

ISSN: 1401-5617

Wheels

—
On Division by Zero

Jesper Carlström

RESEARCH REPORTS IN MATHEMATICS
NUMBER 11, 2001

DEPARTMENT OF MATHEMATICS
STOCKHOLM UNIVERSITY

Electronic versions of this document are available at
<http://www.matematik.su.se/reports/2001/11>

Date of publication: September 10, 2001

2000 Mathematics Subject Classification: Primary 16Y99, Secondary 13B30, 13P99,
03F65, 08A70.

Keywords: fractions, quotients, localization.

Postal address:

Department of Mathematics
Stockholm University
S-106 91 Stockholm
Sweden

Electronic addresses:

<http://www.matematik.su.se>
info@matematik.su.se

WHEELS

—

ON DIVISION BY ZERO

Jesper Carlström

Department of Mathematics

Stockholm University

<http://www.matematik.su.se/~jesper/>

Filosofie licentiatavhandling

Abstract

We show how to extend any commutative ring (or semiring) so that division by any element, including 0, is in a sense possible. The resulting structure is what is called a *wheel*. Wheels are similar to rings, but $0x = 0$ does not hold in general; the subset $\{x \mid 0x = 0\}$ of any wheel is a commutative ring (or semiring) and any commutative ring (or semiring) with identity can be described as such a subset of a wheel.

The main goal of this paper is to show that the given axioms for wheels are natural and to clarify how valid identities for wheels relate to valid identities for commutative rings and semirings.

Contents

1	Introduction	3
1.1	Why invent the wheel?	3
1.2	A sketch	4
2	Involution-monoids	7
2.1	Definitions and examples	8
2.2	The construction of involution-monoids from commutative monoids	9
2.3	Insertion of the parent monoid	11
2.4	The role of X in the structure of M_X^*	12
2.5	The construction as a functor	14
3	Applications to semirings	15
3.1	Applications to additive monoids of semirings	16
3.2	Applications to multiplicative monoids of commutative semirings	19
4	Wheels	23
4.1	Basic properties	24
4.2	Applications to wheels of fractions	27
4.2.1	The semiring in a wheel of fractions	27
4.2.2	Obtaining $/$ -invertible wheels of fractions	28
4.2.3	Axiomatizing the class of wheels of fractions	29
4.2.4	Cancellation of zero-terms	30
4.3	Equational logic for wheels of fractions	32
5	Wheel-modules	44

Acknowledgements

I would like to thank Clas L ofwall for his support during this work. He followed the process through all of its stages, listening to my vague ideas and reading many preliminary versions of the manuscript. He put me in the right direction many times. If anything is easy to grasp in this paper, it is probably because he asked me to simplify that part and to give some examples.

I would also like to thank Per Martin-L of, who spent many hours discussing constructive aspects with me. He also read several versions of the manuscript and gave me many valuable comments, on the content as well as on how it should be expressed. The first idea, of which wheel theory is a generalization, originates from him.

Thanks also to J orgen Backelin, with whom I discussed possible choices of axioms, before the final selection was made.

1 Introduction

1.1 Why invent the wheel?

The fact that multiplicative inversion of real numbers is a partial function is annoying for any beginner in mathematics: “Who has forbidden division by zero?”. This problem seems not to be a serious one from a professional point of view, but the situation remains as an unaesthetic fact. We know how to extend the semiring of natural numbers so that we get solutions to equations like $5 + x = 2$, $2x = 3$, $x^2 = 2$, $x^2 = -1$ etc., we even know how to get completeness, but not how to divide by 0.

There are also concrete, pragmatic aspects of this problem, especially in connection with exact computations with real numbers. Since it is not in general decidable whether a real number is non-zero, one can’t in general tell whether it is invertible. Edalat and Potts [EP00, Pot98] suggested that two extra ‘numbers’, $\infty = 1/0$ and $\perp = 0/0$, be adjoined to the set of real numbers (thus obtaining what in domain theory is called the ‘lifting’ of the real projective line) in order to make division always possible. In a seminar, Martin-Löf proposed that one should try to include these ‘numbers’ already in the construction of the rationals from the integers, by allowing not only non-zero denominators, but arbitrary denominators, thus ending up not with a *field*, but with a field with two extra elements. Such structures were called ‘wheels’ (the term inspired by the topological picture \odot of the projective line together with an extra point $0/0$) by Setzer [Set97], who showed how to modify the construction of fields of fractions from integral domains so that wheels are obtained instead of fields.

In this paper, we generalize Setzer’s construction, so that it applies not only to integral domains, but to any commutative semiring. Our construction introduces a ‘reciprocal’ to every element, the resulting structure being what will be called a ‘wheel of fractions’. We use the term ‘wheel’ in a more general sense than Setzer, but a wheel in our sense is still a structure in which addition, multiplication and division can always be performed. A wheel in Setzer’s sense will be recognized as what we denote by $\odot_{S_0} A$, where A is an integral domain, S_0 the subset $A \setminus \{0\}$.¹

Beside applications to exact computations, there are other applications to computer programming: algorithms that split into cases depending on whether their arguments are zero or not, can sometimes be simplified using a total division function.

We also hope for applications in algebraic geometry. Instead of considering ‘rational maps’ as partial functions from a variety to a field, one may consider ‘rational functions’ from the variety to the wheel extending the field. The set of all such rational functions is a wheel in our sense (but not in Setzer’s) with the operations defined point-wise. We say that rational functions are ‘equal as maps’ if they agree on some non-empty Zariski open subset. This is a congruence relation (it is transitive since the Zariski topology of a variety is irreducible) and the set of congruence classes is again a wheel — the ‘wheel of rational maps’. In the same way, there is a ‘wheel of regular functions’. Every regular function is equal as a map to some rational function. Finally, the most important

¹The notation S_0 will be used for the set of cancellable elements in a monoid, i.e., $a \in S_0$ means that $ax = ay \Rightarrow x = y$ for all x, y . When considering semirings, we use the multiplicative monoid for this definition. In integral domains, we get $S_0 = A \setminus \{0\}$.

methods in algebraic geometry are the projection onto a quotient of the ring of functions considered (the geometric interpretation being restricting attention to a Zariski closed subset of the variety) and the construction of a ring of fractions (localization in the sense that attention is restricted to what happens *near* some Zariski closed subset). Usually, those constructions are formally very different, even though their interpretations are similar. Using wheels instead of rings, those two situations are both handled by the same method: by projection onto a quotient. That suggests possible simplifications when wheels are used instead of rings.²

Finally, there should be applications to the development of constructive mathematics, in which total functions are preferred to partial functions.³

1.2 A sketch

We indicate briefly what the basic ideas are. Proofs and technical verifications are postponed to the following sections.

The natural way of trying to introduce inverses to all elements of a ring, is to modify the usual construction of rings of fractions. If A is a commutative ring with identity, S a multiplicative submonoid of it, then the usual construction is as follows: Define the relation \sim_S on $A \times S$ (here the product is taken in the category of sets) as

$$(x, s) \sim_S (x', s') \quad \text{means} \quad \exists s'' \in S : s''(xs' - x's) = 0.$$

Then \sim_S is an equivalence relation and $A \times S / \sim_S$ is a commutative ring with (the class containing (x, y) is denoted by $[x, y]$)

$$\begin{aligned} 0 &= [0, 1] \\ 1 &= [1, 1] \\ [x, s] + [x', s'] &= [xs' + x's, ss'] \\ [x, s][x', s'] &= [xx', ss']. \end{aligned}$$

Clearly, 0 can't be inverted unless $0 \in S$, but if that is the case, then \sim_S is the improper relation, so that $A \times S / \sim_S$ is trivial.

The obvious thing to try is to replace $A \times S$ by $A \times A$. Considering the multiplicative structure only, $A \times S$ is a product of monoids and \sim_S a congruence relation on it. Let \equiv_S be the congruence relation that is generated on the monoid $A \times A$ by \sim_S . Then

$$(x, y) \equiv_S (x', y') \iff \exists s, s' \in S : (sx, sy) = (s'x', s'y').$$

²We do not treat algebraic geometry in this paper, but the statements in this paragraph are rather obvious corollaries of the theory developed. It is also possible to treat the more general situation of spectrums and schemes.

³That a function f (think of the inversion of real numbers) is 'partial' on a set A means that $f(x)$ is defined provided a certain predicate holds for x , say, that $P(x)$ is true. In constructive type theory [ML84], this is interpreted as that f takes pairs (x, p) as arguments, with x an element of A and p a proof of $P(x)$. Thus, formally, f is not at all in the usual sense a partial function, but total on a set of pairs. In that sense, there are no partial functions in constructive type theory, but the term 'partial' can still be useful as a way of explaining that one informally thinks of f as defined on some 'part' of A . Since a proof of $P(x)$ is needed as an argument for f , it is clearly preferable to have f replaced by a total function. That is a reason for asking for a program of reworking mathematics without partial functions. This paper is in the line with such a program.

We define $\odot_S A$ (the wheel of fractions with respect to S) as $A \times A / \equiv_S$ with the operations

$$\begin{aligned} 0 &= [0, 1] \\ 1 &= [1, 1] \\ [x, y] + [x', y'] &= [xy' + x'y, yy'] \\ [x, y][x', y'] &= [xx', yy'] \\ /[x, y] &= [y, x]. \end{aligned}$$

This structure is *not* a ring (unless it is trivial), since $0x = 0$ is not valid in general: with $x = [0, 0]$, we get $0x = [0, 0]$ which is not equal to $[0, 1]$ unless $0 \in S$, but then \equiv_S is improper and $\odot_S A$ is trivial.

The additive structure is a commutative monoid, as well as the multiplicative. However, the group structure of addition is destroyed, since $[0, 0] + x = 0$ has no solution in non-trivial cases. Instead, one has the formula $x - x = 0x^2$ if $x - y$ is defined as $x + (-1)y$, where $-1 = [-1, 1]$. Thus $x - x = 0$ is true for any x with $0x = 0$; and in many wheels, there are many such x 's.

The unary operation $/$ is an involution on the multiplicative monoid, i.e. $//x = x$ and $/(xy) = /y/x$. We call it the ‘reciprocal’ operation, with $/x$ being the ‘reciprocal’ of x . One does not have $x/x = 1$ in general, but $x/x = 1 + 0x/x$.⁴

Wheels of fractions are *wheels*, abstractly defined as follows.

Definition 1.1 (wheel). A wheel is a structure $\langle H, 0, 1, +, \cdot, / \rangle$ in which the following holds:

$$\langle H, 0, + \rangle \text{ is a commutative monoid} \tag{1}$$

$$\langle H, 1, \cdot, / \rangle \text{ is a commutative monoid with involution } / \tag{2}$$

Distributivity

$$(x + y)z + 0z = xz + yz \tag{3}$$

$$\frac{x}{y} + z + 0y = \frac{x + yz}{y} \tag{4}$$

Rules for zero-terms

$$0 \cdot 0 = 0 \tag{5}$$

$$(x + 0y)z = xz + 0y \tag{6}$$

$$/(x + 0y) = /x + 0y \tag{7}$$

$$x + 0/0 = 0/0. \tag{8}$$

Here, H is a set (we will use the same symbol for the wheel), 0 and 1 are constants, $+$ and \cdot are binary operations and $/$ is a unary operation. We often omit the dot for multiplication and we sometimes write $\frac{x}{y}$ for x/y . The usual priority rules apply: lower -arity gives higher priority and multiplication is prior to addition.

⁴That $x/x = 1$ is not in general true, is the reason why we avoid the notation x^{-1} for $/x$. The reciprocal should be thought of as a unary version of division, like negation is a unary version of subtraction. The unary negation corresponds to the binary subtraction by $x - y = x + (-y)$ and $-y = 0 - y$. In the same way, there is a correspondence between the unary reciprocal and the binary division by $x/y = x(/y)$ and $/y = 1/y$.

Note that the usual rule ‘ $0x = 0$ ’, which states that “zero-terms can be erased”, is replaced by rules stating that zero-terms can be moved in certain ways in an expression. Indeed, (6) and (7) state that addition by a zero-term commutes with multiplication and reciprocal, so that if a zero-term occurs somewhere inside an expression, then it can be moved outside.

Example 1.2. $((x+4+0y)(2+0z)+0x)(2+0z) = ((x+4)2)2+0x+0y+0z+0z$.

As a derived rule (rule (10) on page 24), we have $0x + 0y = 0xy$, so that several zero-terms can be merged together in one.

Example 1.3. $0x + 0y + 0z + 0z = 0xyz^2$.

The distributivity rule (3) looks different from the usual one, since we have a zero-term on the left-hand side. But it reduces to the usual rule when $0z = 0$. Since e.g. $0 \cdot 2 = 0$,⁵ we have $(x + y)2 = x2 + y2$ and hence we get from the examples above that $((x + 4 + 0y)(2 + 0z) + 0x)(2 + 0z) = 4x + 16 + 0xyz^2$.

Some examples of wheels of fractions are (we list only the underlying sets):

1. $\odot_{\mathbb{Z} \setminus \{0\}} \mathbb{Z} = \mathbb{Q} \cup \{/0, 0/0\}$.
2. $\odot_A A = \{0\}$. This is the *trivial* wheel.
3. $\odot_{\{1\}}(\mathbb{Z}/2\mathbb{Z}) = \{0, 1, /0, 0/0\}$.
4. $\odot_{\{1\}} \mathbb{Z}$, the set of fractions of integers where no identifications are made. Hence two fractions are regarded as equal only if they have the same numerators and the same denominators.
5. $\odot_{\{1,3\}}(\mathbb{Z}/4\mathbb{Z}) = \{0, 1, 2, 3, /0, /2, 0/0, 0/2, 2/0, 2/2\}$. Note that this wheel extends the ring $\mathbb{Z}/4\mathbb{Z}$ with six new elements. Thus, it is possible for such an extension to have more “new” elements than “old” ones.
6. $\odot_{S_0} A$, where $S_0 = \{x \in A \mid xy = xz \Rightarrow y = z\}$, is the ‘total wheel of fractions’. It contains the well-known total *ring* of fractions as the subset $\{x \mid 0x = 0\}$. Moreover, $\odot_{S_0} A$ is what we will call ‘/*-invertible*’: if $xy = 1$, then $y = /x$. Hence $/$ can be used to compute multiplicative inverses whenever such exist.
7. $\odot_U A$, where U is the set of units in A , is */-invertible* and the subset $\{x \mid 0x = 0\}$ is an isomorphic copy of A . This shows that A can be extended to a */-invertible* wheel in a structure-preserving way (as opposed to the construction of a total wheel of fractions, which often kills a lot of ideals).

An advantage of wheels as compared to rings is that several rules that are valid in rings only in special cases, will have counterparts that are generally valid in wheels. One example is the rule

$$xz = yz \wedge z \neq 0 \Rightarrow x = y$$

of integral domains, whose general counterpart for wheels is

$$xz = yz \Rightarrow x + 0z/z = y + 0z/z$$

⁵ $0(1+1) = 0(1+1) + 0 \cdot 0 = 0 \cdot 1 + 0 \cdot 1 = 0 + 0 = 0$.

(derived rule (13) on page 24).

Since any ring can be extended to a wheel in a structure-preserving way, one may always switch to wheel theory if one likes, even if one works with a problem which originates from a context of rings. Suppose for instance that a, b, c are elements of a ring A and that we have concluded that

$$ac = bc.$$

We may then think of a, b, c as elements of a $/$ -invertible wheel that extends A and use $0c = 0$, concluding that

$$a + 0/c = b + 0/c.$$

This does always make sense. Additional information about c can later be used to go further in the calculation.

We arrived at wheels of fractions by a modification of a well-known construction, and it was by no means clear that the chosen construction yields the best result. We will however show that, in a certain sense, this construction is very natural. We do that in the following steps.

First, we forget the operations 0 and $+$ of the ring we started with, so that we are left with the multiplicative monoid. We show that the construction can be carried out in this setting, and that it solves a universal problem for monoids. This shows that it is very natural from the point of view of the multiplicative monoid. We then show that there is a unique way of defining 0 and $+$ such that the construction is functorial from the category of semirings to the category of ‘weak wheels’, which is a very general category. That unique way is precisely the one described above.

Convention. Any category with structures as objects is assumed to have algebraic morphisms as arrows, i.e., mappings that preserve all operations, including constants which are to be viewed as nullary operations.

2 Involution-monoids

The construction sketched in the previous section was motivated by a wish to make certain changes to the multiplicative monoid of a commutative ring. Fortunately, it turns out that the construction is compatible with the additive structure too, but that is not needed for the construction to be made. In fact, if one forgets about the additive structure and treats the multiplicative monoid only, then the construction becomes even more natural. A few sections will therefore be devoted to a study of commutative monoids. In this context, the motivation from the previous section amounts to the following.

Every monoid M comes with a partial inversion function, defined on the group of units in M . M can be extended to a group only when every element of it is cancellable. However, we will show that if M is commutative, then it can always be extended to a commutative monoid with an involution $*$, such that the partial inversion function of M is the restriction of $*$. This will follow as an application of the more general construction given below.

2.1 Definitions and examples

We use the following notion of involution, the canonical example being the inverse of a group.

Definition 2.1 (involution). *An involution on M is a mapping $*$: $M \rightarrow M$ such that*

$$\begin{aligned}x^{**} &= x \\(xy)^* &= y^*x^*.\end{aligned}$$

Note that $e^* = e$ for the identity e , since $e^* = ee^* = e^{**}e^* = (ee^*)^* = e^{**} = e$. Therefore, an involution is a homomorphism $M^{\text{op}} \rightarrow M$. If M is commutative, then an involution is precisely an automorphism of order 2.

Definition 2.2 (involution-monoid). *An involution-monoid is a pair $\langle M, * \rangle$ where M is a monoid and $*$ an involution on it. A morphism of involution-monoids $\varphi : \langle M, * \rangle \rightarrow \langle N, \star \rangle$ is a monoid morphism $M \rightarrow N$ with $\varphi(x^*) = \varphi(x)^\star$ for every x .*

- Example 2.3.**
1. A group with $x^* = x^{-1}$.
 2. An abelian group (or any commutative monoid) with $*$ being the identity morphism.
 3. The multiplicative monoid of a field together with

$$x^* = \begin{cases} 0 & (x = 0) \\ x^{-1} & (x \neq 0). \end{cases}$$

Note however that this definition is constructive only if $x = 0$ is decidable. That is not the case in e.g. \mathbb{R} .

4. The monoid of $n \times n$ matrices with $*$ being transposition.
5. The monoid of strings from a given alphabet (e being the empty string and the composition being concatenation) with x^* being the string x in reversed order.

As is seen, involutions can sometimes be used as approximations of inversion functions. We therefore use the following different notions of inversion.

