal{C}, \otimes, I) \to (\mathcal{D}, \otimes', I')$ is precisely a monoid morphism because the requirement of $\eta_{a,b}$ forces $f(a) \otimes' f(b)$ to be the same as $f(a \otimes b)$ and that of η_I forces $I' = f(I)$.

17.3 Strict monoidal categories as 2-categories Just as a monoid may be viewed as one-object category under categorification (§16.1), a strict monoidal category may be viewed as one-object 2-category. The 2-category has a unique object $\star$. Each object A of the strict monoidal category is viewed as a 1-cell $A : \star \to \star$ with the tensor product $A \otimes B$ serving as the composition BA and the tensor unit I serving as the identity 1-cell, $I = 1_\star : \star \to \star$. The 2-cells are just the morphisms $f : A \to B$ of the strict monoidal category and the interchange law says that "$\otimes$" is a bifunctor.

17.4 Monoidal categories A *monoidal category* weakens the conditions of the strict monoidal category, replacing the equations with coherent isomorphisms:

$$\alpha_{A,B,C} : A \otimes (B \otimes C) \cong (A \otimes B) \otimes C$$
$$\lambda_A : A \otimes I \cong A$$
$$\rho_A : I \otimes A \cong A$$

The isomorphisms involved must be coherent by satisfying a number of equations [Mac Lane, 1991]. The definitions of monoidal functors and monoidal natural transformations are then adjusted to use the coherent isomorphisms for mediation. Monoidal categories, monoidal functors and monoidal natural transformations form a two category **MonCAT**.

17.5 Category actions The *action* of a category $\mathcal{C}$ on a category $\mathcal{X}$ is a functor:

$$\oslash : \mathcal{C} \times \mathcal{X} \to \mathcal{X}$$

A *strong functor* between two categories with $\mathcal{C}$-actions is a functor $F : \mathcal{X} \to \mathcal{Y}$ together with a natural transformation called *strength*:

$$\theta_{A,X} : A \oslash' F(X) \to F(A \oslash X)$$

Once again, we use a linear term calculus with zoned variable contexts to write terms involving strength:

$$\frac{\Sigma \mid \ \vdash \mathbf{a} : A \quad \mid \Gamma \vdash \mathbf{t} : FX}{\Sigma \mid \Gamma \vdash \mathbf{a} \bullet \mathbf{t} : F(A \oslash X)}$$

If $F = (F, \bullet)$ and $G = (G, \bullet')$ are two strong functors, a *strong natural transformation* $\varphi : F \Rightarrow G$ is a natural transformation $\varphi : F \stackrel{\cdot}{\to} G$ satisfying:

$$a : A \mid x : FX \vdash \varphi_{A \oslash X}(a \bullet x) = a \bullet' \varphi_X(x) : G(A \oslash X)$$

Categories with $\mathcal{C}$-actions, strong functors and strong natural transformations form a 2-category $\mathcal{C}$-**Act**.

17.6 Bicategories Bicategories are a typed" version of monoidal categories, where the objects of the monoidal category become morphisms $X : A \rightsquigarrow B$ of another category. The bifunctor of the monoidal category then becomes a form of composition that is "associative upto isomorphism" and the morphisms of the monoidal category $\rho : X \Rightarrow Y$ become 2-cells between these the one cells $X : A \rightsquigarrow B$ and $Y : A \rightsquigarrow B$. Thus, bicategories are a form of "weak 2-categories."

A bicategory $\mathcal{B}$ has objects (0-cells) A, B, ..., 1-cells X, Y, ... and 2-cells ρ, σ, ... For each pair of objects A, B, there is a category $[A, B]$ whose objects are the 1-cells $X : A \to B$ and morphisms are 2-cells $\rho : X \Rightarrow Y$. The composition of 1-cells is a bifunctor:

$$* : [A, B] \times [B, C] \to [A, C]$$

which we write in the diagrammatic order. It allows us to compose 1-cells $X * Y$ and 2-cells $\rho * \sigma$ in the usual way. There is an "identity" 1-cell for each object A, denoted 1_A.

The "associativity" of composition is given by a natural isomorphism α:

$$\begin{array}{ccccc}
[A,B] \times [B,C] \times [C,D] & \xrightarrow{* \times 1} & [A,C] \times [C,D] & \xrightarrow{\quad * \quad} & [A,D] \\
 & & \Downarrow \alpha & & \\
[A,B] \times [B,C] \times [C,D] & \xrightarrow{1 \times *} & [A,B] \times [B,D] & \xrightarrow{\quad * \quad} & [A,D]
\end{array}$$

(Note that α is natural in the *functors* inhabiting the categories $[A, B]$ etc.) The "identity" 1-cell similarly has natural isomorphisms

$$\lambda_X : X * 1_B \Rightarrow X \qquad \rho_X : 1_A * X \Rightarrow X$$

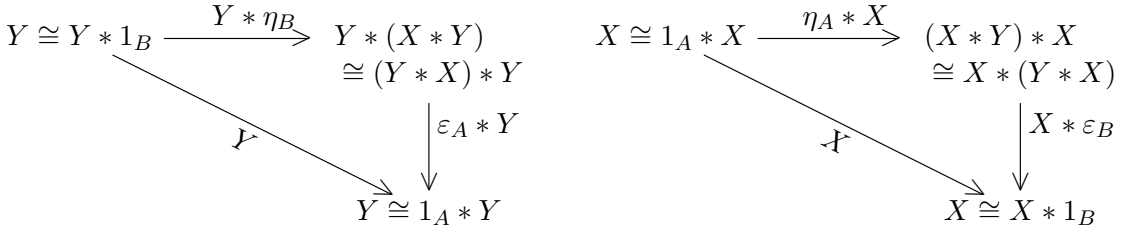
The isomorphisms must be coherent by satisfying two equations [Mac Lane, 1991, XII.6].

Bicategories differ from 2-categories only in that the "composition" of 1-cells is associative up to isomorphism (and "identity" 1-cells are identities up to isomorphism). The rest everything is the same. For this reason, bicategories are called "weak 2-categories." Conversely, 2-categories are "strict bicategories."

17.7 Monoidal categories as bicategories A one-object bicategory is nothing but a monoidal category (cf. §17.4). The unique object may be written as $\star$ and every object A of the monoidal category may be viewed as a 1-cell $A : \star \to \star$. The tensor product $\otimes$ is the composition of these 1-cells and the tensor unit I is the identity 1-cell $1_\star = I : \star \to \star$. The morphisms in the monoidal category $f : A \to B$ become the 2-cells in the bicategory. Contrast this with one-object 2-categories which are nothing but *strict* monoidal categories.

17.8 Bimodules form bicategories Monoids and bimodules (cf. §7.14) provide another example of bicategories. The objects are monoids A, B, ..., 1-cells are bimodules $X : A \rightsquigarrow B$ and 2-cells are bimodule homomorphisms $f : X \to Y$. The tensor product of bimodules $X \otimes_B Y : A \to C$ gives the composition of 1-cells and each monoid A serves as the identity 1-cell $A : A \rightsquigarrow A$.

17.9 Duals in bicategories Adjunctions internal to a bicategory, i.e., adjoint pairs of 1-cells, are referred to as duals. Two 1-cells $X : B \to A$ and $Y : A \to B$ form a dual pair $X \dashv Y$ if there are 2-cells $\eta_A : 1_A \to X * Y$ and $\varepsilon_B : Y * X \to 1_B$ (the "unit" and the "counit") such that the triangle identities hold:



Note that we are writing the horizontal composition in the diagrammatic order $X * Y$. The analogy with standard adjunctions in **Cat** may be seen by using juxtaposition instead: the unit and counit have types $\eta : 1 \to YX$ and $\varepsilon : XY \to 1$ and the triangle equalities say, in essence, that the composites $\eta Y; Y\varepsilon$ and $X\eta; \varepsilon X$ are identities.