Definition 2.4. *Let $\langle M, * \rangle$ be an involution-monoid.*

1. *An element $x \in M$ is invertible or a unit if there exists a $y \in M$ with $xy = yx = e$. This is the ordinary notion for monoids and we use the ordinary notation x^{-1} for the element with $xx^{-1} = x^{-1}x = e$. We call x^{-1} the multiplicative inverse of x .*
2. *An element $x \in M$ is $*$ -invertible or a $*$ -unit if $xx^* = x^*x = e$ (hence $*$ -units are units).*
3. *$\langle M, * \rangle$ is said to be $*$ -invertible if all units are $*$ -invertible, i.e., if $x^{-1} = x^*$ whenever x^{-1} is defined.*

Example 2.5. ‘Orthogonal’ is a more common term for $*$ -invertible $n \times n$ -matrixes.

2.2 The construction of involution-monoids from commutative monoids

From now on, M is assumed to be a *commutative* monoid and X a subset of it. S is assumed to be the submonoid generated by X , i.e., S consists of all finite products of elements of X (e being the empty product). Variables $s, s', s'', \dots, s_0, s_1, \dots$ are assumed to vary over S when nothing else is stated.

The ordinary construction of commutative monoids of fractions is precisely like that of commutative rings of fractions, except that one needs a minor modification to take care of the fact that no subtraction is present. Bourbaki [Bou73] defines M_X as the monoid $M \times S / \sim_S$, where $(m, s) \sim_S (m', s')$ means $\exists s'' : s''ms' = s''m's'$. That construction solves the following universal problem:

Find a monoid M_X with $\iota_{(M,X)} : M \rightarrow M_X$, having the property that whenever $\varphi : M \rightarrow N$ is a monoid-morphism with $\varphi(x)$ a unit for every $x \in X$, then there is a unique $\hat{\varphi} : M_X \rightarrow N$ with $\varphi = \hat{\varphi} \circ \iota_{(M,X)}$.

$$\begin{array}{ccc}
 & M_X & \\
 \iota_{(M,X)} \nearrow & & \searrow \hat{\varphi} \\
 M & \xrightarrow{\varphi} & N
 \end{array}$$

Informally, one asks that the elements of a subset X should ‘turn into units’ when mapped into N and defines M_X to be universal with that property. Bourbaki does not assume that N is commutative, but the generality is somewhat illusory, since the homomorphic image of a commutative monoid is again commutative. Since everything takes place inside the image of M , one has all commutativity that is needed. One could therefore as well require that N be commutative and handle a morphism $M \rightarrow N'$ (N' not commutative) by letting N be the image, factorizing the morphism as $M \rightarrow N \rightarrow N'$.

An analogue for involution-monoids is to ask that the elements $x \in X$ should ‘turn into \star -units’ when mapped into an involution-monoid $\langle N, \star \rangle$. Formally, we state it as follows, with T denoting the forgetful functor taking the monoid out of an involution-monoid. Motivated by the arguments in the previous paragraph, we assume that N is commutative.⁶

Let M be a commutative monoid and X an arbitrary subset of M . Then find an involution-monoid M_X^* with $\eta_{(M,X)} : M \rightarrow T(M_X^*)$, having the property that whenever $\langle N, \star \rangle$ is a commutative involution-monoid and $\varphi : M \rightarrow N$ a monoid-morphism with $\varphi(x)\varphi(x)^* = \varphi(x)^*\varphi(x) = e$ for every $x \in X$, then there is a unique $\hat{\varphi} : M_X^* \rightarrow \langle N, \star \rangle$ with $\varphi = T(\hat{\varphi}) \circ \eta_{(M,X)}$.

⁶In our situation, this requirement is really essential, since we do not any longer work inside the image of M , but in the generated involution-monoid, which is not automatically commutative.

$$\begin{array}{ccc}
& T(M_X^*) & \\
\eta_{(M,X)} \nearrow & & \searrow T(\hat{\varphi}) \\
M & \xrightarrow{\varphi} & N
\end{array}$$

We show how to find M_X^* . Consider the involution-monoid $\langle M \times M, * \rangle$ where $(x, y)^* = (y, x)$ and define \equiv_S on it by

$$(x, y) \equiv_S (x', y') \quad \text{means} \quad \exists s_1, s_2 : (s_1, s_1)(x, y) = (s_2, s_2)(x', y').$$

Then \equiv_S is clearly reflexive and symmetric, but it is also transitive, since if $(s_1, s_1)(x, y) = (s_2, s_2)(x', y')$ and $(s_3, s_3)(x', y') = (s_4, s_4)(x'', y'')$, then it follows that $(s_1 s_3, s_1 s_3)(x, y) = (s_2 s_4, s_2 s_4)(x'', y'')$. It is easily seen that \equiv_S preserves the operations, so that it is in fact a congruence relation. The congruence class containing (x, y) will be denoted by $[x, y]$.

Definition 2.6. Let $M_X^* = \langle M \times M, * \rangle / \equiv_S$, and let $\eta_{(M,X)} : M \rightarrow T(M_X^*)$ be defined by $x \mapsto [x, e]$.

Theorem 2.7 (solution to the universal problem). Suppose $\langle N, * \rangle$ is a commutative involution-monoid and $\varphi : M \rightarrow N$ a monoid-morphism. Then $\hat{\varphi}([x, y]) = \varphi(x)\varphi(y)^*$ defines a morphism of involution-monoids $M_X^* \rightarrow \langle N, * \rangle$ and if ψ is such a morphism too, then $\varphi = T(\psi) \circ \eta_{(M,X)}$ if and only if $\psi = \hat{\varphi}$.

Proof. We first prove uniqueness of ψ . In order for it to extend φ , we must have

$$\begin{aligned}
\psi([x, y]) &= \psi([x, e][y, e]^*) = \psi([x, e])\psi([y, e])^* = \psi(\eta_{(M,X)}(x))\psi(\eta_{(M,X)}(y))^* \\
&= \varphi(x)\varphi(y)^*.
\end{aligned}$$

Hence, $\hat{\varphi}$ is the only candidate. It is well-defined, since if $[x, y] = [x', y']$ then there are s_1, s_2 with $(s_1, s_1)(x, y) = (s_2, s_2)(x', y')$ and therefore

$$\begin{aligned}
\varphi(x)\varphi(y)^* &= \varphi(x)e\varphi(y)^* = \varphi(x)\varphi(s_1)\varphi(s_1)^*\varphi(y)^* = \varphi(xs_1)\varphi(ys_1)^* \\
&= \varphi(x's_2)\varphi(y's_2)^* = \varphi(x')\varphi(s_2)\varphi(s_2)^*\varphi(y')^* = \varphi(x')e\varphi(y')^* \\
&= \varphi(x')\varphi(y')^*.
\end{aligned}$$

It is a monoid-morphism since

$$\begin{aligned}
\hat{\varphi}([x, y][x', y']) &= \hat{\varphi}([xx', yy']) = \varphi(xx')\varphi(yy')^* = \varphi(x)\varphi(x')\varphi(y)^*\varphi(y')^* \\
&= \varphi(x)\varphi(y)^*\varphi(x')\varphi(y')^* = \hat{\varphi}([x, y])\hat{\varphi}([x', y'])
\end{aligned}$$

and

$$\hat{\varphi}([e, e]) = \varphi(e)\varphi(e)^* = ee^* = e.$$

It preserves the involution since

$$\begin{aligned}
\hat{\varphi}([x, y]^*) &= \hat{\varphi}([y, x]) = \varphi(y)\varphi(x)^* = \varphi(y)^{**}\varphi(x)^* = (\varphi(x)\varphi(y)^*)^* \\
&= \hat{\varphi}([x, y])^*.
\end{aligned}$$

□

2.3 Insertion of the parent monoid

The mapping $\eta_{(M,X)}$ differs from the usual $\iota_{(M,X)}$ only in the choice of codomain, since we will see in theorem 2.10 that M_X is included in M_X^* . Hence, many properties of ι carry over to η . One such is that $\eta_{(M,X)}$ is not in general injective.

Proposition 2.8. $\eta_{(M,X)}(a) = \eta_{(M,X)}(b)$ iff there is some $s \in S$ with $sa = sb$.

Proof.

$$\begin{aligned} & [a, e] = [b, e] \\ \iff & \exists s_1, s_2 : (s_1, s_1)(a, e) = (s_2, s_2)(b, e) \\ \iff & \exists s : sa = sb. \end{aligned}$$

□

Corollary 2.9. $\eta_{(M,X)}$ is injective iff every $x \in X$ is cancellable.

Proof. If every $x \in X$ is cancellable, then every $s \in S$ is too, since S is generated by X . Hence $sa = sb \Rightarrow a = b$ and the rest follows from the proposition. □

It is convenient to use the same symbol for an element of a monoid M and its image under $\eta_{(M,X)}$, even when $\eta_{(M,X)}$ is not injective (remembering that lack of injectivity means that equality in M_X^* need not imply equality in M). We will use that notation frequently. In particular, it allows us to write xy^* instead of $[x, y]$.

Theorem 2.10. The homomorphism $\hat{\eta}_{(M,X)} : M_X \rightarrow T(M_X^*)$, induced by $\eta_{(M,X)}$ according to the universal property of M_X , is injective.

$$\begin{array}{ccc} & M_X & \\ \iota_{(M,X)} \nearrow & & \searrow \hat{\eta}_{(M,X)} \\ M & \xrightarrow{\eta_{(M,X)}} & T(M_X^*) \end{array}$$

Proof. This is best seen in a very concrete way, examining how the constructions of M_X and M_X^* are related.

Let \sim_S be the restriction of \equiv_S to $M \times S$. Then

$$(x, s) \sim_S (x', s') \iff \exists s'' : s''xs' = s''x's,$$

because on the one hand, if $s''xs' = s''x's$, then we have

$$(s''s', s''s')(x, s) = (s''xs', s''ss') = (s''x's, s''ss') = (s''s, s''s)(x', s')$$

and on the other, if $(s_1, s_1)(x, s) = (s_2, s_2)(x', s')$, then

$$s_1xs' = s_2x's' = s_1x's.$$

Since \sim_S is a restriction of \equiv_S , it follows that the mapping $M \times S/ \sim_S \rightarrow M \times M/ \equiv_S$ is injective. This mapping is clearly $\hat{\eta}_{(M,X)}$.

$$\begin{array}{ccc}
M \times S & \xrightarrow{\quad\quad\quad} & M \times M \\
\downarrow & & \downarrow \\
M \times S/ \sim_S & \dashrightarrow & M \times M/ \equiv_S
\end{array}$$

□

We therefore regard M_X as a submonoid of $T(M_X^*)$.

2.4 The role of X in the structure of M_X^*

When exploring possible structures of M_X^* , we don't have to consider arbitrary subsets X , but only submonoids S , since $M_X^* = M_S^*$ when S is the submonoid generated by X . When comfortable, we may even restrict our attention to a special class of submonoids: those "closed under division".

Definition 2.11. A submonoid S of M is closed under division if it holds that

$$sx \in S \Rightarrow x \in S. \quad (s \in S, x \in M)$$

Definition 2.12. The divisional closure of a submonoid S of M is the smallest submonoid of M that contains S and is closed under division.

Lemma 2.13. $\tilde{S} = \{x \in M \mid \exists s : sx \in S\}$ is the divisional closure of S .

Proof. Since $es \in S$, we have $s \in \tilde{S}$ and hence $S \subset \tilde{S}$.

It is obvious that \tilde{S} cannot be smaller, the question is whether it is closed under multiplication and division. Suppose $x, y \in \tilde{S}$, say $s_1x, s_2y \in S$. Then $(s_1s_2)(xy) \in S$, so that $xy \in \tilde{S}$. Hence it is a submonoid of M .

To see that it is closed under division, suppose that $\tilde{s}x \in \tilde{S}$, with $\tilde{s} \in \tilde{S}$. Then there is an s with $s(\tilde{s}x) \in S$. But there is also an s' such that $s'\tilde{s} \in S$. Then $ss'\tilde{s} \in S$ and $(ss'\tilde{s})x = (s's)(\tilde{s}x) \in S$, so that $x \in \tilde{S}$. □

This is an algebraic closure operator (in the sense of e.g. [BS81]).

Proposition 2.14. $M_X^* = M_S^* = M_{\tilde{S}}^*$ (they are identical as sets and have identical operations).

Proof. It is clear by definition of M_X^* that it is identical to M_S^* . We prove $M_S^* = M_{\tilde{S}}^*$.

Suppose that $[x, y] = [x', y']$ in M_S^* , say $(s_1, s_1)(x, y) = (s_2, s_2)(x', y')$. Then, since $S \subset \tilde{S}$, $[x, y] = [x', y']$ in $M_{\tilde{S}}^*$. On the other hand, suppose $[x, y] = [x', y']$ in $M_{\tilde{S}}^*$, say $(x_1, x_1)(x, y) = (x_2, x_2)(x', y')$ with $x_1, x_2 \in \tilde{S}$. Then there are by definition of \tilde{S} some $s_1, s_2 \in S$ with $s_1x_1, s_2x_2 \in S$. Hence $((s_1x_1)s_2, (s_1x_1)s_2)(x, y) = (s_1(s_2x_2), s_1(s_2x_2))(x', y')$, which means that $[x, y] = [x', y']$ in M_S^* .

Thus M_S^* and $M_{\tilde{S}}^*$ are identical as sets. Their operations are identical since they are defined in the same way. □

We now investigate the connection between the involution of M_S^* and the partial inversion function on its underlying monoid.

Proposition 2.15. $\eta_{(M,X)}(x)$ is $*$ -invertible if and only if x belongs to the divisional closure of the submonoid S generated by X .

Proof. Let $x \in \tilde{S}$, say $sx = s' \in S$. Then $(s, s)(x, x) = (s', s')(e, e)$, so that $xx^* = e$ in M_S^* .

On the other hand, suppose $xx^* = e$ in M_S^* , say $(s, s)(x, x) = (s', s')(e, e)$. Then $sx = s' \in S$, so that $x \in \tilde{S}$. \square

In particular, we will often (and without explanation) use that $ss^* = e$ for any $s \in S$.

Definition 2.16. A submonoid $S \subset M$ is saturated if $xy \in S \Rightarrow x, y \in S$.

Example 2.17. The group of units is saturated as a subset of M .

Example 2.18. Let $S_0 = \{x \in A \mid xy = xz \Rightarrow y = z\}$. It is saturated as a subset of M . We call it “the submonoid of cancellable elements”.

Proposition 2.19. M_S^* is $*$ -invertible if and only if \tilde{S} is saturated in M .

Proof. By proposition 2.14 we may assume that $S = \tilde{S}$.

Suppose that S is saturated and that $[a, b]$ is a unit in M_S^* , say $[a, b][a', b'] = e$, which is to say that $(s_1, s_1)(aa', bb') = (s_2, s_2)(e, e)$ for some s_1, s_2 . Then $s_1aa' = s_1bb' = s_2 \in S$, so that $a, b \in S$ by saturation. Hence $ab \in S$ so that $[a, b][a, b]^* = [ab, ab] = e$.

Suppose on the other hand that every unit is $*$ -invertible and take $xy \in S$. We shall prove $x \in S$. In M_S^* , x is a unit with $x^{-1} = y(xy)^*$, but then it is $*$ -invertible by assumption, so that $y(xy)^* = x^*$. That means that there are s_1, s_2 with $(s_1, s_1)(y, xy) = (s_2, s_2)(e, x)$, in particular $s_2x = s_1(xy)$. But xy was taken from S , so we conclude that $s_2x \in S$ and hence that $x \in \tilde{S}$. Since we have assumed that $S = \tilde{S}$, we get $x \in S$. A similar argument proves that $y \in S$. \square

Corollary 2.20. Let U be the group of units in M . Then M_U^* is $*$ -invertible and it contains M as a submonoid (this solves the problem of finding an extension of M together with an involution extending the partial inversion function of M).

Proof. Since U is saturated, M_U^* is $*$ -invertible by the previous proposition. By corollary 2.9, $\eta_{(M,U)}$ is injective, since units are cancellable. \square

Proposition 2.21. The restriction to M of the involution of M_U^* , is the partial inversion function of M .

Proof. The restriction of $*$ to M extends the partial inversion function since M_U^* is $*$ -invertible. We shall prove that if $x, y \in M$ with $y = x^*$ in M_U^* , then $xy = e$. Suppose therefore that $x, y \in M$ and that $y = x^*$. The latter means that there are units $u_1, u_2 \in M$ with $(u_1, u_1)(y, e) = (u_2, u_2)(e, x)$, but then $u_2x = u_1$, so that $x \in U$. Hence $xx^* = e$, thus $xy = e$. \square

2.5 The construction as a functor

We may turn the construction $(M, X) \mapsto M_X^*$ into a functor F in the following way.

Let \mathbf{C} be the category of pairs (M, X) where M is a commutative monoid and X a subset of M . A \mathbf{C} -arrow $(M, X) \rightarrow (M', X')$ is a monoid-morphism $\varphi : M \rightarrow M'$ with $\varphi(X) \subset X'$. Define $F(M, X)$ to be M_X^* and $F(\varphi)$ to be the mapping $[x, y] \mapsto [\varphi(x), \varphi(y)]$. It is well-defined since if $[x, y] = [x', y']$, then there are s_1, s_2 such that $(s_1, s_1)(x, y) = (s_2, s_2)(x', y')$. Since $\varphi(X) \subset X'$, it follows that $\varphi(s_1)$ and $\varphi(s_2)$ are elements of the submonoid generated by X' . Since $(\varphi(s_1), \varphi(s_1))(\varphi(x), \varphi(y)) = (\varphi(s_2), \varphi(s_2))(\varphi(x'), \varphi(y'))$, it follows that $[\varphi(x), \varphi(y)] = [\varphi(x'), \varphi(y')]$. It is easily seen that $F(\varphi)$ is a morphism of involution-monoids. It is now easy to check that F is a functor from \mathbf{C} to the category $\mathbf{CInvMon}$ of commutative involution-monoids.

Proposition 2.22. *F has a right adjoint G , which is defined on objects by $\langle M, * \rangle \mapsto (M, X)$, with X the set of $*$ -units in $\langle M, * \rangle$, and on morphisms by $\alpha \mapsto T(\alpha)$. η is a natural transformation from the identity to GF and it is the unit of the adjunction. The counit is ϵ with $\epsilon_{\langle M, * \rangle} : [x, y] \mapsto xy^*$.*

Proof. $\eta_{(M, X)}$ maps X into the set U of $*$ -units, thus is a \mathbf{C} -arrow $(M, X) \rightarrow T(M_X^*, U)$. The naturalness is obvious.