If $X \dashv Y : A \to B$, we say that X is a *left dual* of Y and Y is a *right dual* of X. As in **Cat**, X and Y determine each other uniquely up to natural isomorphism. The left dual of Y is often written as Y^*.

17.10 Duals in monoidal categories The above definition immediately specializes to monoidal categories, which are bicategories with a single object. The unique object is thus ignored and the 1-cells are regarded as the objects in the monoidal category.

An object with a left dual (right dual) is said to be *left dualizable* (*right dualizable*), and an object with both duals is just *dualizable*.

A monoidal category (i.e., a one-object bicategory) in which all objects are left dualizable (right dualizable) is called a *left autonomous category* (*right autonomous category*). One in which all objects are both left and right dualizable is called an *autonomous category*. (Note that this terminology differs from Michael Barr's notion of "autonomous category" as a closed symmetric monoidal category.) An alternative terminology for autonomous categories is "*rigid* monoidal categories".

17.11 Duals in symmetric monoidal categories In a *symmetric* monoidal category, left and right duals are the same. So, a symmetric monoidal category is left autonomous if only if it is right autonomous, or just autonomous. A symmetric autonomous category is also called a *compact closed category*. The terminology is apt because such a category is always closed with the internal homs given by $X \multimap Y = X^* \otimes Y$.

17.12 Dualizing object If there is an object $\perp$ in a closed monoidal category such that the left dual of every object Y is expressible as $Y^* = Y \multimap \perp$, then $\perp$ is called a *dualizing object*.

18 Exactness

18.1 Kernel In **Grp**, the kernel of a morphism $f : A \to B$ is the inverse image of 1_B. Since the inverse image is a subgroup of A, write it as a morphism $k : N \to A$. The composite $k; f : N \to B$ sends every element of N to 1_B. In other words, it is the zero morphism $0 : N \to \mathbf{0} \to B$.

In general, in any category with a null object, the *kernel* of an arrow $f : A \to B$ is the universal arrow $k : N \to A$ such that $k; f = 0 : N \to B$. It is universal in the sense that any other arrow $h : X \to A$ such that $h; f = 0$ uniquely factors through k, i.e., $h = h'; k$ for a unique $h' : X \to N$.

A more direct way of describing the kernel of f is as the *equalizer* of $f : A \to B$ and $0 : A \to B$. Hence, any category with null objects and equalizers has kernels.

In **Mon**, every morphism $f : A \to B$ has a kernel, which selects the inverse image of 1_B. It is evidently a submonoid of A. In $\mathbf{CPO}_\perp$, the kernel of $f : A \to B$ is the inverse image of $\perp_B$.

Note that kernels are necessarily monics (obviously from the fact that they are equalizers). However, not all monics are kernels.

18.2 Cokernel By duality, the *cokernel* of $f : A \to B$ is a morphism $u : B \to Q$ that is universal among morphisms satisfying $f; u = 0 : A \to Q$. More simply, it is the coequalizer of $f : A \to B$ and $0 : A \to B$. Being a coequalizer, it is always an epi.

In **Ab**, the cokernel of $f : A \to B$ is the projection to the quotient $u : B \to B/\mathrm{Im}(f)$. In the **Grp**, the cokernel of $f : A \to B$ identifies the quotient $B/\mathrm{Im}^*(f)$, where $\mathrm{Im}^*(f)$ is the *conjugate closure* of the image of f.

19 Regular categories

In the following all categories will be finitely complete, i.e., have all finite limits.

19.1 Relations as jointly monic spans A relation in a finitely complete category is a pair of morphisms $X \xleftarrow{p_1} R \xrightarrow{p_2} Y$ such that $\langle p_1, p_2 \rangle : R \to X \times Y$ is a monic.

19.2 Congruence relations as internal equivalence relations An internal equivalence relation on an object X is a relation $X \xleftarrow{p_1} R \xrightarrow{p_2} X$ equipped with the following morphisms:

- *internal reflexivity*: a common section $r : X \to R$ for p_1 and p_2,
- *internal symmetry*: a morphism $s : R \to R$ which interchanges p_1 and p_2, i.e., $p_1 \circ s = p_2$ and $p_2 \circ s = p_1$, and
- *internal transitivity*: a morphism $t : R \times_X R \to R$

19.3 Kernel pair of a morphism The notion of a "kernel congruence" of a morphism (§2.21) generalizes to arbitrary categories as follows. If $h : A \to B$ is a morphism, then the pullback of h with itself:

$$\begin{array}{ccc} A \times_B A & \xrightarrow{p_1} & A \\ \downarrow{\scriptstyle p_2} & & \downarrow{\scriptstyle h} \\ A & \xrightarrow{h} & B \end{array}$$

gives a pair of morphisms $(p_1, p_2) : A \times_B A \to A$ which are called the *kernel pair* of h. Note that, in **SGrp**, this coincides with the notion of kernel in §2.21: $A \times_B A = \{\, (a, b) : h(a) = h(b) \,\}$, and p_1 and p_2 are the two projections.

A Normal submonoids

We embed monoid A in $\mathcal{R}_A = \mathbf{Rel}(|A|, |A|)$, the collection of binary relations over $|A|$, treated as an ordered monoid with relational composition RS as the multiplication and the identity relation I_A as the unit. The monoid $\mathcal{R}_A$ also has a monotone *converse* operation $R^\smile = \{\, (y, x) \mid (x, y) \in R \,\}$ satisfying $I_A \subseteq RR^\smile$, $I_A \subseteq R^\smile R$ and $(RS)^\smile = S^\smile R^\smile$.

For $a \in A$, the right multiplication operator is treated as a relation $\xrightarrow{a}_R$, its converse as $\xleftarrow{a}_R$ and the symmetric closure as $\xleftrightarrow{a}_R$;

$$\begin{array}{lcl} x \xrightarrow{a}_R y & \iff & ax = y \\ x \xleftarrow{a}_R y & \iff & y \xrightarrow{a}_R x \\ x \xleftrightarrow{a}_R y & \iff & x \xleftarrow{a}_R y \ \vee \ x \xrightarrow{a}_R y \end{array}$$

Similar notations are also used for left multiplication: $\xrightarrow{a}_L$, $\xleftarrow{a}_L$ and $\xleftrightarrow{a}_L$. However, we will deal almost exclusively with right multiplication relations in this section and, so, drop the subscript R for notational simplicity. By virtue of duality, the same conclusions also hold for left multiplication.

Note that $\xrightarrow{a}$ is a *many-to-one* relation, $\xleftarrow{a}$ is a *one-to-many* relation and $\xleftrightarrow{a}$ is a *many-to-many* relation.

All three relations are extended to submonoids $S \subseteq A$ in the obvious fashion:

$$x \xrightarrow{S} y \iff \exists a \in S.\ x \xrightarrow{a} y$$

These relations are *many-to-many.*

Since the elements of A are being embedded in $\mathcal{R}_A$, we have the faithfulness condition:

$$a = b \iff \xrightarrow{a} \ = \ \xrightarrow{b} \tag{A.1}$$

We can also express statements of the form $x \xrightarrow{S} y$ wholely in terms of relations:

$$\begin{array}{lcl} x \xrightarrow{a} y & \iff & \xrightarrow{y} \ = \ \xrightarrow{x}\xrightarrow{a} \\ x \xrightarrow{S} y & \iff & \xrightarrow{y} \ \subseteq \ \xrightarrow{x}\xrightarrow{S} \\ x \xleftarrow{a} y & \iff & \xrightarrow{x} \ = \ \xrightarrow{y}\xrightarrow{a} \\ x \xleftarrow{S} y & \iff & \xrightarrow{x} \ \subseteq \ \xrightarrow{y}\xrightarrow{S} \end{array} \tag{A.2}$$