We have to prove that

$$(\epsilon F)_{(M, X)}(F\eta)_{(M, X)} = \mathbf{Id}_{M_X^*}$$

and

$$(G\epsilon)_{\langle M, * \rangle}(\eta F)_{\langle M, * \rangle} = \mathbf{Id}_{G(\langle M, * \rangle)}.$$

$$\begin{aligned} (\epsilon F)_{(M, X)} &: FG(M_X^*) \rightarrow M_X^* \text{ by } [[x, y], [z, w]] \mapsto [xw, yz] \\ (F\eta)_{(M, X)} &: M_X^* \rightarrow FG(M_X^*) \text{ by } [x, y] \mapsto [[x, e], [y, e]] \\ (G\epsilon)_H &: GFG(\langle M, * \rangle) \rightarrow G(\langle M, * \rangle) \text{ by } [x, y] \mapsto xy^* \\ (\eta G)_H &: G(\langle M, * \rangle) \rightarrow GFG(\langle M, * \rangle) \text{ by } x \mapsto [x, e] \end{aligned}$$

Then

$$(\epsilon F)_{(M, X)}(F\eta)_{(M, X)}([x, y]) = (\epsilon F)_{(M, X)}([[x, e], [y, e]]) = [xe, ye] = [x, y]$$

and

$$(G\epsilon)_{\langle M, * \rangle}(\eta G)_{\langle M, * \rangle}(x) = (G\epsilon)_{\langle M, * \rangle}([x, e]) = xe^* = x.$$

□

$\hat{\eta}_{(M, X)}$ is a transformation from M_X to $T(M_X^*)$, natural in (M, X) .

Remark. Involutions on monoids $\langle M, 0, + \rangle$ will throughout be denoted by $-$. We write $-x$ for x^* in this case. Further, we write $x - y$ for $x + (-y)$.

Involutions on monoids $\langle M, 1, \cdot \rangle$ will throughout be denoted by $/$. We write $/x$ for x^* in this case. Further, we write x/y for $x \cdot (/y)$.

3 Applications to semirings

The consideration of commutative monoids leads naturally to consideration of semirings in the sense of e.g. Golan [Gol99] (we define this notion below, the word ‘semiring’ has no uniform meaning in the literature), in the following way.

When addition has been defined on the natural numbers, turning \mathbb{N} into a commutative monoid, one finds that \mathbb{N} can be used to add, not only finite collections of elements, but also finite collections of equally large finite collections of elements. This motivates the introduction of multiplication on \mathbb{N} .

The argument in the previous paragraph is an elementary way of saying that \mathbb{N} can be viewed as its own endomorphism monoid with multiplication being composition. The combination with addition turns \mathbb{N} into a semiring.

In general, suppose that $M = \langle M, 0, + \rangle$ is a commutative monoid. Then its monoid $\mathbf{End}(M)$ of endomorphisms has a natural additive structure inherited from M by

$$\begin{aligned} 0(x) &= 0, \\ (f + g)(x) &= f(x) + g(x). \end{aligned}$$

These definitions make $\mathbf{End}(M)$ a semiring.

Definition 3.1. *A semiring is a structure $\langle M, 0, 1, +, \cdot \rangle$ such that*

1. $\langle M, 0, + \rangle$ is a commutative monoid.
2. $\langle M, 1, \cdot \rangle$ is a monoid.
3. $(x + y)z = xz + yz$ and $x(y + z) = xy + xz$.
4. $0x = x0 = 0$.

We do not exclude the trivial case $0 = 1$.

If there is a solution a to the equation $1 + x = 0$ in a semiring A , then we may define $-x = ax$ and we get $x + (-x) = 1x + ax = (1 + a)x = 0x = 0$, so that A is a ring in this case.

Beside the important example $\mathbf{End}(M)$, we have many mathematical structures that are semirings, for instance:

1. Any ring with identity.
2. The (left or right) ideals of a ring, with $I + J$ defined to be $\{i + j \mid i \in I, j \in J\}$ and IJ defined to be the (left or right) ideal generated by $\{ij \mid i \in I, j \in J\}$. 0 is the trivial ideal and 1 is the improper ideal.
3. Any bounded distributive lattice, e.g. Heyting algebras and Boolean algebras. Here $\langle 0, 1, +, \cdot \rangle$ is $\langle \perp, \top, \vee, \wedge \rangle$.

We need some definitions, which are taken from [Gol99], except that we use the word ‘module’ for what there is called ‘semimodule’. The reason for this choice is that ‘module’ is shorter and we see no point in distinguishing the notions since if A happens to be a ring, then an A -module in the following sense is automatically an A -module in the sense of rings, since if we define $-m = (-1)m$, then $m + (-m) = (1 + (-1))m = 0m = 0$.

Definition 3.2. Let A be a semiring. A left A -module is a commutative monoid $\langle M, 0, + \rangle$ with multiplication by A -elements to the left defined (formally, a function $A \times M \rightarrow M$ written $(a, m) \mapsto am$) such that for any $a, a' \in A, m, m' \in M$,

$$\begin{aligned} (aa')m &= a(a'm) \\ 1m &= m \\ (a + a')m &= am + a'm \\ a(m + m') &= am + am' \\ 0m &= 0 \\ a0 &= 0 \end{aligned}$$

where 0 to the left is in A , while 0 to the right or alone is in M .

Every commutative monoid $\langle M, e, \cdot \rangle$ is an \mathbb{N} -module with the multiplication $(n, m) \mapsto m^n$. This is in analogy with the fact that every abelian group is a \mathbb{Z} -module.

Definition 3.3. A left ideal of A is a left A -module inside A ; i.e., it is a submonoid of the additive monoid of A and it is closed under multiplication to the left by any element from A .⁷

The notions of right A -module and right ideal are defined analogously.

3.1 Applications to additive monoids of semirings

We show how our construction can be used for extending a semiring so that an additive involution $-$ can be defined on the result. The process is similar to the construction of \mathbb{Z} from \mathbb{N} , but it does not need that all elements are additively cancellable.

The idea is that we first apply our construction to the additive monoid, then defining multiplication on the result.

Suppose that $\langle A, 0, + \rangle$ is the additive monoid of a semiring A and that X is a subset. Let S be the right ideal generated by X ; i.e., S consists of all finite sums

$$\sum_i x_i a_i \quad (x_i \in X, a_i \in A).$$

We use the construction $\langle A, 0, + \rangle_S^*$, given in section 2.2, writing $-x$ for x^* . In section 2.2, we noted that each element of M_X^* is of the form xy^* for $x, y \in M$; the corresponding statement now is that each element is of the form $x + (-y)$, which will be written as $x - y$. Such elements $x - y$ and $x' - y'$ (with $x, y, x', y' \in A$) are equal when there exists s, s' with $(s, s) + (x, y) = (s', s') + (x', y')$ in $A \times A$ (in particular, $1 - 1 = 0$ holds precisely when there is some s such that $s + 1 \in S$, see proposition 2.15).

We now define a ‘multiplication’ by

$$(x - y)(z - w) = (xz + yw) - (xw + yz). \quad (x, y, z, w \in A)$$

⁷We accept improper ideals, which [Gol99] does not. This is natural since we accept trivial semirings.

It is well-defined, since if $x - y = x' - y'$, then there are $s, s' \in S$ with $(s, s) + (x, y) = (s', s') + (x', y')$ in $A \times A$, so that

$$\begin{aligned} (sx' + sy', sx' + sy') + (xz + yw, xw + yz) \\ &= ((s+x)z + (s+y)w, (s+x)w + (s+y)z) \\ &= ((s'+x')z + (s'+y')w, (s'+x')w + (s'+y')z) \\ &= (s'z + s'w, s'w + s'z) + (x'z + y'w, x'w + y'z) \end{aligned}$$

and since S is a right ideal, $sx' + sy' \in S$ and $s'z + s'w \in S$. Hence $(xz + yz) - (xw + yz) = (x'z + y'z) - (x'w + y'z)$.

The structure obtained, together with the constant 1, will be denoted by A_X^- . It is clearly a semiring: indeed, it is a quotient of A_\emptyset^- which is the ‘convolution algebra’, or ‘monoid-semiring’, on A defined by the monoid $\{1, -1\}$.⁸ A natural morphism $\eta_{(A,X)} : A \rightarrow A_X^-$ is given by $\eta_{(A,X)}(x) = \eta_{((A,0,+),X)}(x) = x$.

Example 3.4. $\mathbb{Z} = \mathbb{N}_\mathbb{N}^-$.

Example 3.5. An example when $1 - 1 \neq 0$ is obtained by thinking of machines not capable of counting with elements larger than the natural number N .

Let \sim be defined on \mathbb{N} as “equal or large”, formally:

$$x \sim y \quad \text{means} \quad x = y \vee (x \geq N \wedge y \geq N).$$

Then \sim is a congruence relation and \mathbb{N}/\sim consists of $N + 1$ elements. Since no element except 0 is additively cancellable, we can’t introduce an additive inverse, but we may construct $(\mathbb{N}/\sim)_\emptyset^-$, in which $1 - 1 \neq 0$.

Now, consider $(\mathbb{N}/\sim) \times (\mathbb{Z}/7\mathbb{Z})$ (it can be thought of as a data structure with elements being pairs representing numbers together with a day of the week). An element (x, y) of it is additively cancellable if and only if $x = 0$. Let S_0 be the set of such elements. It holds in $((\mathbb{N}/\sim) \times (\mathbb{Z}/7\mathbb{Z}))_{S_0}^-$ that $(0, x) - (0, x) = 0$, but not that $1 - 1 = 0$.

Example 3.6. Let \mathbb{N}^∞ be \mathbb{N} extended with an element ∞ with

$$\begin{aligned} x + \infty &= \infty + x = \infty \\ x\infty &= \infty x = \begin{cases} 0 & (x = 0) \\ \infty & (x \neq 0). \end{cases} \end{aligned}$$

This notion of infinity differs from that in the introduction, since here we have $0\infty = 0$, while we had $0\infty = \perp$ in the introduction.

\mathbb{N}^∞ is clearly a semiring, thus we may construct $(\mathbb{N}^\infty)_X^-$. However, if $a \in X$ for some $a \neq 0$, then ∞ is an element of the right ideal generated by X , hence $(\mathbb{N}^\infty)_X^-$ is trivial, since $(\infty, \infty) + (x, y) = (\infty, \infty) + (x', y')$ for all x, y, x', y' . Hence, if we want non-triviality, we need to take $X \subset \{0\}$, ending up with the structure $(\mathbb{N}^\infty)_\emptyset^-$. In that structure, $x - x = 0$ is true only for $x = 0$.

The operation $-$ makes A_X^- a semiring⁻:

⁸See [Gol99, example 3.3, p. 29]. The notation is different.

Definition 3.7. A semiring⁻ is a semiring with an additive involution $-$ such that

$$-x = (-1)x = x(-1).$$

A homomorphism of semirings⁻ is a semiring homomorphism which also preserves $-$.

One could also describe a semiring⁻ as a semiring with a constant -1 such that $(-1)(-1) = 1$ and $(-1)x = x(-1)$ for all x . An involution with the required properties is then given by $-x = (-1)x$.

Note that a ring is precisely a semiring⁻ in which $1 - 1 = 0$ holds. Then $x - x = (1 - 1)x = 0x = 0$ for all x .

The construction solves the following universal problem. Let T be the functor which forgets the $-$ of semirings⁻, turning them into semirings.

Theorem 3.8. Suppose that A is a semiring and X a subset of it. Suppose also that B is a semiring⁻ and $\varphi : A \rightarrow T(B)$ a semiring morphism such that $\varphi(x) - \varphi(x) = 0$ for all $x \in X$. Then there exists a unique morphism $\hat{\varphi} : A_X^- \rightarrow B$ of semirings⁻ such that $\varphi = T(\hat{\varphi}) \circ \eta_{(A,X)}$.

$$\begin{array}{ccc} & T(A_X^-) & \\ \eta_{(A,X)} \nearrow & & \searrow T(\hat{\varphi}) \\ A & \xrightarrow{\varphi} & T(B) \end{array}$$

Proof. Note first that if $s \in S$, then $\varphi(s) - \varphi(s) = 0$, since

$$\begin{aligned} \varphi\left(\sum_i x_i a_i\right) - \varphi\left(\sum_i x_i a_i\right) &= \sum_i (\varphi(x_i) - \varphi(x_i))\varphi(a_i) \\ &= \sum_i (0\varphi(a_i)) = 0. \end{aligned}$$

Hence, we may replace X by S everywhere in the statement of the theorem. Then theorem 2.7 gives the unique candidate $\hat{\varphi} : [x, y] \mapsto \varphi(x) - \varphi(y)$ and proves that it preserves $0, +, -$. It also preserves 1 , since $\hat{\varphi}(1) = \hat{\varphi}([1, 0]) = \varphi(1) = 1$. Preservation of multiplication is checked thus: Let $x, y, z, w \in A$.

$$\begin{aligned} \hat{\varphi}((x - y)(z - w)) &= \hat{\varphi}((xz + yw) - (xw + yz)) \\ &= \varphi(xz + yw) - \varphi(xw + yz) \\ &= \varphi(x)\varphi(z) + \varphi(y)\varphi(w) - (\varphi(x)\varphi(w) + \varphi(y)\varphi(z)) \\ &= (\varphi(x) - \varphi(y))(\varphi(z) - \varphi(w)) \\ &= \hat{\varphi}(x - y)\hat{\varphi}(z - w). \end{aligned}$$

The rules $-x = (-1)x = x(-1)$ were applied in the fourth step. \square

3.2 Applications to multiplicative monoids of commutative semirings

In the following, all monoids and semirings are assumed to be commutative (a semiring is commutative if $+$ and \cdot are commutative). When we use notions like units, divisional closures etc., we refer to the multiplicative monoid (when nothing else is stated).

We will sometimes write $\frac{x}{y}$ for x/y .

Given a commutative semiring, one may first apply the construction to the additive monoid, introducing an additive involution as was explained in the previous section. One may then continue by applying the construction to the multiplicative monoid, introducing also a multiplicative involution $/$. We show in this section how that second step is carried out. It is not necessary that an additive involution is present, one may as well start from a plain semiring. However, if both $-$ and $/$ are wanted, then $-$ should be constructed first, since the result of the construction in this section will not be a semiring anymore.

In section 1.2, it was sketched how this construction is made. We there supposed that it was applied to a ring, but semirings work as well. We now show that the choice of definition of $+$ was not arbitrary, it is the unique choice which yields a functorial definition (using some very general conditions). More precisely, let M be the multiplicative monoid of a semiring. Then there is a unique way of defining on M_X^* a binary operation $+$ with neutral element 0 , such that the functor $F : M \mapsto M_X^*$ acts functorially also with respect to 0 and $+$ and such that $\eta_{(M,X)}$ preserves also these operations.

Technically, we state it in the theorem below. We need some preliminaries.

Definition 3.9. Let \mathbf{C}' be the category with objects (A, X) where A is a (commutative) semiring and X a subset of it. An arrow $(A, X) \rightarrow (A', X')$ is a semiring morphism $\varphi : A \rightarrow A'$ with $\varphi(X) \subset X'$.

Note that there is a forgetful functor $T_1 : \mathbf{C}' \rightarrow \mathbf{C}$ (the category \mathbf{C} was defined on page 14), forgetting the additive structure.

Definition 3.10 (weak wheel). A weak wheel is a structure $\langle H, 0, 1, +, \cdot, / \rangle$ with $\langle H, 1, \cdot, / \rangle$ a commutative involution-monoid and 0 neutral for $+$, i.e. $0 + x = x + 0 = x$.

Let T_2 be the forgetful functor (forgetting 0 and $+$) from the category \mathbf{WW} of weak wheels to the category $\mathbf{CInvMon}$ of commutative involution-monoids.

Theorem 3.11. There is a unique functor $F' : \mathbf{C}' \rightarrow \mathbf{WW}$ such that the diagram

$$\begin{array}{ccc}
 \mathbf{C}' & \xrightarrow{\quad F' \quad} & \mathbf{WW} \\
 T_1 \downarrow & & \downarrow T_2 \\
 \mathbf{C} & \xrightarrow{\quad F \quad} & \mathbf{CInvMon}
 \end{array}$$

commutes and such that, for each (A, X) , there is an operation-preserving function $\eta_{(A,X)}$ which makes the following diagram commute (here M is the multiplicative monoid of A).

$$\begin{array}{ccc}
(A, X) & \overset{\eta_{(A,X)}}{\dashrightarrow} & F'(A, X) \\
\downarrow x \mapsto x & & \downarrow x \mapsto x \\
(M, X) & \xrightarrow{\eta_{(M,X)}} & M_X^*
\end{array}$$

Proof. There is a unique possible definition of $\eta_{(A,X)}$: by the lower diagram, $F'(A, X)$ is M_X^* with additional structure (with the notation $/$ instead of $*$ and with $\frac{x}{y}$ sometimes denoting x/y), so that we must have $\eta_{(A,X)}(x) = \eta_{(M,X)}(x)$.

We therefore define $\eta_{(A,X)}$ as $x \mapsto \eta_{(M,X)}(x)$. We then have what is needed to check that the lower diagram commutes, but it is not clear that $\eta_{(A,X)}$ preserves the operations, since 0 and $+$ have not yet been defined.

Preservation of 0 requires that 0 be defined in $F'(A, X)$ as $\eta_{(A,X)}(0)$, which has to be $\eta_{(M,X)}(0)$ by the previous paragraph. We therefore make this definition. Note that it is compatible with our general use of the notation ' x ' for $\eta_{(M,X)}(x)$.

It remains to investigate how $+$ must be defined in order for $\eta_{(A,X)}$ to preserve $+$ and for the upper diagram to commute.

Uniqueness:

The requirement that $+$ be preserved by $\eta_{(A,X)}$ means that there should be no difference between $\eta_{(A,X)}(x) + \eta_{(A,X)}(y)$ and $\eta_{(A,X)}(x + y)$. Thus, we may safely use the notation ' $x + y$ ' for both.

We have to know how F' must act on arrows. Let φ be an arrow in \mathbf{C}' . How $F'(\varphi)$ acts on elements is given by how it acts after the forgetful functor T_2 has been applied. We have for $x, y \in A$ that $((T_2 F')(\varphi))(x/y) = ((F T_1)(\varphi))(x/y) = \varphi(x)/\varphi(y)$.

Consider $\mathbb{N}[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4]$, the semiring of polynomials in $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\}$ with natural numbers as coefficients.⁹ Clearly, in $F'(\mathbb{N}[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4], \emptyset)$, we have

$$\frac{\mathbf{x}_1}{\mathbf{x}_2} + \frac{\mathbf{x}_3}{\mathbf{x}_4} = \frac{p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)}{q(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)}$$

for some pair of polynomials (p, q) with natural numbers as coefficients. For any (A, X) in \mathbf{C}' and any $a_1, a_2, a_3, a_4 \in A$, there is a unique morphism $\varphi : (\mathbb{N}[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4], \emptyset) \rightarrow (A, X)$ with $\varphi(\mathbf{x}_i) = a_i$, hence

$$\begin{aligned}
\frac{a_1}{a_2} + \frac{a_3}{a_4} &= \frac{\varphi(\mathbf{x}_1)}{\varphi(\mathbf{x}_2)} + \frac{\varphi(\mathbf{x}_3)}{\varphi(\mathbf{x}_4)} = (F'(\varphi)) \left(\frac{\mathbf{x}_1}{\mathbf{x}_2} + \frac{\mathbf{x}_3}{\mathbf{x}_4} \right) \\
&= (F'(\varphi)) \left(\frac{p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)}{q(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)} \right) = \frac{\varphi(p(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4))}{\varphi(q(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4))} \\
&= \frac{p(a_1, a_2, a_3, a_4)}{q(a_1, a_2, a_3, a_4)}.
\end{aligned}$$

⁹We use $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$ as formal symbols, distinguishing them from variables x_1, x_2, x_3, x_4 .