We obtain the following "cancellation" laws:

$$\begin{array}{lcl} \xleftarrow{a}\xrightarrow{ab} & \subseteq & \xrightarrow{b} \\ \xrightarrow{ab}\xleftarrow{b} & \supseteq & \xrightarrow{a} \\ \xleftarrow{ab}\xrightarrow{a} & \subseteq & \xleftarrow{b} \\ \xrightarrow{b}\xleftarrow{ab} & \supseteq & \xleftarrow{a} \end{array} \tag{A.3}$$

For the first formula, the left hand side relation has pairs of the form (xa, xab) for all $x \in A$, which are clearly included in $\xrightarrow{b}$. For the second formula, the left hand side relation contains pairs (x, xa), among others, and so includes $\xrightarrow{a}$. The next two formulas are obtained from the first two by taking converses.

If $N \subseteq A$ is a normal submonoid of A then $aN = Na$ for all $a \in A$. Therefore:

$$\xrightarrow{a}\xrightarrow{N} \ = \ \xrightarrow{N}\xrightarrow{a}$$

Lemma 1 *Let $N \subseteq A$ be a normal submonoid. If $a \sim_N a'$ and $ab \sim_N a'b'$ then $b \sim_N b'$.*

Proof: The assumptions $a \sim_N a'$ and $ab \sim_N a'b'$ can be written as:

$$\begin{array}{lcl} \xrightarrow{a'} & \subseteq & \xrightarrow{a}\xleftarrow{N}^* \\ \xrightarrow{a'b'} & \subseteq & \xrightarrow{ab}\xleftarrow{N}^* \end{array}$$

By composing both sides of the second inclusion formula with $\xleftarrow{a'}$ we obtain:

$$\begin{array}{lcll} \xleftarrow{a'}\xrightarrow{a'b'} & \subseteq & \xleftarrow{a}\xleftarrow{N}^*\xrightarrow{ab}\xleftarrow{N}^* & \\ & = & \xleftarrow{a}\xrightarrow{ab}\xleftarrow{N}^*\xleftarrow{N}^* & \text{since } N \text{ is a normal submonoid} \\ & = & \xleftarrow{a}\xrightarrow{ab}\xleftarrow{N}^* & \end{array}$$

The relation $\xleftarrow{a'}\xrightarrow{a'b'}$ relates pairs $(xa', xa'b')$ for all $x \in A$ with the middle term being x. For the same middle term x, the relation $\xleftarrow{a}\xrightarrow{ab}$ relates the pair (xa, xab).

References

[Abramsky and Jung, 1994] Abramsky, S. and Jung, A. (1994). Domain theory. In Abramsky, S., Gabbay, D. M., and Maibaum, T. S. E., editors, *Handbook of Logic in Computer Science*, volume 3, pages 1–168. Clarendon Press, Oxford.

[Aluffi, 2009] Aluffi, P. (2009). *Algebra: Chapter 0*, volume 104 of *Graduate Studies in Math.* Amer. Math. Soc.

[Armstrong, 2007] Armstrong, J. (2007). Groupoids (and more group actions). In *Unapologetic Mathematician*. wordpress.com.

[Borceux and Janelidze, 2001] Borceux, F. and Janelidze, G. (2001). *Galois Theories*. Cambridge Univ. Press.

[Clifford and Preston, 1961] Clifford, A. H. and Preston, G. B. (1961). *The Algebraic Theory of Semigroups*. AMS.

[Cohn, 1982] Cohn, P. M. (1982). *Algebra, Volume 1*. John Wiley, second edition.

[Cox, 2004] Cox, D. A. (2004). *Galois Theory*. Wiley Interscience.

[Dixmier, 1984] Dixmier, J. (1984). *General Topology*. Springer-Verlag.