So, addition has to be defined by p and q in $F'(A, X)$, for any choice of (A, X) . The polynomials p and q are homogeneous of the same degree, since in $F'(\mathbb{N}, \{2\})$ we must have (for $x, y, z, w \in \mathbb{N}$)

$$\begin{aligned} \frac{p(2x, 2y, 2z, 2w)}{q(2x, 2y, 2z, 2w)} &= \frac{2x}{2y} + \frac{2z}{2w} = \frac{2}{2} \cdot \frac{x}{y} + \frac{2}{2} \cdot \frac{z}{w} \\ &= 1 \cdot \frac{x}{y} + 1 \cdot \frac{z}{w} = \frac{x}{y} + \frac{z}{w} = \frac{p(x, y, z, w)}{q(x, y, z, w)} \end{aligned}$$

so that there exist n, m with

$$(2^n, 2^n)(p(2x, 2y, 2z, 2w), q(2x, 2y, 2z, 2w)) = (2^m, 2^m)(p(x, y, z, w), q(x, y, z, w)).$$

But then p and q must be homogeneous of degree $m - n$.

We now compute q . Since the $\eta_{(A, X)}$ are required to preserve $+$, we must have

$$\frac{x}{1} + \frac{y}{1} = \frac{x+y}{1}$$

so that $q(x, 1, z, 1) = 1$. Hence $q(x, y, z, w) = y^i w^j$ for some i, j with $i + j = m - n$. Since 0 is required to be neutral for $+$, we must also have

$$\frac{x}{y} + \frac{0}{1} = \frac{x}{y}, \quad \frac{0}{1} + \frac{z}{w} = \frac{z}{w},$$

hence $q(x, y, 0, 1) = y$ and $q(0, 1, z, w) = w$. Hence $i = j = 1$ and we conclude that $q(x, y, z, w) = yw$.

We now compute p , which has to be homogeneous of degree 2 , as q is. Preservation of $+$ gives (see above) that $p(x, 1, z, 1) = x + z$, hence $p(x, y, z, w) = xy^k w^{1-k} + zy^\ell w^{1-\ell}$ for some $k, \ell \in \{0, 1\}$. That 0 is neutral gives $p(x, y, 0, 1) = x$ and $p(0, 1, z, w) = z$, hence $k = 0$ and $\ell = 1$, so that $p(x, y, z, w) = xw + yz$.

Existence:

Let $F'(A, X)$ be M_X^* (again with the notation $/$ instead of $*$ and with $\frac{x}{y}$ sometimes denoting x/y) with $0 = [0, 1]$ and $+$ defined as

$$\frac{x}{y} + \frac{z}{w} = \frac{xw + yz}{yw}. \quad (x, y, z, w \in A)$$

It is well-defined, since if $x', y', z', w' \in A$ and $x'/y' = x/y$ and $z'/w' = z/w$, then there are s_1, s_2, s_3, s_4 with $(s_1, s_1)(x, y) = (s_2, s_2)(x', y')$ and $(s_3, s_3)(z, w) = (s_4, s_4)(z', w')$, hence

$$\begin{aligned} \frac{x'w' + y'z'}{y'w'} &= \frac{s_2 s_4}{s_2 s_4} \cdot \frac{x'w' + y'z'}{y'w'} = \frac{s_2 x' s_4 w' + s_2 y' s_4 z'}{s_2 y' s_4 w'} = \frac{s_1 x s_3 w + s_1 y s_3 z}{s_1 y s_3 w} \\ &= \frac{s_1 s_3}{s_1 s_3} \cdot \frac{xw + yz}{yw} = \frac{xw + yz}{yw}. \end{aligned}$$

If $\varphi : (A, X_A) \rightarrow (B, X_B)$ is a \mathbf{C}' -arrow, then let $F'(\varphi)$ be $x/y \mapsto \varphi(x)/\varphi(y)$ (for $x, y \in A$). Let us check that it is well-defined. Let S_A be the multiplicative monoid generated by X_A and S_B the one generated by X_B . That φ is a \mathbf{C}' -arrow means that $\varphi(X_A) \subset X_B$, hence $\varphi(s) \in S_B$ for all $s \in S_A$, so that every

$\varphi(s)$ is $/$ -invertible in $F'(B, X)$. Suppose $x/y = x'/y'$ with $x', y' \in A$. That means that there are $s, s' \in S$ such that $(s, s)(x, y) = (s', s')(x', y')$ and hence

$$\begin{aligned} \frac{\varphi(x')}{\varphi(y')} &= \frac{\varphi(s')}{\varphi(s')} \cdot \frac{\varphi(x')}{\varphi(y')} = \frac{\varphi(s')\varphi(x')}{\varphi(s')\varphi(y')} = \frac{\varphi(s'x')}{\varphi(s'y')} = \frac{\varphi(sx)}{\varphi(sy)} \\ &= \frac{\varphi(s)\varphi(x)}{\varphi(s)\varphi(y)} = \frac{\varphi(s)}{\varphi(s)} \cdot \frac{\varphi(x)}{\varphi(y)} = \frac{\varphi(x)}{\varphi(y)}. \end{aligned}$$

F' is now clearly a functor. \square

According to the proof of this theorem, the construction of wheels of fractions, as given in section 1.2, is in a sense the most well-behaving one can have as a solution to the problem of making division a total function. We used the category **WW** as a tool for showing that, and we will not need this category anymore, neither the functor F' , as defined in the proof above. Instead, we use the full subcategory **Wheel**, with objects being the wheels (definition 1.1). We regard \odot as a functor $\mathbf{C}' \rightarrow \mathbf{Wheel}$, given by $(A, X) \mapsto \odot_X A$ and $\varphi \mapsto (x/y \mapsto \varphi(x)/\varphi(y))$ (it is easy to verify that wheels of fractions are indeed wheels). The full subcategory of wheels of fractions will be denoted by **WF**.

Note that $\odot_X A$ is defined for any commutative *semiring* A , not only for rings as was the case in section 1.2.

Example 3.12. $\odot_{\emptyset} \mathbb{N}$ is the initial object in **Wheel**.

Example 3.13. $\odot_{\mathbb{N} \setminus \{0\}} \mathbb{N}$ is the wheel which is obtained from the semiring of non-negative rational numbers by adjoining two extra elements $1/0$ and $0/0$.

The next proposition shows that the wheels of fractions are special among the wheels and the theorem that follows states that these rules characterize wheels of fractions completely up to isomorphism.

Proposition 3.14. *Any wheel of fractions is a wheel (definition 1.1 on page 5) and it holds in it that*

1. $\forall x \exists y \exists z : x = y/z \wedge 0y = 0z = 0$,
2. $\forall x \forall y \forall x' \forall y' : (x/y = x'/y' \wedge 0x = 0y = 0x' = 0y' = 0 \Rightarrow \exists z : x = zx' \wedge y = zy' \wedge 0z = 0/z = 0)$.

Proof. That a wheel of fractions is a wheel is straight-forward to check.

To show $\forall x \exists y \exists z : x = y/z \wedge 0y = 0z = 0$, use that every element $x \in \odot_X A$ is of the form y/z , where $y, z \in A$.

See theorem 4.16 for the proof of 2. \square

Theorem 3.15 (completeness for wheels of fractions). *A wheel is isomorphic to a wheel of fractions if and only if it has the properties given in the previous proposition.*

Proof. See theorem 4.16. \square

Remark. We refer in the proofs to theorem 4.16 which is not yet given. There is no risk for circular reasoning, since we will not use these results for deriving others, just for constructions of some examples.

We state them now because it is good to know a complete axiomatization in order to understand the class of structures under consideration.

4 Wheels

Since the class of wheels is defined by a set of equations (it is an ‘equational class’), it follows that substructures of wheels are wheels (‘subwheels’), that homomorphic images of wheels are wheels, and that products of wheels are wheels (for formal definitions, see some textbook in universal algebra, e.g. [BS81, Grä79]). It is also complete and cocomplete as a category (see [Bor94, p. 138]).

The corresponding is however not true for wheels of fractions: there are such which have subwheels or quotients that are not isomorphic to any wheel of fractions.

Example 4.1. Let A be the ring of functions $\mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{Z}$, with the operations defined point-wise. Let \mathbf{x} be the element $[\pi_1, \pi_2]$ of $\odot_{\emptyset} A$, with $\pi_1(x, y) = x$ and $\pi_2(x, y) = y$. Let H be the subwheel generated by \mathbf{x} , i.e., the elements of H are those which are given by terms composed of (zero or more occurrences of) \mathbf{x} and the operations. We will show that $\forall x \exists y \exists z : x = y/z \wedge 0y = 0z = 0$ fails to hold in H . We do that by showing that for any $y, z \in H$ with $0y = 0z = 0$, we have $\mathbf{x} \neq y/z$.

Every element $[f, g]$ of $\odot_{\emptyset} A$ defines a function $\mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{Z} \times \mathbb{Z}$ by $[f, g](x, y) = (f(x, y), g(x, y))$. \mathbf{x} defines the identity function. Note that

$$\begin{aligned} \mathbf{x}(0, 0) &= (0, 0) \\ (\mathbf{x} + [f, g])(0, 0) &= (0, 0) \\ (\mathbf{x} \cdot [f, g])(0, 0) &= (0, 0) \\ (/ \mathbf{x})(0, 0) &= (0, 0) \end{aligned}$$

so that the elements of H which are given by terms in which \mathbf{x} occurs at least once, define functions which takes the value $(0, 0)$ in at least one point.

Suppose $0y = 0z = 0$ and $y, z \in H$. Take f, g such that $y = [f, g]$. Then $0y = [0, g] = [0, 1]$, but then $g = 1$ and hence $y = [f, 1]$, so that y defines a function that does not take $(0, 0)$ as value. Hence y is described by a term which does not contain \mathbf{x} . The same argument applies to z . But then y/z is described by a term which does not contain \mathbf{x} , so that y/z defines a constant function. Then $\mathbf{x} \neq y/z$, since the function defined by \mathbf{x} is not constant.

Hence H is not isomorphic to any wheel of fractions.

Example 4.2. Let $H = \odot_{\{1,3\}}(\mathbb{Z}/4\mathbb{Z})$ and define the congruence relation

$$x \equiv y \quad \text{means} \quad x = y \vee x, y \in \{0/0, 0/2, 2/0, 2/2\}.$$

Then the rule $\forall x, y, x', y' : (x/y = x'/y' \wedge 0x = 0y = 0x' = 0y' = 0 \Rightarrow \exists z : x = zx' \wedge y = zy' \wedge 0z = 0/z = 0)$ fails in H/\equiv , since $0/0 = 2/2$, but $2 \neq z0$ for every z .

Hence H/\equiv is not isomorphic to any wheel of fractions.

Hence, the class of wheels isomorphic to wheels of fractions is not axiomatizable by equations. It is therefore not an accident that we needed non-equational rules to characterize (up to isomorphism) the wheels of fractions among the wheels.

It is not satisfactory to work in a class that is not closed under such operations, sooner or later one is lead to consider constructions like those given above.

Also, it is not satisfactory to succeed in making division total, but introducing partiality of important class operations.

That is the reason why we choose to study wheels, instead of wheels of fractions only.¹⁰

4.1 Basic properties

We devote some pages to an abstract study of wheels (definition 1.1, p. 5). Wheels of fractions are studied more in section 4.2.

Remember that the distributivity rule (3) for wheels looks different from the one for rings (or semirings). However, when $0z = 0$, it gives that $(x + y)z = xz + yz$. We emphasize this fact now once and for all, in the following we will apply it freely in computing things like $0(x + y) = 0x + 0y$.

Proposition 4.3 (derived rules). *The following rules hold in wheels:*

$$/1 = 1 \tag{9}$$

$$0x + 0y = 0xy \tag{10}$$

$$(0/0)x = 0/0 \tag{11}$$

$$x/x = 1 + 0x/x \tag{12}$$

$$xz = yz \Rightarrow x + 0z/z = y + 0z/z. \tag{13}$$

Proof. (9) Since $/$ is an involution. To repeat the argument: $/1 = 1/1 = //1/1 = /(1 \cdot 1) = //1 = 1$.

$$(10) \quad 0x + 0y \stackrel{(6)}{=} (0 + 0y)x = 0xy.$$

$$(11) \quad (0/0)x = 0x/0 \stackrel{(10)}{=} 0x + 0/0 \stackrel{(8)}{=} 0/0.$$

$$(12) \quad x/x = (0 + x \cdot 1)/x \stackrel{(4)}{=} 0/x + 1 + 0x \stackrel{(10)}{=} 1 + 0x/x.$$

(13) One should really learn to do without implicational rules like this, but we state it since it is characteristic for computations in wheel algebra.

$$\begin{aligned} xz &= yz \\ xz/z &= yz/z \\ x(1 + 0z/z) &= y(1 + 0z/z) && \text{by (12)} \\ x + 0z/z &= y + 0z/z. && \text{by (6)} \end{aligned}$$

□

Note that by (8) and (11), a term is equal to ‘0/0’ if it contains ‘0/0’ as a subterm.

A notion of negation can sometimes be introduced. If H is a wheel and $a \in H$ with $0a = 0$ and $a^2 = 1$, then $- : x \mapsto ax$ is an additive involution since $a(x + y) = (x + y)a + 0a = xa + ya = ay + ax$ and $aa x = 1x = x$. There are at least three natural ways in which such an a can appear.

1. $a = 1$. This gives the identity (thus not an interesting involution to use).

¹⁰See section 4.3 for further comments on the choice of axioms

2. If $H = \odot_X A$ with A a semiring⁻. Then $a = -1$ gives an involution which extends the $-$ of A . If A is a ring, then this case coincides with the next.
3. If a is a solution to the equation $1+x = 0$, then $0a = 1 \cdot 0 + a \cdot 0 = (1+a)0 = 0 \cdot 0 = 0$ and $a^2 = 1+a+a^2 = 1+(1+a)a+0a = 1+0a+0a = 1+0+0 = 1$. With this choice of a , we have the nice formula

$$x - x = 0x^2,$$

$$\text{since } x - x = 1x + ax = (1+a)x + 0x = 0x + 0x = 0x^2.$$

The wheel of fractions $\odot_{\emptyset}(\mathbb{Z}/2\mathbb{Z})$ is the (up to isomorphism unique) smallest non-trivial wheel, since it contains only four elements:

Proposition 4.4. *If any two of the elements 0 , 1 , $/0$ and $0/0$ are equal in a wheel H , then H is trivial.*

Proof. There are six cases to go through.

- | | |
|------------|--|
| $0 = 1$ | Then $1 = 1/1 = 0/0$. See below for the rest. |
| $0 = /0$ | Then $0 = 0 \cdot 0 = 0/0$. See below. |
| $0 = 0/0$ | Then $x = 0 + x = 0/0 + x = 0/0$. |
| $1 = /0$ | Then $1 = 1/1 = /0//0 = 0/0$. See below. |
| $1 = 0/0$ | Then $x = 1x = (0/0)x = 0/0$. |
| $/0 = 0/0$ | Then $x = 0 + x = /0 + x = /(0/0) + x = 0/0 + x = 0/0$. |

□

The two notions of reciprocal and multiplicative inverse are connected by the operations $0, +$ as follows.

Proposition 4.5. *Let x be a unit.*

1. $x^{-1} + 0/x = /x + 0x^{-1}$.
2. $x^{-1} = /x + 0\frac{x^{-1}}{x-1}$.
3. $/x = x^{-1} + 0\frac{x}{x}$.

Proof. 1. $x^{-1} + 0/x = x^{-1}/(xx^{-1}) + 0/x = /x + 0x^{-1}/(xx^{-1}) = /x + 0x^{-1}$.

2. Add $0/(x^{-1})$ to each side of 1.

3. Add $0x$ to each side of 1.

□

Note also that if xy is $/$ -invertible, then x and y are units with $x^{-1} = y/(xy) = /x + 0y/y$ and $y^{-1} = x/(xy) = /y + 0x/x$.

We may construct a right adjoint to the functor \odot in the following steps.

Definition 4.6. *Let H be a wheel. Then define the following subsets:*

$$\begin{aligned} \mathcal{R}H &= \{x \mid 0x = 0\}, \\ \mathcal{S}H &= \{x \mid 0x = 0/x = 0\}. \end{aligned}$$

Proposition 4.7. $\mathcal{R}H$ is a semiring under the operations $0, 1, +$ and \cdot . $\mathcal{S}H$ is a multiplicative group in $\mathcal{R}H$.

Proof. Suppose that $x, y \in \mathcal{R}H$. Then

1. $0(x + y) = 0(x + y) + 0 \cdot 0 = 0x + 0y = 0 + 0 = 0$,
2. $0(xy) = (0x)y = 0y = 0$.

Thus $\mathcal{R}H$ is closed under addition and multiplication. $0 \in \mathcal{R}H$ by (5) and $1 \in \mathcal{R}H$ since 1 is the identity in the multiplicative monoid.

That $\mathcal{S}H$ is closed under multiplication follows from the fact that if $x, y \in \mathcal{S}H$, then $0/(xy) = (0/y)/x = 0/x = 0$. We have $1 \in \mathcal{S}H$ by (9). Given an element $x \in \mathcal{S}H$, then $/x \in \mathcal{S}H$ (since $/$ is an involution) and it is the multiplicative inverse of x by the computation $x/x = 1+0x/x = 1+0/x = 1$. \square

We will use the notation $\mathcal{R}H$ also for the semiring it becomes when the operations $0, 1, +, \cdot$ are induced by H . It will be called ‘the semiring of H ’ or simply ‘the semiring’ when the context is clear. It is the largest semiring in H since $0x = 0$ holds in any semiring.

Remarks. We showed that every element of $\mathcal{S}H$ is $/$ -invertible. On the other hand, $\mathcal{S}H$ contains every $/$ -unit that is inside the subset $\mathcal{R}H$, since if x is $/$ -invertible and $0x = 0$, then $0/x = 0x/x = 0 \cdot 1 = 0$.

An element of $\mathcal{R}H$ is a unit in $\mathcal{R}H$ iff it is a unit in H . To see this, note that if $x \in \mathcal{R}H$ and $xy = 1$, then also $y \in \mathcal{R}H$, since $0y = 0xy = 0 \cdot 1 = 0$.

The question whether $x \in H$ is a unit or not is partially answered by the information whether or not $x \in \mathcal{R}H$ and/or $/x \in \mathcal{R}H$:

$x \in \mathcal{R}H?$	$/x \in \mathcal{R}H?$		x unit?	x $/$ -unit?
No	No	\Rightarrow	Maybe	Maybe
Yes	No	\Rightarrow	Maybe	No
No	Yes	\Rightarrow	Maybe	No
Yes	Yes	\Rightarrow	Yes	Yes

Let us consider $\eta_{(A,X)} : x \mapsto [x, 0]$ as a \mathbf{C}' -arrow $(A, X) \rightarrow (\mathcal{R} \odot_X A, \mathcal{S} \odot_X A)$, which we may do since, if $x \in X$, then $0/\eta_{(A,X)}(x) = [0, 1]/[x, 1] = [0, x] = [0, 1] = 0$, so that $\eta_{(A,X)}(X) \subset \mathcal{S} \odot_X A$.

Let $(\mathcal{R}, \mathcal{S})$ be the functor **Wheel** $\rightarrow \mathbf{C}'$ given by $(\mathcal{R}, \mathcal{S})(H) = (\mathcal{R}H, \mathcal{S}H)$ and $(\mathcal{R}, \mathcal{S})(\varphi)$ being the restriction of φ to $\mathcal{R}H$. This works since if $0x = 0$, then $0\varphi(x) = \varphi(0)\varphi(x) = \varphi(0x) = \varphi(0) = 0$ (thus φ actually maps the semiring of the domain into the semiring of the codomain), and if in addition $0/x = 0$, then $0/\varphi(x) = \varphi(0)/\varphi(x) = \varphi(0/x) = \varphi(0) = 0$ (thus $(\mathcal{R}, \mathcal{S})(\varphi)$ behaves as it should with respect to \mathcal{S}).

Theorem 4.8. The functor $\odot : \mathbf{C}' \rightarrow \mathbf{Wheel}$ is left adjoint to $(\mathcal{R}, \mathcal{S}) : \mathbf{Wheel} \rightarrow \mathbf{C}'$ with unit η and counit ϵ , given by $\epsilon_H : [x, y] \mapsto x/y$.

Proof. ϵ_H is well-defined since if $[x, y] = [x', y']$, then there exist $s_1, s_2 \in \mathcal{S}H$ such that $(s_1, s_1)(x, y) = (s_2, s_2)(x', y')$ and hence $x/y = (x/y)(1 + 0s_1/s_1) =$

$(s_1x)/(s_1y) = (s_2x')/(s_2y') = (x'/y')(1 + 0s_2/s_2) = x'/y'$. It is easily seen that ϵ is a natural transformation.

By basic category theory, the result follows if we show that for any \mathbf{C} -object (A, X) and any wheel H , we have¹¹

$$(\epsilon \odot)_{(A, X)}(\odot \eta)_{(A, X)} = \mathbf{Id}_{\odot_X A}$$

and

$$((\mathcal{R}, \mathcal{S})\epsilon)_H(\eta(\mathcal{R}, \mathcal{S}))_H = \mathbf{Id}_{(\mathcal{R}H, \mathcal{S}H)}.$$

$$(\epsilon \odot)_{(A, X)} : \odot_{S \odot_X A} \mathcal{R} \odot_X A \rightarrow \odot_X A \text{ by } [[x, y], [z, w]] \mapsto [xw, yz]$$

$$(\odot \eta)_{(A, X)} : \odot_X A \rightarrow \odot_{S \odot_X A} \mathcal{R} \odot_X A \text{ by } [x, y] \mapsto [[x, 1], [y, 1]]$$

$$((\mathcal{R}, \mathcal{S})\epsilon)_H : (\mathcal{R} \odot_{SH} \mathcal{R}H, \mathcal{S} \odot_{SH} \mathcal{R}H) \rightarrow (\mathcal{R}H, \mathcal{S}H) \text{ by } [x, y] \mapsto x/y$$

$$(\eta(\mathcal{R}, \mathcal{S}))_H : (\mathcal{R}H, \mathcal{S}H) \rightarrow (\mathcal{R} \odot_{SH} \mathcal{R}H, \mathcal{S} \odot_{SH} \mathcal{R}H) \text{ by } x \mapsto [x, 1]$$

Then

$$(\epsilon \odot)_{(A, X)}(\odot \eta)_{(A, X)}([x, y]) = (\epsilon \odot)_{(A, X)}([[x, 1], [y, 1]]) = [x \cdot 1, y \cdot 1] = [x, y]$$

and

$$((\mathcal{R}, \mathcal{S})\epsilon)_H(\eta(\mathcal{R}, \mathcal{S}))_H(x) = ((\mathcal{R}, \mathcal{S})\epsilon)_H([x, 1]) = x/1 = x.$$

□

4.2 Applications to wheels of fractions

4.2.1 The semiring in a wheel of fractions

The following fact will be used freely without reference to the proposition.

Proposition 4.9. *If $x \in A$, then $0/x = 0$ in $\odot_X A$ if and only if $x \in \tilde{S}$, which is the divisional closure of S , which is the multiplicative monoid generated by X .*

Proof. Proposition 2.15 gives us that $x/x = 1$ is equivalent to $x \in \tilde{S}$. We need to prove $x/x = 1 \iff 0/x = 0$ under the condition $0x = 0$.

Suppose $x/x = 1$. Then $0/x = 0x/x = 0 \cdot 1 = 0$.

Suppose $0/x = 0$. Then $x/x = 1 + 0x/x = 1$. □

The elements of $\mathcal{R} \odot_X A$ are of a very special form:

Lemma 4.10. *$x \in \mathcal{R} \odot_X A$ iff $x = r/s$ for some $r \in A$, $s \in S$, with S the multiplicative monoid generated by X .*

Proof. If $x = r/s$, then $0x = 0r/s = 0/s = 0$. On the other hand, suppose that $0x = 0$, say with $x = y/z$ ($y, z \in A$). Then $0/z = 0x = 0$, so that $z \in \tilde{S}$, say $s'z \in S$. With $r = s'y, s = s'z$, we get $r/s = (s'y)/(s'z) = (s'/s')(y/z) = x$. □

In commutative algebra and algebraic geometry, one often uses the ring of fractions $S^{-1}A$, where S is a multiplicative submonoid of a commutative ring A (often the complement of a prime ideal). One defines $S^{-1}A$ as the monoid A_S together with $0 = [0, 1]$ and $[x, s] + [x', s'] = [xs' + x's, ss']$. As is seen, this construction is similar to ours. In fact, we have factorized it in two parts: as \odot followed by \mathcal{R} .

¹¹The naturalness must also be checked, but our constructions are obviously natural.

Theorem 4.11. *If A is a commutative ring with identity and S a multiplicative submonoid, then $S^{-1}A \cong \mathcal{R}_{\odot_S}A$ by the isomorphism $r/s \mapsto r/s$.*

Proof. The mapping $\hat{\eta}_{(A,S)} : A_S \rightarrow \mathcal{R}_{\odot_S}A, r/s \mapsto r/s$ is an injective morphism of the multiplicative monoids by theorem 2.10. It is easy to see that it preserves 0 and + too. It is surjective by the previous lemma and thus a bijection. $S^{-1}A$ is a ring so $\hat{\eta}_{(A,S)}$ is an isomorphism of rings. \square

Any wheel of fractions is (according to proposition 2.14) of the form $\odot_S A$ with S being closed under division. For that presentation, we have a stronger version of the previous theorem.

Theorem 4.12. *$S^{-1}A$ and $\mathcal{R}_{\odot_S}A$ are identical iff S is closed under division.*

Proof. Take $[x, y] \in \mathcal{R}_{\odot_S}A$. Then $0/y = 0x/y = 0$, so that $y \in \tilde{S}$. If S is closed under division, we get $y \in S$, so that $(x, y) \in A \times S$. Hence $\mathcal{R}_{\odot_S}A = S^{-1}A$ in that case.

Suppose instead that $\mathcal{R}_{\odot_S}A$ and $S^{-1}A$ are identical as sets. Take $x \in \tilde{S}$; we shall prove $x \in S$. We have $(x, x) \in [x, x] = x/x$ but since $x/x \in \mathcal{R}_{\odot_S}A$ by the previous lemma, we get that $(x, x) \in A \times S$ when $\mathcal{R}_{\odot_S}A$ and $S^{-1}A$ are identical as sets. \square

We summarize the results above in the following corollary. It states one isomorphism and two identities:

Corollary 4.13. *Let A be a semiring, S a multiplicative monoid in it, \tilde{S} its divisional closure. Then $S^{-1}A \cong \mathcal{R}_{\odot_S}A = \mathcal{R}_{\odot_{\tilde{S}}}A = \tilde{S}^{-1}A$.*

Proof. By theorem 4.11, $S^{-1}A \cong \mathcal{R}_{\odot_S}A$. By theorem 2.14, $\mathcal{R}_{\odot_S}A = \mathcal{R}_{\odot_{\tilde{S}}}A$ and by theorem 4.12, $\mathcal{R}_{\odot_{\tilde{S}}}A = \tilde{S}^{-1}A$. \square

4.2.2 Obtaining /-invertible wheels of fractions

In the choice between $S^{-1}A$ and $\odot_S A$ in practical computing when A is a ring, it may seem to be a drawback of the latter that one has to distinguish between reciprocals and multiplicative inverses. One should then remember that $S^{-1}A$ and $\mathcal{R}_{\odot_S}A$ are isomorphic as rings and that the multiplicative inversion therefore is the same in both structures. The difficulty appears first at the moment when we start using the involution $/$. Hence, there is no real drawback, since one can stick to multiplicative inversion when that is more comfortable.

But that does not explain why we introduced the operation $/$. If finding inverses is what we want it for, then we would like to have no difference between reciprocals and inverses, i.e., we would like to have a /-invertible wheel, so that $x^{-1} = /x$ whenever x^{-1} is defined. By proposition 2.19, we obtain that if and only if \tilde{S} is saturated as a submonoid of the multiplicative monoid of A . The question is therefore if a given S can be replaced by one having saturated divisional closure.

Definition 4.14 (saturated closure). *The saturated closure \bar{S} of a submonoid S of a commutative monoid M is the set $\{x \in M \mid \exists y \in M : xy \in S\}$ (it is the smallest saturated submonoid of M that contains S).*

Since $\tilde{S} = \bar{S}$, \bar{S} is a submonoid with saturated divisional closure. In the following, \bar{S} will denote the saturated closure of S in the multiplicative monoid of A .

There is a well-known natural isomorphism $S^{-1}A \cong \bar{S}^{-1}A$ given by $r/s \mapsto r/\bar{s}$. Theorem 4.12 shows that $\bar{S}^{-1}A = \mathcal{R}_{\odot_{\bar{S}}A}$, so that the isomorphism is in fact $S^{-1}A \rightarrow \mathcal{R}_{\odot_{\bar{S}}A}$. Since $\odot_{\bar{S}}A$ is $/$ -invertible, it is in a sense a better extension of $S^{-1}A$ than $\odot_S A$ is.¹²

It is well-known that any ring A may be extended to its ‘total ring of fractions’ $S_0^{-1}A$, so that every multiplicatively cancelable element in A becomes a unit. However, when doing that, some structure of A can be lost. For instance, in extending \mathbb{Z} to \mathbb{Q} , one loses all non-trivial proper ideals. Hence, there is a reason for not always extending rings as much as one can. The construction of wheels of fractions have the advantage that it need not destroy any structure. The following corollary shows that a semiring can always be extended to a $/$ -invertible wheel in such a way that it can be retrieved using the \mathcal{R} operator.

Corollary 4.15. *Let U be the set of units in a semiring A . Then $\odot_U A$ is $/$ -invertible and it extends A in the sense that*

1. $\mathcal{R}\odot_U A$ is isomorphic to A ,
2. the operations $0, 1, +, \cdot$ of A are extended to corresponding operations of $\odot_U A$,
3. if A is a semiring⁻, then the involution $-$ of A is extended to an involution $-$ of $\odot_U A$.
4. the restriction of $/$ to A is the partial function \cdot^{-1} .

Proof. This is just a restatement of 2.20 and 2.21, together with information on how $+$ and $-$ are defined (see page 24 for the latter). \square

4.2.3 Axiomatizing the class of wheels of fractions

We now restate and prove the completeness theorem for the axioms of wheels of fractions. The symbols \mathcal{R} and \mathcal{S} simplify the notation compared to that used in proposition 3.14.

Theorem 4.16. *A wheel H is isomorphic to some wheel of fractions if and only if it has the following properties:*

$$\forall x \exists y, z \in \mathcal{R}H : x = y/z \quad (14)$$

$$\forall x, y, x', y' \in \mathcal{R}H : (x/y = x'/y' \Rightarrow \exists s \in \mathcal{S}H : x = sx' \wedge y = sy') \quad (15)$$

Proof. (14) holds in any $\odot_X A$ since every x is of the form y/z for some $y, z \in A$ so we have $x = y/z$ for the corresponding elements in $\mathcal{R}_{\odot_X A}$.

To show that (15) holds in $\odot_X A$, suppose that $x/y = x'/y' \wedge 0x = 0y = 0x' = 0y' = 0$. By lemma 4.10, we may assume that $x = a/s_1, y = b/s_2, x' = c/s_3,$

¹²It is true that $\mathcal{R}_{\odot_S A}$ and $\mathcal{R}_{\odot_{\bar{S}} A}$ are isomorphic as rings by the mapping $r/s \mapsto r/\bar{s}$, but not in general that $/$ is preserved under this isomorphism. Consider e.g. $\mathcal{R}_{\odot_{\{1,2,3,\dots\}}\mathbb{Z}}$ and $\mathcal{R}_{\odot_{\{\pm 1, \pm 2, \pm 3, \dots\}}\mathbb{Z}}$. They are isomorphic as rings, but $/(-1) = -1$ is true only in the latter. $(-1)^{-1} = -1$ is of course true whenever $-$ is obtained in one of the ways mentioned on page 24.

$y' = d/s_4$ with $a, b, c, d \in A$. Then $(s_2/s_1)a/b = x/y = x'/y' = (s_4/s_3)(c/d)$, so there exist s', s'' such that

$$(s', s')(s_2, s_1)(a, b) = (s'', s'')(s_4, s_3)(c, d).$$

Taking

$$s = \frac{s_3 s_4 s''}{s_1 s_2 s'}$$

(which is in $\mathcal{S}H$), we get

$$sx' = \frac{s_3 s_4 s''}{s_1 s_2 s'} \cdot \frac{c}{s_3} = \frac{s_3 s_4 s'' c}{s_1 s_2 s_3 s'} = \frac{s_3 s_2 s' a}{s_1 s_2 s_3 s'} = \frac{a}{s_1} = x.$$

and

$$sy' = \frac{s_3 s_4 s''}{s_1 s_2 s'} \cdot \frac{d}{s_4} = \frac{s_3 s_4 s'' d}{s_1 s_2 s_4 s'} = \frac{s_1 s_4 s' b}{s_1 s_2 s_4 s'} = \frac{b}{s_2} = y.$$

On the other hand, suppose that (14) and (15) are true in a wheel H . Then ϵ_H (defined in theorem 4.8) is surjective by (14) and injective by (15). Hence it is an isomorphism $\odot_{\mathcal{S}H} \mathcal{R}H \rightarrow H$. \square

4.2.4 Cancellation of zero-terms

In computing with (semi-)rings, one does not only use equational rules, but additional information, for instance, that a certain element a is ‘multiplicatively cancellable’ (\cdot -cancellable): $ax = ay \Rightarrow x = y$.

We have seen that any semiring can be viewed as a part of a wheel; technically, that any semiring A is embedded in $\odot_U A$ (where U is the set of units of A) and that an isomorphic copy of A can be recovered from $\odot_U A$ by the operator \mathcal{R} . It is important that we can transfer information about cancellation to the wheel.

Consider the equation $ax = ay$ in a semiring A . The information ‘ a is \cdot -cancellable’ lets us conclude that $x = y$. When A is considered as embedded in a wheel, it is instead natural to reason as follows (without any use of cancellation):

$$\begin{aligned} ax &= ay \\ ax/a &= ay/a \\ x + 0a/a &= y + 0a/a \\ x + 0/a &= y + 0/a. \end{aligned}$$

The question of cancellation now comes in since we need to know whether $0/a$ is *additively* cancellable ($+$ -cancellable). In this way, the problem of \cdot -cancellation is reduced to a problem of $+$ -cancellation. One might however suspect that there could be situations in which a is \cdot -cancellable in A , but $0/a$ is not $+$ -cancellable in $\odot_X A$. Fortunately, there are no such cases, according to the following proposition.

Proposition 4.17. *If a is \cdot -cancellable in A , then $0/a$ is $+$ -cancellable in $\odot_X A$.*

Proof. Suppose that a is \cdot -cancellable in A . Suppose also that $[x, y] + 0/a = [x', y'] + 0/a$, that is, that $[ax, ay] = [ax', ay']$. Then there are s, s' such that $sax = s'ax'$ and $say = s'ay'$ but since a is \cdot -cancellable, we conclude that $sx = s'x$ and $sy = s'y'$, hence $[x, y] = [x', y']$. \square

Suppose now that if $a \neq 0$ in A , then a is \cdot -cancellable, i.e., that $S_0 = A \setminus \{0\}$. We say (following [Gol99]) that such a semiring is *multiplicatively cancellative*.¹³ This is a maximality property, since if $a = 0$, then a cannot be \cdot -cancellable unless the semiring is trivial. A corresponding maximality property for wheels is $0x \neq x \Rightarrow (0/x \text{ is } +\text{-cancellable})$, since if $0x = x$, then $0/x$ can't be $+$ -cancellable unless the wheel is trivial ($y + 0/x = z + 0/x$ holds for any y and z if $0x = x$, since then $y + 0/x = y + 0/(0x) = 0/0 = z + 0/(0x) = z + 0/x$). These properties are connected as follows.

Proposition 4.18. 1. *If A is multiplicatively cancellative, then $0x \neq x \Rightarrow (0/x \text{ is } +\text{-cancellable})$ holds in $\odot_X A$.*

2. *If $0x \neq x \Rightarrow (0/x \text{ is } +\text{-cancellable})$ holds in H , then $\mathcal{R}H$ is multiplicatively cancellative.*

Proof. 1. We shall prove that $0/x$ is $+$ -cancellable whenever $0x \neq x$. Let $x = [x', x'']$ and assume $0x \neq x$. Suppose that $x' = 0$. Then $x = [0, x''] = [0, 1][0, x''] = 0x$, which is refuted, hence $x' \neq 0$. If A is multiplicatively cancellative, then x' is \cdot -cancellable. Hence $0/x'$ is $+$ -cancellable, but $0/x = 0/x'$ and we are done.

2. Take $x \in \mathcal{R}H$ with $x \neq 0$. Then $0x = 0 \neq x$, so that $0/x$ is $+$ -cancellable. We may then compute as follows:

$$\begin{aligned} xy &= xz \\ xy/x &= xz/x \\ y + 0x/x &= z + 0x/x \\ y + 0/x &= z + 0/x \\ y &= z. \end{aligned}$$

Note that we did not need to assume that $y, z \in \mathcal{R}H$. □

Example 4.19. Solve the equation $3x = 12$.