[Dubuc, 2003] Dubuc, E. J. (2003). Localic galois theory. *Advances in Mathematics*, 175(1):144–167.

[Dubuc and de la Vega, 2000] Dubuc, E. J. and de la Vega, C. S. (2000). On the Galois theory of Grothendieck. Manuscript on arXiv.

[Eckmann and Hilton, 1962] Eckmann, B. and Hilton, P. J. (1962). Group-like structures in general categories I: Multiplication and comultiplication. *Math. Ann.*, 145:227–255.

[Edwards, 1984] Edwards, H. M. (1984). *Galois theory*. Springer-Verlag.

[Eilenberg, 1974] Eilenberg, S. (1974). *Automata, Languages, and Machines*. Academic Press. (Volumes A and B).

[Eilenberg, 1976] Eilenberg, S. (1976). *Automata, Languages, and Machines; Vol. B*. Academic Press.

[Escofier, 2001] Escofier, J.-P. (2001). *Galois Theory*. Springer-Verlag.

[Gierz et al., 2003] Gierz, G., Hoffmann, K. H., Keimel, K., Lawson, J. D., Mislove, M. W., and Scott, D. S. (2003). *Continuous Lattices and Domains*. Cambridge Univ. Press.

[Ginzburg, 1968] Ginzburg, A. (1968). *Algebraic Theory of Automata*. Academic Press, New York.

[Ginzburg and Yoeli, 1965] Ginzburg, A. and Yoeli, M. (1965). Products of automata and the problem of covering. *Trans. Amer. Math. Soc*, 116:253–266.

[Golan, 1999] Golan, J. S. (1999). *Semirings and their Applications*. Kluwer.

[Goldblatt, 1984] Goldblatt, R. (1984). *Topoi, the Categorical Analysis of Logic*. North-Holland. available online from *Historical Math Monographs*, Cornell University.

[Grassmann, 1979] Grassmann, H. (1979). On factor S-sets of monoids modulo submonoids. *Semigroup Forum*, 19(1):163–172.

[Gray and Ruškuc, 2008] Gray, R. and Ruškuc, N. (2008). Green index and finiteness conditions for semigroups. *J. Algebra*, 320:3145–3164.

[Grillet, 1995] Grillet, P. A. (1995). *Semigroups: An Introduction to the Structure Theory*. Marcel Dekker, New York.

[Hartmanis and Stearns, 1966] Hartmanis, J. and Stearns, R. E. (1966). *Algebraic Structure Theory of Sequential Machines*. Prentice-Hall.

[Hoare et al., 2011a] Hoare, C. A. R., Hussain, A., Moller, B., O'Hearn, P. W., Petersen, R. L., and Struth, G. (2011a). On locality and the exchange law for concurrent processes. In *CONCUR 2011*, pages 250–264. Springer-Verlag.

[Hoare et al., 2011b] Hoare, C. A. R., Moller, B., Struth, G., and Wehrman, I. (2011b). Concurrent Kleene algebra and its foundations. *J. Logic and Algeb. Program.*, 80(6):266–296.

[Holcombe, 1982] Holcombe, W. M. L. (1982). *Algebraic Automata Theory*. Cambridge Studies in Advanced Mathematics. Cambridge Univ. Press, Cambridge.

[Howie, 1976] Howie, J. M. (1976). *An Introduction to Semigroup theory*. Academic Press.

[Johnstone, 1982] Johnstone, P. T. (1982). *Stone Spaces*, volume 3 of *Cambridge Studies in Advanced Mathematics*. Cambridge Univ. Press.

[Johnstone, 2002] Johnstone, P. T. (2002). *Sketches of an Elephant: A Topos Theory Compendium (two volumes)*. Clarendon Press.

[Kilp et al., 2000] Kilp, M., Knauer, U., and Mikhalev, A. V. (2000). *Monoids, Acts and Categories with Applications to Wreath Products and Graphs*, volume 29 of *Expositions in Mathematics*. Walter de Gruyter, Berlin.