We have $x + 0/3 = 3x/3 = 12/3 = 3 \cdot 4/3 = 4 + 0/3$. That gives $x = 4$ whenever $0/3$ is $+$ -cancellable, which happens for instance if the wheel in question is $\odot_X \mathbb{Z}$, for any subset $X \subset \mathbb{Z}$, since 3 is \cdot -cancellable in \mathbb{Z} . In general we may not conclude more than $x + 0/3 = 4 + 0/3$, which is sound since the equation does not have a unique solution in general.

Ring algebra does not allow us to derive anything from $3x = 12$ when 3 is not \cdot -cancellable (e.g. when the equation have several solutions), but wheel algebra always lets us develop to $x + 0/3 = 4 + 0/3$, which can be used in further calculations: if x and $0/3$ appear in the left hand side of an equation, then we may use the axioms for zero-terms to move $0/3$ close to x , then replacing $x + 0/3$ by $4 + 0/3$ and finally move the term $0/3$ out using the rules for zero-terms again. If x appears, but not $0/3$, then we may add $0/3$ to both sides of the equation, hence introducing $0/3$ but eliminating an occurrence of x . Note that when such an introduction has been made, then one can use the introduced term over and

¹³The term 'integral' is unfortunately already used in another sense in the context of semi-rings [Gol99]. E.g. $\mathbb{Z}/6\mathbb{Z}$ is integral as a semiring in this sense.

over again, until all occurrences of x are eliminated. One may view these $0/3$ -terms as flags indicating that the result is worthless if e.g. $0 = 3$, since then the rule ' $x + 0/0 = 0/0$ ' can be used for reducing the equation to $0/0 = 0/0$.

Say that we want to compute the value of $x^2 + x$, given that $3x = 12$. We can't give a definite answer, since the value is not uniquely determined unless we specify which semiring we think of. What we can do is to compute $x^2 + x + 0/3$:

$$\begin{aligned} x^2 + x + 0/3 &= x^2 + 4 + 0/3 = x(x + 0/3) + 4 = x(4 + 0/3) + 4 \\ &= (x + 0/3)4 + 4 = (4 + 0/3)4 + 4 = 4 \cdot 4 + 4 + 0/3 \\ &= 20 + 0/3. \end{aligned}$$

If we get information at a later stage, letting us conclude that $0/3$ is +-cancellable, then we can get something explicit for $x^2 + x$. This might happen if we get more knowledge about the semiring we started with, but also if we pass to a quotient of that semiring. For instance, suppose that we conclude that we need the value modulo 16. We then pass to a quotient defined by $0 = 16$. In such a quotient, 3 is a unit, and hence $0/3$ is +-cancellable.¹⁴ We conclude that $x^2 + x = 20 = 4 \pmod{16}$.

We may view the wheel alternative as computing modulo everything at the same time.

4.3 Equational logic for wheels of fractions

One may ask if our definition of 'wheel' is 'the right one'. Are the axioms chosen ad hoc, or do they axiomatize a class of structures that is in some sense naturally delimited?

There is some evidence that we can't improve the definition:

1. The wheels of fractions are wheels according to our definition.
2. The axioms are a set of equations. That is by Birkhoff's theorem [Bir35] necessary for the axiomatized class to be closed under substructures, quotients, products and isomorphisms. Further, by 'compactness for equational logic', if it is finitely axiomatizable in first order predicate logic, then it is finitely axiomatizable by equations. Hence, there is no need to consider other candidates than equations as axioms.
3. The axioms are finitely many.
4. We have empirical experience that the axioms are precisely what we need to make natural computations.

On the other hand, there is some counterevidence too: There are identities, like $(x + 1)^2 = x(x + 2) + 1$, that hold in any wheel of fractions but does not seem to be provable from our axioms. But $(x + 1)^2 = x(x + 2) + 1$ is not a very natural identity and may be thought of as valid in wheels of fractions 'by accident': that it just 'happens to be' and that there is nothing conceptual that suggests that it should be valid in all wheels.¹⁵

¹⁴If x is a unit, then $0x$ and $0/x$ are +-cancellable since $0x + 0x^{-1} = 0xx^{-1} = 0$ and $0/x + 0/(x^{-1}) = 0/(xx^{-1}) = 0$ in that case.

¹⁵However, by completeness, (14) (see theorem 4.16) suffices (together with wheel axioms) for derivation of any identity that is valid in all wheels of fractions, since every wheel which has this property is a homomorphic image of a wheel of fractions. This is because ϵ_H (defined in theorem 4.8) is surjective by (14).

Anyway, one might wish to include more axioms for wheels, perhaps those needed to prove every identity that is valid in the class **WF** of wheels of fractions. By Birkhoff's theorem, we should have equational axioms only, so that the class axiomatized is still closed under products, substructures and quotients. This amounts to axiomatizing the class $V(\mathbf{WF})$, called 'the variety generated by **WF**', consisting of structures that can be obtained in finitely many steps from wheels of fractions, taking products, substructures, quotients and isomorphic copies.¹⁶

We will study the class $V(\mathbf{WF})$, but also simultaneously the following. A *wheel with negation* is a wheel with an additional constant -1 such that $1 + (-1) = 0$. By the discussion on page 24, such a constant determines an additive involution by $-x = (-1)x$ and it holds that $x - x = 0x^2$, hence the semiring in such a wheel is in fact a ring. A wheel of fractions with negation is obtained for instance when the semiring used in the construction is a ring. We denote by **Wheelneg** the class of wheels with negation and by **WFneg** the class of wheels of fractions with negation. A morphism of wheels with negation should preserve -1 , but that is automatically satisfied by any morphism of wheels: suppose H and H' are wheels with negation and $\varphi : H \rightarrow H'$ a wheel morphism. Then it holds in H' that $\varphi(-1) = 1 + (-1) + \varphi(-1) = -1 + \varphi(1 + (-1)) = -1 + \varphi(0) = -1$.

We will emphasize on the classes **Wheel**, **WF** and $V(\mathbf{WF})$ in the following, but the treatment of **Wheelneg**, **WFneg** and $V(\mathbf{WFneg})$ is precisely analogous, just replace \mathbb{N} by \mathbb{Z} , $\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n]$ by $\mathbb{Z}[\mathbf{x}_1, \dots, \mathbf{x}_n]$, $\odot_{\emptyset}\mathbb{N}$ by $\odot_{\emptyset}\mathbb{Z}$, "wheel axioms" by "wheel axioms and ' $1 + (-1) = 0$ '", "semirings" by "rings with identity", and so on. One has to check some details of negation in the proofs, but that is straight-forward and omitted for the purpose of readability (we include a remark at a crucial point).

Definition 4.20. A term $p(x_1, \dots, x_n)$ is a polynomial term if it contains no occurrence of $/$. We say that such terms are equivalent if they can be proved to be equal using the axioms for commutative semirings. This is an equivalence relation and its equivalence classes are called polynomials.

Example 4.21. The polynomial terms ' $(x + 1)(x + 1)$ ' and ' $(xx) + ((2x) + 1)$ ' are not identical, but equivalent. Hence they define the same polynomial.

When wheels with negation are treated, polynomial terms are equivalent if they can be proved to be equal using axioms for commutative rings with identity.

Theorem 4.22. The following are equivalent for a wheel H .

1. $H \in V(\mathbf{WF})$.
2. $H \in V(\{\odot_{\emptyset}\mathbb{N}\})$.
3. Any identity which is valid in all wheels of fractions is valid in H too.
4. Any identity which is valid in $\odot_{\emptyset}\mathbb{N}$ is valid in H too.
5. H is isomorphic to a quotient of a subwheel of a power of $\odot_{\emptyset}\mathbb{N}$.
6. H is isomorphic to a quotient of a subwheel of $\odot_{\emptyset}(\mathbb{N}^I)$, where \mathbb{N}^I is the semiring of functions $I \rightarrow \mathbb{N}$, for some set I .

¹⁶See some textbook in universal algebra, e.g. [BS81, Grä79].

Proof. We first prove $1 \iff 2 \iff 3 \iff 4$ by showing $1 \Rightarrow 3 \Rightarrow 4 \Rightarrow 2 \Rightarrow 1$. The step $4 \Rightarrow 2$ follows from Birkhoff's theorem and the other steps are obvious, except for $3 \Rightarrow 4$ which needs a proof.

Suppose $t(x_1, \dots, x_n) = s(x_1, \dots, x_n)$ is valid in $\odot_{\emptyset}\mathbb{N}$. We shall prove that it is valid in all wheels of fractions. Substitute $[x'_i, x''_i]$ for every x_i in the identity and compute the values of each side. The result is $[p(x_1, \dots, x_n), q(x_1, \dots, x_n)] = [p'(x_1, \dots, x_n), q'(x_1, \dots, x_n)]$ for polynomials p, p', q, q' with coefficients from \mathbb{N} . That $t(x_1, \dots, x_n) = s(x_1, \dots, x_n)$ is valid in $\odot_{\emptyset}\mathbb{N}$ means that $(p(x_1, \dots, x_n), q(x_1, \dots, x_n))$ and $(p'(x_1, \dots, x_n), q'(x_1, \dots, x_n))$ takes the same values if we substitute natural numbers for the variables. But then $p = p'$ and $q = q'$ as polynomials, so that the identity is valid in any wheel of fractions.

A theorem by Tarski gives $2 \Rightarrow 5$ (see e.g. [BS81]) and $5 \Rightarrow 2$ is obvious.

To establish $5 \iff 6$, we need to show $(\odot_{\emptyset}\mathbb{N})^I \cong \odot_{\emptyset}(\mathbb{N}^I)$. We get this from the identity $(\mathbb{N} \times \mathbb{N})^I \cong \mathbb{N}^I \times \mathbb{N}^I$ for the multiplicative monoid of \mathbb{N} . It is obvious that this isomorphism preserves also 0 and +. \square

One says that $\odot_{\emptyset}\mathbb{N}$ is *generic* for $V(\mathbf{WF})$, because of the equivalence $1 \iff 4$. It is moreover the \mathbf{WF} -free wheel over \emptyset . In general, if K is a class of wheels, then the K -free wheel over a set X is the (up to isomorphism unique) wheel $F_K(X)$ with the property that for any function $X \rightarrow H$, where $H \in K$, there is a morphism $F_K(X) \rightarrow H$ that extends it. It is easily seen from Birkhoff's construction that $F_K(X) = F_{V(K)}(X)$ and in our case we have $V(\mathbf{WF}) = V(\{\odot_{\emptyset}\mathbb{N}\})$, so that $F_{\mathbf{WF}}(X) = F_{\{\odot_{\emptyset}\mathbb{N}\}}(X)$ and the latter is (again by Birkhoff's construction) a subwheel of some wheel of the form $(\odot_{\emptyset}\mathbb{N})^I$ — which, as we proved in the theorem, is isomorphic to $\odot_{\emptyset}(\mathbb{N}^I)$. One may consider it as the wheel of wheel-terms with identification made when two terms yield identical values in $\odot_{\emptyset}\mathbb{N}$. Hence an equation $t(x_1, \dots, x_n) = s(x_1, \dots, x_n)$ is *valid* in \mathbf{WF} if and only if $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is *true* in $F_{\mathbf{WF}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$.

Remark. Note how we use the terms ‘valid’ and ‘true’. We only apply ‘true’ to a closed formula (which in our case means a formula in which no variable occur, since we consider quantifier-free formulas only). An open formula is *valid* if it becomes true after *any* substitution with elements from *any* of the structures considered.

The corresponding theorem for wheels with negation reads as follows.

Theorem 4.23. *The following are equivalent for a wheel H with negation ($\odot_{\emptyset}\mathbb{Z}$ is here considered as a wheel with negation, that means for instance that $\odot_{\emptyset}\mathbb{N}$ is not considered to be a substructure of it).*

1. $H \in V(\mathbf{WFneg})$.
2. $H \in V(\{\odot_{\emptyset}\mathbb{Z}\})$.
3. Any identity which is valid in all wheels of fractions with negation is valid in H too.
4. Any identity which is valid in $\odot_{\emptyset}\mathbb{Z}$ is valid in H too.
5. H is isomorphic to a quotient of a subwheel of a power of $\odot_{\emptyset}\mathbb{Z}$.
6. H is isomorphic to a quotient of a subwheel of $\odot_{\emptyset}(\mathbb{Z}^I)$, where \mathbb{Z}^I is the ring of functions $I \rightarrow \mathbb{Z}$, for some set I .

Proof. As the previous proof. \square

Since validity in **WF** is equivalent to truth in $F_{\mathbf{WF}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$, we get a completeness result for **WF** if we give a complete method for deriving formulas which are true in $F_{\mathbf{WF}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$. In this section, we will study such derivations, and similar ones.

We work in a language which is obtained from the language of wheels by adjoining constant symbols $\mathbf{x}_1, \dots, \mathbf{x}_n$ (and when wheels with negation are considered, the language has an additional constant symbol ‘ -1 ’). In the following, \mathbf{x}_i will denote

1. a constant symbol in the language,
2. an element of $F_K(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$ for several choices of K ,
3. an element of $\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n]$,
4. an element of $\odot_{\emptyset}(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$,

so the reader should be careful to observe which is the case for every occurrence. As a consequence, when t is a term, $t(\mathbf{x}_1, \dots, \mathbf{x}_n)$ can be interpreted as a closed term in the language or as an element of $F_K(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$ for several choices of K , or as an element of $\odot_{\emptyset}(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$.

When we say things like “ $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ can be proved using the wheel axioms together with an axiom ‘ $0\mathbf{x}_i = 0$ ’ for each i ”, we mean that there is a formal derivation of the formula $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ in the system of equational logic with the axioms indicated (here, ‘axioms’ should be interpreted in the sense of formal systems). The formula ‘ $1 + (-1) = 0$ ’ is an axiom when we work with wheels with negation.

Theorem 4.24 (completeness). *The equation $t(x_1, \dots, x_n) = s(x_1, \dots, x_n)$ is valid in **WF** (and hence valid in $V(\mathbf{WF})$) if and only if $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is provable using the axioms for wheels together with the axioms ‘ $0\mathbf{x}_i$ is +-cancellable’ ($i = 1, \dots, n$).*

Proof. Validity of the equation $t(x_1, \dots, x_n) = s(x_1, \dots, x_n)$ is equivalent to the truth of $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ in $F_{\mathbf{WF}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$ (by definition of the latter wheel). Lemma 4.47 will show that if the latter equation is true, then it is provable using the axioms for wheels together with the axioms ‘ $0\mathbf{x}_i$ is +-cancellable’.

On the other hand, if it is provable by that technique, then it is true by lemma 4.39. \square

Before we state and prove some technical lemmas, we state the following theorems, which are alternatives to the previous ones. By distinguishing between ‘balanced’ and ‘unbalanced’ equations, they make more precise statements than the previous ones.

Definition 4.25. *An equation is balanced if each variable occurs equally many times on each side. Otherwise, it is unbalanced.*

Note that all axioms for wheels except ‘ $x + 0/0 = 0/0$ ’ are balanced. Remember that x occurs n times in ‘ x^n ’, since ‘ x^n ’ is just a short notation for a product of n x ’s. A warning: the equation ‘ $0\mathbf{x}_i = 0$ ’ is balanced, since \mathbf{x}_i is *not*

a variable in our formal sense, but a constant symbol, interpreted as a special element of the structure considered in the context.

Theorem 4.26. *An unbalanced equation $t(x_1, \dots, x_n) = s(x_1, \dots, x_n)$ is valid in \mathbf{WF} (and hence valid in $V(\mathbf{WF})$) if and only if $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = 0/0$ and $s(\mathbf{x}_1, \dots, \mathbf{x}_n) = 0/0$ are true in $\odot_\emptyset(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$.*

Proof. Use lemma 4.49 and the equivalence between validity and truth. \square

Example 4.27. The formula $1 + \frac{1}{x} = \frac{x+1}{x}$ is valid in $\mathbb{Q} \setminus \{0\}$. Is it perhaps valid in all wheels of fractions? The answer is no, since it is not balanced, and one easily checks that $1 + \frac{1}{\mathbf{x}} \neq \frac{\mathbf{x}+1}{\mathbf{x}}$ in $\odot_\emptyset(\mathbb{N}[\mathbf{x}])$. A balanced analogue is

$$1 + \frac{1}{x} + 0x = \frac{x+1}{x},$$

which is easily proved by the wheel axioms.

Theorem 4.28. *A balanced equation $t(x_1, \dots, x_n) = s(x_1, \dots, x_n)$ is valid in \mathbf{WF} (and hence valid in $V(\mathbf{WF})$) if and only if $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is true in $\odot_\emptyset(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$.*

Proof. If $t(x_1, \dots, x_n) = s(x_1, \dots, x_n)$ is valid, $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is clearly true.

For the other direction, use lemma 4.50 and the equivalence between validity and truth. \square

Example 4.29. $\frac{(x+1)^2}{x} = \frac{x(x+2)+1}{x}$ is valid in all wheels of fractions since it is balanced (x occurs three times on each side) and $\frac{(\mathbf{x}+1)^2}{\mathbf{x}} = \frac{\mathbf{x}(\mathbf{x}+2)+1}{\mathbf{x}}$ is true in $\odot_\emptyset(\mathbb{N}[\mathbf{x}])$.

Example 4.30. $(x+y)(x-y) = x^2 - y^2$ is valid in all wheels of fractions with negation. This follows since it is balanced and $(\mathbf{x}_1 + \mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{x}_1^2 - \mathbf{x}_2^2$ is true in $\odot_\emptyset(\mathbb{Z}[\mathbf{x}_1, \mathbf{x}_2])$.

These two results give a completeness result for \mathbf{WF} together with the following theorem:

Theorem 4.31. *The equation $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is true in the wheel $\odot_\emptyset(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$ if and only if it is provable by the wheel axioms together with the rules ‘ $0\mathbf{x}_i = 0$ ’ for each $i = 1, \dots, n$.*

Proof. $0\mathbf{x}_i = 0$ is evidently true in $\odot_\emptyset(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$, so provable equations are true.

Suppose on the other hand that $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is true in $\odot_\emptyset(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$. Then $t(\mathbf{x}_1, \dots, \mathbf{x}_n) \equiv s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ in $F_{\mathbf{Wheel}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$ (definition 4.40 and lemma 4.44), say $t(\mathbf{x}_1, \dots, \mathbf{x}_n) + \mathbf{z}\mathbf{t}_1 = s(\mathbf{x}_1, \dots, \mathbf{x}_n) + \mathbf{z}\mathbf{t}_2$ is true in $F_{\mathbf{Wheel}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$ for some zero-terms $\mathbf{z}\mathbf{t}_1$ and $\mathbf{z}\mathbf{t}_2$. Then (by completeness) it is provable from the wheel axioms. The rules ‘ $0\mathbf{x}_i = 0$ ’ can finally be used to prove that $\mathbf{z}\mathbf{t}_1 = 0$ and $\mathbf{z}\mathbf{t}_2 = 0$. \square

The acceptance of the rules ‘ $0\mathbf{x}_i = 0$ ’ lets us count as in rings with polynomial expressions in $\mathbf{x}_1, \dots, \mathbf{x}_n$.

Example 4.32. $(\mathbf{x} + 1)^2 = \mathbf{x}(\mathbf{x} + 2) + 1$ is provable using wheel axioms and ‘ $0\mathbf{x} = 0$ ’:

$$\begin{aligned} (\mathbf{x} + 1)^2 &= (\mathbf{x} + 1)(\mathbf{x} + 1) + 0\mathbf{x} + 0 \cdot 1 = (\mathbf{x} + 1)(\mathbf{x} + 1) + 0(\mathbf{x} + 1) \\ &= \mathbf{x}(\mathbf{x} + 1) + 1(\mathbf{x} + 1) = \mathbf{x}(\mathbf{x} + 1) + \mathbf{x} + 1 \\ &= \mathbf{x}(\mathbf{x} + 1 + 1) + 0\mathbf{x} + 1 = \mathbf{x}(\mathbf{x} + 2) + 1 \end{aligned}$$

The theorems above have some corollaries. The first states how identities from semirings can be transferred to wheels of fractions.

Corollary 4.33. *If $t(x_1, \dots, x_n)$ and $s(x_1, \dots, x_n)$ are polynomial terms, then $t(x_1, \dots, x_n) = s(x_1, \dots, x_n)$ is valid in **WF** if and only if it is balanced and valid in \mathbb{N} .*

Proof. Suppose it is valid in **WF**. Then it is valid in $\odot_{\emptyset}\mathbb{N}$ and hence in \mathbb{N} . Since $t(\mathbf{x}_1, \dots, \mathbf{x}_n) \in \mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n] \cong \mathcal{R}\odot_{\emptyset}(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$, we conclude that it cannot be equal to $0/0$. So the equation is balanced.

On the other hand, if the equation is valid in \mathbb{N} , then $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ in

$$\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n] \subset \odot_{\emptyset}(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n]).$$

If it is also balanced, then it is valid in **WF**. \square

Example 4.34. $(x + 1)^2 = x(x + 2) + 1$ is balanced and valid in \mathbb{N} , hence it is valid in wheels of fractions.

Example 4.35. $(x + y)(x - y) = x^2 - y^2$ is balanced and valid in \mathbb{Z} , hence it is valid in wheels of fractions with negation.

The next result states how equations that have partial meaning and partial validity in fields of characteristic zero can be transferred to wheels of fractions.

Corollary 4.36. *Suppose there is an infinite subset X of \mathbb{Q} such that the terms $t(x_1, \dots, x_n)$ and $s(x_1, \dots, x_n)$ are meaningful and \mathbb{Q} -valued when values from X are substituted for the variables and $/$ is interpreted as inversion. Suppose also that t and s yield the same values when such substitutions are made. Then*

$$t(x_1, \dots, x_n) + 0s(x_1, \dots, x_n) = s(x_1, \dots, x_n) + 0t(x_1, \dots, x_n)$$

*is valid in **WF**.*

Proof. Since the latter equation is balanced, we use that its validity in **WF** is equivalent to the truth of

$$t(\mathbf{x}_1, \dots, \mathbf{x}_n) + 0s(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n) + 0t(\mathbf{x}_1, \dots, \mathbf{x}_n)$$

in $\odot_{\emptyset}(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$.

If each side is computed, we get $[p_t(\mathbf{x}_1, \dots, \mathbf{x}_n), q_t(\mathbf{x}_1, \dots, \mathbf{x}_n)]$ on the left for some polynomials p_t, q_t and $[p_s(\mathbf{x}_1, \dots, \mathbf{x}_n), q_s(\mathbf{x}_1, \dots, \mathbf{x}_n)]$ on the right for some polynomials p_s, q_s . Clearly, q_t and q_s will be equal as polynomials. We shall prove that also p_t and p_s must be equal as polynomials.

When $\mathbf{x}_1, \dots, \mathbf{x}_n$ are replaced by values from X , then we get values in $\mathbb{Q} = \mathcal{R}\odot_{S_0}\mathbb{Z}$. Hence q_t and q_s are non-zero for such substitutions. Since we also know that the values for t and s are equal for such substitutions, we conclude that the values for p_t and p_s are equal. But since X is infinite, that implies that p_t and p_s are equal as polynomials. \square

Example 4.37. The equation $x/x = 1$ is valid in $\mathbb{Q} \setminus \{0\}$, hence $x/x + 0 \cdot 1 = 1 + 0x/x$ is valid in wheels of fractions (this is of course a known rule for wheels, but it is instructive to see how it is related to validity in \mathbb{Q}).

Example 4.38. The equation $x/x - 1 = 0$ is valid in $\mathbb{Q} \setminus \{0\}$, hence $x/x - 1 = 0x/x$ is valid in wheels of fractions with negation.

We now prove the lemmas for the theorems.

Lemma 4.39. $0\mathbf{x}_i$ is $+$ -cancellable in $F_{\mathbf{WF}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$, for $i = 1, \dots, n$.

Proof. We shall prove that if $t(x_1, \dots, x_n) + 0x_i = s(x_1, \dots, x_n) + 0x_i$ is valid in $\odot_{\emptyset}\mathbb{N}$, then $t(x_1, \dots, x_n) = s(x_1, \dots, x_n)$ is too.

Substitute therefore x'_j/x''_j for every variable x_j and compute the values of the terms, ending up with the following form, where p_t, q_t, p_s, q_s are polynomials:

$$\begin{aligned} t(x_1, \dots, x_n) &= \frac{p_t(x'_1, \dots, x'_n, x''_1, \dots, x''_n)}{q_t(x'_1, \dots, x'_n, x''_1, \dots, x''_n)} \\ s(x_1, \dots, x_n) &= \frac{p_s(x'_1, \dots, x'_n, x''_1, \dots, x''_n)}{q_s(x'_1, \dots, x'_n, x''_1, \dots, x''_n)}. \end{aligned}$$

The values of $t(x_1, \dots, x_n) + 0x_i$ and $s(x_1, \dots, x_n) + 0x_i$ are then

$$\begin{aligned} t(x_1, \dots, x_n) + 0x_i &= \frac{p_t(x'_1, \dots, x'_n, x''_1, \dots, x''_n)x''_i}{q_t(x'_1, \dots, x'_n, x''_1, \dots, x''_n)x''_i} \\ s(x_1, \dots, x_n) + 0x_i &= \frac{p_s(x'_1, \dots, x'_n, x''_1, \dots, x''_n)x''_i}{q_s(x'_1, \dots, x'_n, x''_1, \dots, x''_n)x''_i}. \end{aligned}$$

Since $t(x_1, \dots, x_n) + 0x_i = s(x_1, \dots, x_n) + 0x_i$ is valid in $\odot_{\emptyset}\mathbb{N}$, we have identical values for *any* substitution in $\odot_{\emptyset}\mathbb{N}$, hence $p_t = p_s$ and $q_t = q_s$ as polynomials. But then $t(x_1, \dots, x_n)$ and $s(x_1, \dots, x_n)$ take the same values for any substitution with elements from $\odot_{\emptyset}\mathbb{N}$, hence $t(x_1, \dots, x_n) = s(x_1, \dots, x_n)$ is valid in $\odot_{\emptyset}\mathbb{N}$. \square

Definition 4.40. Let $t(\mathbf{x}_1, \dots, \mathbf{x}_n) \equiv s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ mean that there are zero-terms $0\mathbf{x}_1^{m_1} \dots \mathbf{x}_n^{m_n}$ and $0\mathbf{x}_1^{m_1} \dots \mathbf{x}_n^{m_n}$ such that

$$t(\mathbf{x}_1, \dots, \mathbf{x}_n) + 0\mathbf{x}_1^{m_1} \dots \mathbf{x}_n^{m_n} = s(\mathbf{x}_1, \dots, \mathbf{x}_n) + 0\mathbf{x}_1^{m_1} \dots \mathbf{x}_n^{m_n}$$

in $F_{\text{Wheel}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$.

Lemma 4.41. \equiv is a congruence relation.

Proof. It is clearly reflexive and symmetric. The transitivity follows from $0x + 0y = 0xy$ (10). We must prove that it preserves $+$, \cdot and $/$. Suppose therefore that $t \equiv s$ (let us suppress the part $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ everywhere). Then there are zero-terms, say $\mathbf{z}\mathbf{t}_1$ and $\mathbf{z}\mathbf{t}_2$ for short, such that $t + \mathbf{z}\mathbf{t}_1 = s + \mathbf{z}\mathbf{t}_2$. We then have (using (6) and (7) in the second and third row)

$$\begin{aligned} t + u + \mathbf{z}\mathbf{t}_1 &= s + u + \mathbf{z}\mathbf{t}_2 \\ tu + \mathbf{z}\mathbf{t}_1 &= (t + \mathbf{z}\mathbf{t}_1)u = (s + \mathbf{z}\mathbf{t}_2)u = su + \mathbf{z}\mathbf{t}_2 \\ /t + \mathbf{z}\mathbf{t}_1 &= /(t + \mathbf{z}\mathbf{t}_1) = /(s + \mathbf{z}\mathbf{t}_2) = /s + \mathbf{z}\mathbf{t}_2. \end{aligned}$$

\square

Lemma 4.42. *If the polynomial terms $p(x_1, \dots, x_n)$ and $q(x_1, \dots, x_n)$ are equivalent, then $p(\mathbf{x}_1, \dots, \mathbf{x}_n) \equiv q(\mathbf{x}_1, \dots, \mathbf{x}_n)$.*

Proof. We first prove that $0p(\mathbf{x}_1, \dots, \mathbf{x}_n) \equiv 0$ by induction on the number of occurrences of operations composing p . It is obvious if p is \mathbf{x}_i or a constant. If p is $t + s$, then $0p(\mathbf{x}_1, \dots, \mathbf{x}_n) = 0(t(\mathbf{x}_1, \dots, \mathbf{x}_n) + s(\mathbf{x}_1, \dots, \mathbf{x}_n)) = 0t(\mathbf{x}_1, \dots, \mathbf{x}_n) + 0s(\mathbf{x}_1, \dots, \mathbf{x}_n) \equiv 0 + 0 = 0$. If p is ts , then $0p(\mathbf{x}_1, \dots, \mathbf{x}_n) = 0t(\mathbf{x}_1, \dots, \mathbf{x}_n)s(\mathbf{x}_1, \dots, \mathbf{x}_n) \equiv 0s(\mathbf{x}_1, \dots, \mathbf{x}_n) \equiv 0$.

Hence, if p is a polynomial term and $z = p(\mathbf{x}_1, \dots, \mathbf{x}_n)$, then for every $x, y \in F_{\mathbf{Wheel}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$ we have $(x + y)z = (x + y)z + 0z \equiv xz + yz$, so that we have

$$\begin{aligned} x + y &= y + x \\ x + (y + z) &= (x + y) + z \\ 0 + x &= x \\ xy &= yx \\ x(yz) &= (xy)z \\ 1x &= x \\ (x + y)z &\equiv xz + yz \\ 0z &\equiv 0 \end{aligned}$$

when counting with polynomial terms. We conclude that every derivation of $p(x_1, \dots, x_n) = q(x_1, \dots, x_n)$ that uses the axioms for commutative semirings only, is transformed into an argument that proves $p(\mathbf{x}_1, \dots, \mathbf{x}_n) \equiv q(\mathbf{x}_1, \dots, \mathbf{x}_n)$, by changing each '=' to '≡'.

Remark: The analogous case of wheels with negation is treated by use of $z - z \equiv 0$, which is true since $z - z = 0z^2$. \square

On the other hand, if $p(\mathbf{x}_1, \dots, \mathbf{x}_n) \equiv q(\mathbf{x}_1, \dots, \mathbf{x}_n)$ for polynomial terms p and q , then they are equivalent. Hence \equiv can be thought of as an extension of the equivalence for polynomial terms.¹⁷

Lemma 4.43. *For any term $t(x_1, \dots, x_n)$, one can find two polynomial terms $p(x_1, \dots, x_n)$ and $q(x_1, \dots, x_n)$ with*

$$t(\mathbf{x}_1, \dots, \mathbf{x}_n) \equiv \frac{p(\mathbf{x}_1, \dots, \mathbf{x}_n)}{q(\mathbf{x}_1, \dots, \mathbf{x}_n)}$$

in $F_{\mathbf{Wheel}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$.

Proof. We use induction on the number of occurrences of operations composing t .

0 occurrences: $t(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is \mathbf{x}_i . Then choose $p(\mathbf{x}_1, \dots, \mathbf{x}_n) = \mathbf{x}_i$ and $q(\mathbf{x}_1, \dots, \mathbf{x}_n) = 1$ and the result is obvious.

$N+1$ occurrences, assuming that we have proved the result for N occurrences (let us suppress the part $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ everywhere): If t is a constant, then choose $p = t$ and $q = 1$. Otherwise, we have the following possibilities.

¹⁷Note however that this statement is informal, since \equiv is not a relation between terms, but between elements of a certain wheel.

1. t is $t_1 + t_2$. By the induction hypotheses, we have $t_1 \equiv p_1/q_1$ and $t_2 \equiv p_2/q_2$. Hence

$$\begin{aligned} t = t_1 + t_2 &\equiv \frac{p_{t_1}}{q_{t_1}} + \frac{p_{t_2}}{q_{t_2}} \equiv \frac{p_{t_1}}{q_{t_1}} + \frac{p_{t_2}}{q_{t_2}} + 0q_{t_1} = \frac{p_{t_1} + \frac{p_{t_2}q_{t_1}}{q_{t_2}}}{q_{t_1}} \\ &\equiv \frac{p_{t_1} + \frac{p_{t_2}q_{t_1}}{q_{t_2}}}{q_{t_1}} + 0q_{t_2} = \frac{p_{t_1} + \frac{p_{t_2}q_{t_1}}{q_{t_2}} + 0q_{t_2}}{q_{t_1}} = \frac{\frac{p_{t_1}q_{t_2} + p_{t_2}q_{t_1}}{q_{t_2}}}{q_{t_1}} \\ &= \frac{p_{t_1}q_{t_2} + p_{t_2}q_{t_1}}{q_{t_1}q_{t_2}}. \end{aligned}$$

2. t is $t_1 t_2$.

$$t = t_1 t_2 \equiv \left(\frac{p_{t_1}}{q_{t_1}} \right) \left(\frac{p_{t_2}}{q_{t_2}} \right) = \frac{p_{t_1} p_{t_2}}{q_{t_1} q_{t_2}}$$

3. t is $/t_1$.

$$t = /t_1 \equiv /(p_{t_1}/q_{t_1}) = q_{t_1}/p_{t_1}$$

□

Lemma 4.44. $(F_{\mathbf{Wheel}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})/\equiv) \cong \odot_{\emptyset}(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$.

Proof. The evaluation morphism $F_{\mathbf{Wheel}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\}) \rightarrow \odot_{\emptyset}(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$ is surjective, we shall prove that its kernel is \equiv .

Clearly, if $t(\mathbf{x}_1, \dots, \mathbf{x}_n) \equiv s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is true in $F_{\mathbf{Wheel}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$, then $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is true in $\odot_{\emptyset}(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$. We need to prove the other direction.

Suppose that $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ in $\odot_{\emptyset}(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$. By the previous lemma, there are polynomial terms p_t, q_t, p_s, q_s such that

$$t(\mathbf{x}_1, \dots, \mathbf{x}_n) \equiv \frac{p_t(\mathbf{x}_1, \dots, \mathbf{x}_n)}{q_t(\mathbf{x}_1, \dots, \mathbf{x}_n)}$$

and

$$s(\mathbf{x}_1, \dots, \mathbf{x}_n) \equiv \frac{p_s(\mathbf{x}_1, \dots, \mathbf{x}_n)}{q_s(\mathbf{x}_1, \dots, \mathbf{x}_n)}.$$

Hence it holds in $\odot_{\emptyset}(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$ that

$$\frac{p_t(\mathbf{x}_1, \dots, \mathbf{x}_n)}{q_t(\mathbf{x}_1, \dots, \mathbf{x}_n)} = t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n) = \frac{p_s(\mathbf{x}_1, \dots, \mathbf{x}_n)}{q_s(\mathbf{x}_1, \dots, \mathbf{x}_n)}.$$

Therefore, p_t and p_s are equivalent, as well as q_t and q_s . Hence, by lemma 4.42, we have

$$p_t(\mathbf{x}_1, \dots, \mathbf{x}_n) \equiv p_s(\mathbf{x}_1, \dots, \mathbf{x}_n)$$

and

$$q_t(\mathbf{x}_1, \dots, \mathbf{x}_n) \equiv q_s(\mathbf{x}_1, \dots, \mathbf{x}_n),$$

so that

$$t(\mathbf{x}_1, \dots, \mathbf{x}_n) \equiv \frac{p_t(\mathbf{x}_1, \dots, \mathbf{x}_n)}{q_t(\mathbf{x}_1, \dots, \mathbf{x}_n)} \equiv \frac{p_s(\mathbf{x}_1, \dots, \mathbf{x}_n)}{q_s(\mathbf{x}_1, \dots, \mathbf{x}_n)} \equiv s(\mathbf{x}_1, \dots, \mathbf{x}_n).$$

□

Lemma 4.45. Suppose p and q are polynomial terms, not both equivalent to 0. If

$$\frac{p(\mathbf{x}_1, \dots, \mathbf{x}_n)}{q(\mathbf{x}_1, \dots, \mathbf{x}_n)} + 0\mathbf{x}_1^{n_1} \dots \mathbf{x}_n^{n_n} = \frac{p(\mathbf{x}_1, \dots, \mathbf{x}_n)}{q(\mathbf{x}_1, \dots, \mathbf{x}_n)} + 0\mathbf{x}_1^{m_1} \dots \mathbf{x}_n^{m_n}$$

is true in $F_{\mathbf{WF}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$, then it is true in $F_{\mathbf{Wheel}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$ too.

Proof. We prove that $(n_1, \dots, n_n) = (m_1, \dots, m_n)$.

Choose distinct positive prime numbers $a_1, \dots, a_n, b_1, \dots, b_n$ such that

$$\frac{p(a_1/b_1, \dots, a_n/b_n)}{q(a_1/b_1, \dots, a_n/b_n)} = \frac{a}{b} \neq \frac{0}{0}$$

in $\odot_\emptyset \mathbb{N}$.