[Kock, 2007] Kock, J. (2007). Note on commutativity in double semigroups and two-fold monoidal categories. *J. Homotopy Relat. Struct.*, 2(2):217–228.

[Lambek and Scott, 1986] Lambek, J. and Scott, P. J. (1986). *Introduction to Higher Order Categorical Logic*. Cambridge Univ. Press, Cambridge.

[Lang, 2000] Lang, S. (2000). *Algebra*. Springer-Verlag.

[Ljapin, 1974] Ljapin, E. S. (1974). *Semigroups*, volume 3 of *Translations of Mathematical Monographs*. AMS.

[Mac Lane, 1991] Mac Lane, S. (1991). *Categories for the Working Mathematician*. Springer-Verlag, second edition.

[Mac Lane and Birkhoff, 1967] Mac Lane, S. and Birkhoff, G. (1967). *Algebra*. Chelsea, New York.

[Mitchell, 1965] Mitchell, B. (1965). *Theory of Categories*, volume 17 of *Pure and Appl. Math.* Academic Press.

[Rhodes and Steinberg, 2009] Rhodes, J. and Steinberg, B. (2009). *The q-theory of Finite Semigroups*. Springer-Verlag.

[Robalo, 2009] Robalo, M. A. D. (2009). Galois theory towards Dessins d'Enfants. Master's thesis, Instituto Superior Técnico, Universidade Técnica de Lisboa.

[Rotman, 1965] Rotman, J. J. (1965). *An Introduction to the Theory of Groups*. Allyn and Bacon, Boston.

[Sangiorgi, 2009] Sangiorgi, D. (2009). On the origins of bisimulation and coinduction. *ACM Trans. Program. Lang. Syst.*, 31(4):15.

[Stewart, 2004] Stewart, I. (2004). *Galois Theory*. Chapman & Hall/CRC.

[Straubing, 1989] Straubing, H. (1989). The wreath product and its applications. In Pin, J. E., editor, *Formal Properties of Finite Automata and Applications*, volume 386 of *LNCS*, pages 15–24. Springer-Verlag, Berlin.

[Street, 2007] Street, R. (2007). *Quantum Groups: A Path to Current Algebra*, volume 19 of *Australian Math. Soc. Lecture Series*. Cambridge Univ. Press.

[Swallow, 2004] Swallow, J. (2004). *Exploratory Galois Theory*. Cambridge Univ. Press.

[Szamuely, 2009] Szamuely, T. (2009). *Galois Groups and Fundamental Groups*. Cambridge Univ. Press.

[Tonini, 2009] Tonini, F. (2009). Notes on Grothendieck-Galois theory. Electronic manuscript.

[van der Waerden, 1949] van der Waerden, B. L. (1949). *Modern Algebra*. Unger, New York, second edition. (Translated from German by Fred Blum, original version 1930-31).

[Wells, 1976] Wells, C. (1976). Some applications of the wreath product construction. *Amer. Math. Monthly*, 83(5):317–338.

[Whitney, 1935] Whitney, H. (1935). On the abstract properties of linear dependence. *Amer. J. Math.*, 57(3):509–533.

[Willard, 1970] Willard, S. (1970). *General Topology*. Addison-Wesley.

[Wraith, 1975] Wraith, G. C. (1975). Lectures on elementary topoi. In Maurer, C., Lawvere, F. W., and Wraith, G. C., editors, *Model Theory and Topoi*, volume 445 of *Lect. Notes Math.*, pages 114–206. Springer-Verlag.

[Yeh, 1968] Yeh, R. T. (1968). On relational homomorphisms of automata. *Inf. Control*, 13:140–155. Reviewed in IEEE Trans. Comp., Oct 1970.

[Yeh, 1970] Yeh, R. T. (1970). Structural equivalence of automata. *Math. Systems Theory*, 4(2):198–211.