Then

$$\frac{p(a_1/b_1, \dots, a_n/b_n)}{q(a_1/b_1, \dots, a_n/b_n)} + 0(a_1/b_1)^{n_1} \dots (a_n/b_n)^{n_n} = \frac{ab_1^{n_1} \dots b_n^{n_n}}{bb_1^{n_1} \dots b_n^{n_n}}$$

and

$$\frac{p(a_1/b_1, \dots, a_n/b_n)}{q(a_1/b_1, \dots, a_n/b_n)} + 0(a_1/b_1)^{m_1} \dots (a_n/b_n)^{m_n} = \frac{ab_1^{m_1} \dots b_n^{m_n}}{bb_1^{m_1} \dots b_n^{m_n}}$$

but by assumption, these should be equal. Since $a \neq 0$ or $b \neq 0$, we conclude that $(n_1, \dots, n_n) = (m_1, \dots, m_n)$. \square

Lemma 4.46. If $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is true in $F_{\mathbf{WF}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$, then there is a term $0\mathbf{x}_1^{n_1} \dots \mathbf{x}_n^{n_n}$ such that $t(\mathbf{x}_1, \dots, \mathbf{x}_n) + 0\mathbf{x}_1^{n_1} \dots \mathbf{x}_n^{n_n} = s(\mathbf{x}_1, \dots, \mathbf{x}_n) + 0\mathbf{x}_1^{n_1} \dots \mathbf{x}_n^{n_n}$ is true in $F_{\mathbf{Wheel}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$.

Proof. Suppose $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is true in $F_{\mathbf{WF}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$. Then it is in $\odot_\emptyset(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$ too. Hence, there are polynomials $p(x_1, \dots, x_n)$ and $q(x_1, \dots, x_n)$ with

$$\begin{aligned} t(\mathbf{x}_1, \dots, \mathbf{x}_n) &= \frac{p(\mathbf{x}_1, \dots, \mathbf{x}_n)}{q(\mathbf{x}_1, \dots, \mathbf{x}_n)} \\ s(\mathbf{x}_1, \dots, \mathbf{x}_n) &= \frac{p(\mathbf{x}_1, \dots, \mathbf{x}_n)}{q(\mathbf{x}_1, \dots, \mathbf{x}_n)} \end{aligned}$$

true in $\odot_\emptyset(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$.

If both p and q are equivalent to ‘0’, then we may suppose that they are actually both the term ‘0’.

Since $F_{\mathbf{Wheel}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})/\equiv \cong \odot_\emptyset(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$, there are zero-terms, say $\mathbf{zt}_1, \mathbf{zt}_2, \mathbf{zt}_3, \mathbf{zt}_4$ for short, such that

$$\begin{aligned} t(\mathbf{x}_1, \dots, \mathbf{x}_n) + \mathbf{zt}_1 &= \frac{p(\mathbf{x}_1, \dots, \mathbf{x}_n)}{q(\mathbf{x}_1, \dots, \mathbf{x}_n)} + \mathbf{zt}_2 \\ s(\mathbf{x}_1, \dots, \mathbf{x}_n) + \mathbf{zt}_3 &= \frac{p(\mathbf{x}_1, \dots, \mathbf{x}_n)}{q(\mathbf{x}_1, \dots, \mathbf{x}_n)} + \mathbf{zt}_4 \end{aligned}$$

in $F_{\mathbf{Wheel}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$.

Since

$$\begin{aligned} \frac{p(\mathbf{x}_1, \dots, \mathbf{x}_n)}{q(\mathbf{x}_1, \dots, \mathbf{x}_n)} + \mathbf{z}\mathbf{t}_2 + \mathbf{z}\mathbf{t}_3 &= t(\mathbf{x}_1, \dots, \mathbf{x}_n) + \mathbf{z}\mathbf{t}_1 + \mathbf{z}\mathbf{t}_3 \\ &= s(\mathbf{x}_1, \dots, \mathbf{x}_n) + \mathbf{z}\mathbf{t}_1 + \mathbf{z}\mathbf{t}_3 \\ &= \frac{p(\mathbf{x}_1, \dots, \mathbf{x}_n)}{q(\mathbf{x}_1, \dots, \mathbf{x}_n)} + \mathbf{z}\mathbf{t}_4 + \mathbf{z}\mathbf{t}_1 \end{aligned}$$

in $F_{\mathbf{WF}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$ and $\mathbf{z}\mathbf{t}_2 + \mathbf{z}\mathbf{t}_3$ and $\mathbf{z}\mathbf{t}_4 + \mathbf{z}\mathbf{t}_1$ compute to zero-terms of the form used in the statement of the previous lemma, we may use that lemma to conclude that if p and q are not both the term '0', then

$$\frac{p(\mathbf{x}_1, \dots, \mathbf{x}_n)}{q(\mathbf{x}_1, \dots, \mathbf{x}_n)} + \mathbf{z}\mathbf{t}_2 + \mathbf{z}\mathbf{t}_3 = \frac{p(\mathbf{x}_1, \dots, \mathbf{x}_n)}{q(\mathbf{x}_1, \dots, \mathbf{x}_n)} + \mathbf{z}\mathbf{t}_4 + \mathbf{z}\mathbf{t}_1$$

is true also in $F_{\mathbf{Wheel}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$. But if p and q are both '0', then this fact follows from the axiom 8 instead of from the lemma.

Now choose (n_1, \dots, n_n) such that $\mathbf{z}\mathbf{t}_1 + \mathbf{z}\mathbf{t}_3 = 0\mathbf{x}_1^{n_1} \dots \mathbf{x}_n^{n_n}$. Then

$$\begin{aligned} t(\mathbf{x}_1, \dots, \mathbf{x}_n) + 0\mathbf{x}_1^{n_1} \dots \mathbf{x}_n^{n_n} &= \frac{p(\mathbf{x}_1, \dots, \mathbf{x}_n)}{q(\mathbf{x}_1, \dots, \mathbf{x}_n)} + \mathbf{z}\mathbf{t}_2 + \mathbf{z}\mathbf{t}_3 \\ &= \frac{p(\mathbf{x}_1, \dots, \mathbf{x}_n)}{q(\mathbf{x}_1, \dots, \mathbf{x}_n)} + \mathbf{z}\mathbf{t}_4 + \mathbf{z}\mathbf{t}_1 \\ &= s(\mathbf{x}_1, \dots, \mathbf{x}_n) + 0\mathbf{x}_1^{n_1} \dots \mathbf{x}_n^{n_n} \end{aligned}$$

□

Lemma 4.47. *Any equation $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ which is true in $F_{\mathbf{WF}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$ can be proved using the wheel axioms together with the axioms '0 \mathbf{x}_i is +-cancellable' ($i = 1, \dots, n$).*

Proof. Suppose $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is true in $F_{\mathbf{WF}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$. By the previous lemma, we get an equation which is true in $F_{\mathbf{Wheel}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$ by adding a zero-term $0\mathbf{x}_1^{n_1} \dots \mathbf{x}_n^{n_n}$ to each side. The resulting equation can be proved using the axioms for wheels. The term $0\mathbf{x}_1^{n_1} \dots \mathbf{x}_n^{n_n}$ can then be cancelled by the use of '0 \mathbf{x}_i is +-cancellable' several times for different i , since it is provable that

$$0\mathbf{x}_1^{n_1} \dots \mathbf{x}_n^{n_n} = \sum_{j=1}^n \sum_{k=1}^{n_j} 0\mathbf{x}_j.$$

□

Lemma 4.48. *Suppose that $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = 0/0$ is true in $\odot_{\emptyset}(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$. Then it is true in $F_{\mathbf{WF}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$ too.*

Proof. Assume the conditions. $t(\mathbf{x}_1, \dots, \mathbf{x}_n) \equiv 0/0$ in $F_{\mathbf{Wheel}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$ according to lemma 4.44, say

$$t(\mathbf{x}_1, \dots, \mathbf{x}_n) + \mathbf{z}\mathbf{t}_1 = 0/0 + \mathbf{z}\mathbf{t}_2 = 0/0 = 0/0 + \mathbf{z}\mathbf{t}_1.$$

Hence $t(\mathbf{x}_1, \dots, \mathbf{x}_n) + \mathbf{z}\mathbf{t}_1 = 0/0 + \mathbf{z}\mathbf{t}_1$ also in $F_{\mathbf{WF}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$, but by lemma 4.39, we conclude that $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = 0/0$. □

Lemma 4.49. *Suppose that $t(x_1, \dots, x_n) = s(x_1, \dots, x_n)$ is an unbalanced equation. Then $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is true in $F_{\mathbf{WF}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$ if and only if each side is equal to $0/0$ in $\odot_\emptyset(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$ (and hence, by the previous lemma, in $F_{\mathbf{WF}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$ too).*

Proof. Suppose that $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is true in $F_{\mathbf{WF}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$. By lemma 4.46, there is a zero-term $0\mathbf{x}_1^{n_1} \dots \mathbf{x}_n^{n_n}$ such that

$$t(\mathbf{x}_1, \dots, \mathbf{x}_n) + 0\mathbf{x}_1^{n_1} \dots \mathbf{x}_n^{n_n} = s(\mathbf{x}_1, \dots, \mathbf{x}_n) + 0\mathbf{x}_1^{n_1} \dots \mathbf{x}_n^{n_n}$$

in $F_{\mathbf{Wheel}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$. By completeness, this equation is provable using the wheel axioms, but since it is unbalanced, one must use an unbalanced axiom. There is just one, namely ‘ $x + 0/0 = 0/0$ ’. That can be used only if ‘ $0/0$ ’ occurs as a subterm in the derivation; but if it does, then every term in the derivation can be proved to be equal to $0/0$.

Hence

$$t(\mathbf{x}_1, \dots, \mathbf{x}_n) + 0\mathbf{x}_1^{n_1} \dots \mathbf{x}_n^{n_n} = 0/0$$

and

$$s(\mathbf{x}_1, \dots, \mathbf{x}_n) + 0\mathbf{x}_1^{n_1} \dots \mathbf{x}_n^{n_n} = 0/0$$

in $F_{\mathbf{WF}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$ and hence also in $\odot_\emptyset(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$. But in the latter, we have $0\mathbf{x}_1^{n_1} \dots \mathbf{x}_n^{n_n} = 0$, hence we conclude that

$$t(\mathbf{x}_1, \dots, \mathbf{x}_n) = t(\mathbf{x}_1, \dots, \mathbf{x}_n) + 0\mathbf{x}_1^{n_1} \dots \mathbf{x}_n^{n_n} = 0/0$$

and

$$s(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n) + 0\mathbf{x}_1^{n_1} \dots \mathbf{x}_n^{n_n} = 0/0$$

holds in $\odot_\emptyset(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$.

The other direction was proved in the previous lemma. \square

Lemma 4.50. *Suppose that $t(x_1, \dots, x_n) = s(x_1, \dots, x_n)$ is a balanced equation and that $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is true in $\odot_\emptyset(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$. Then it is true also in $F_{\mathbf{WF}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$.*

Proof. Assume the conditions. Then $t(\mathbf{x}_1, \dots, \mathbf{x}_n) \equiv s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is true in $F_{\mathbf{Wheel}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$ by lemma 4.44, say that $t(\mathbf{x}_1, \dots, \mathbf{x}_n) + 0\mathbf{x}_1^{n_1} \dots \mathbf{x}_n^{n_n} = s(\mathbf{x}_1, \dots, \mathbf{x}_n) + 0\mathbf{x}_1^{m_1} \dots \mathbf{x}_n^{m_n}$. Since this is true in $F_{\mathbf{Wheel}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$, it is in $F_{\mathbf{WF}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$ too. If $(n_1, \dots, n_n) = (m_1, \dots, m_n)$, then it follows by lemma 4.39 that $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is true in $F_{\mathbf{WF}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$.

If $(n_1, \dots, n_n) \neq (m_1, \dots, m_n)$, then the equation $t(x_1, \dots, x_n) + 0x_1^{n_1} \dots x_n^{n_n} = s(x_1, \dots, x_n) + 0x_1^{m_1} \dots x_n^{m_n}$ is unbalanced, hence, by the previous lemma, we have that $t(\mathbf{x}_1, \dots, \mathbf{x}_n) + 0\mathbf{x}_1^{n_1} \dots \mathbf{x}_n^{n_n} = 0/0$ and $s(\mathbf{x}_1, \dots, \mathbf{x}_n) + 0\mathbf{x}_1^{m_1} \dots \mathbf{x}_n^{m_n} = 0/0$ in $\odot_\emptyset(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$. But since $0\mathbf{x}_1^{n_1} \dots \mathbf{x}_n^{n_n} = 0$ in $\odot_\emptyset(\mathbb{N}[\mathbf{x}_1, \dots, \mathbf{x}_n])$, we have

$$t(\mathbf{x}_1, \dots, \mathbf{x}_n) = 0/0$$

and

$$s(\mathbf{x}_1, \dots, \mathbf{x}_n) = 0/0$$

which by lemma 4.48 implies that $t(\mathbf{x}_1, \dots, \mathbf{x}_n) = s(\mathbf{x}_1, \dots, \mathbf{x}_n)$ is true in $F_{\mathbf{WF}}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\})$. \square

5 Wheel-modules

Since the notion of module is central in the study of rings, and wheels can be used as a replacement of rings, we should introduce the notion of an H -module over a wheel H . We indicate briefly how a theory of modules may be developed.

The notion of A -module, where A is an arbitrary semiring, was defined in section 3 (definition 3.2) and it was noted that when A happens to be a ring, then an A -module in the defined sense is an A -module in the usual sense of rings. The following definition is similar to that definition. Note that the axioms are the wheel axioms that are interpretable in the given context, we mark to the right to which wheel axiom each axiom corresponds.

Definition 5.1. *Let H be a wheel. An H -module is a commutative monoid $\langle M, 0, + \rangle$ with multiplication by H -elements defined (formally, a function $H \times M \rightarrow M$ written $(x, m) \mapsto xm$) such that for any $x, x' \in H, m, m' \in M$,*

$$(xx')m = x(x'm) \quad (\text{as 2})$$

$$1m = m \quad (\text{as 2})$$

$$(x + x')m + 0m = xm + x'm \quad (\text{as 3})$$

$$x(m + m') + x0 = xm + xm' \quad (\text{as 3})$$

$$(/x)m + m' + x0 = (/x)(m + xm') \quad (\text{as 4})$$

$$0 \cdot 0 = 0 \quad (\text{as 5})$$

$$x(m + 0m') = xm + 0m' \quad (\text{as 6})$$

$$m + (/0)0 = (/0)0 \quad (\text{as 8})$$

where 0 to the left is in H , while 0 to the right or alone is in M .

A morphism of H -modules is as usual a monoid-morphism which also preserves the multiplication. Let $\mathcal{R}_H M$ be the subset $\{m \in M \mid 0m = 0\}$, which is clearly an $\mathcal{R}H$ -module and for any H -module-morphism $\varphi : M \rightarrow N$, let $\mathcal{R}_H \varphi$ be the $\mathcal{R}H$ -module-morphism obtained by restriction to $\mathcal{R}_H M$. \mathcal{R}_H is clearly a functor from the category of H -modules to the category of $\mathcal{R}H$ -modules.

The construction of wheels of fractions works also for modules over commutative semirings, so that we may construct “ $\odot_X A$ -modules” from A -modules.

Let A be a commutative semiring, X a subset and S the multiplicative monoid generated by X . Let M be an A -module and define \sim on $M \times A$ by

$$(m, a) \sim (m', a') \iff \exists s_1, s_2 \in S : (s_1 m, s_1 a) = (s_2 m', s_2 a').$$

That this is an equivalence relation is seen in the same way as before. Let $\odot_X M$ be $M \times A / \sim$ with the operations $([c, d] \in \odot_X A)$

$$0 = [0, 1]$$

$$[m, a] + [m', a'] = [a'm + am', aa']$$

$$[c, d][m, a'] = [cm, da'].$$

These operations are well-defined by the same arguments as before, and the resulting structure is clearly a $\odot_X A$ -module, denoted $\odot_X M$.

If $\varphi : M \rightarrow N$ is an A -module homomorphism, let $\odot_X \varphi$ be the $\odot_X A$ -module homomorphism $\odot_X M \rightarrow \odot_X N$ given by $[m, a] \mapsto [\varphi(m), a]$. \odot_X is then

a functor from the category of A -modules to the category of $\odot_X A$ -modules; with right adjoint $\mathcal{R}_{\odot_X A}$. If A is a ring, then the composition $\mathcal{R}_{\odot_X A} \odot_X$ is the well-known functor S^{-1} .

As an application, we may define $\odot_X A$ in the case when A is a commutative ring *without* identity element. Suppose A is such a ring, X a subset of it. Let $A[1]$ be (following Nagata) the additive group $\mathbb{Z} \times A$ together with the extra operations $1 = (1, 0)$ and

$$(n, a)(n', a') = (nn', na' + n'a + aa'),$$

which makes $A[1]$ a commutative ring *with* identity 1, and with A injectively embedded by the operation-preserving injection $x \mapsto (0, x)$. Identify A with its image under this embedding. A is then an ideal of $A[1]$, in particular an $A[1]$ -module. In this way, $\odot_X A$ is defined as a $\odot_X A[1]$ -module.

Another application is exponentiation with fractions as exponents. Since every commutative monoid $\langle M, e, \cdot \rangle$ is an \mathbb{N} -module by multiplication m^n ($m \in M, n \in \mathbb{N}$), it is extended to a $\odot_{S_0} \mathbb{N}$ -module $\odot_{S_0} M$ consisting of elements $[m, n]$ ($m \in M, n \in \mathbb{N}$, write $\sqrt[n]{m}$ for $[m, n]$). One has $\sqrt[n]{m}^{p/q} = \sqrt[q]{m^p}$. In particular, $x^{1/0}$ and $x^{0/0}$ is defined in this way.

References

- [AM69] M. F. Atiyah and I. G. MacDONald. *Introduction to Commutative Algebra*. Addison-Wesley, London, 1969.
- [Bir35] Birkhoff, 1935. See e.g. [BS81] theorem 11.9, p. 75 or [Grä79] theorem 3, p. 171.
- [Bor94] F. Borceux. *Handbook of Categorical Algebra*, volume 2. Cambridge University Press, Cambridge, 1994.
- [Bou73] N. Bourbaki. *Elements of Mathematics, Algebra, Part I*. Addison-Wesley, Reading, Massachusetts, 1973.
- [BS81] S. Burris and H. P. Sankappanavar. *A Course in Universal Algebra*. Springer-Verlag, New York, 1981.
- [EP00] A. Edalat and P. J. Potts. A new representation for exact real numbers. In S. Brookes and M. Mislove, editors, *Electronic Notes in Theoretical Computer Science*, volume 6. Elsevier Science Publishers, 2000.
- [Gol99] J. S. Golan. *Semirings and their Applications*. Kluwer Academic Publishers, Dordrecht, 1999.
- [Grä79] G. Grätzer. *Universal Algebra*. Springer-Verlag, New York, second edition, 1979.
- [Har77] R. Hartshorne. *Algebraic Geometry*. Springer-Verlag, New York, 1977.
- [Hey56] A. Heyting. *Intuitionism: An Introduction*. North-Holland, Amsterdam, 1956.
- [ML84] P. Martin-Löf. *Intuitionistic Type Theory*. Bibliopolis, Naples, 1984.
- [MRR88] R. Mines, F. Richman, and W. Ruitenburg. *A Course in Constructive Algebra*. Springer-Verlag, New York, 1988.
- [Öbe00] M. Öberg. Algebraiska och topologiska hjul. Examensarbete i matematik, Stockholm University, 2000.
- [Pot98] P. J. Potts. *Exact Real Arithmetic using Möbius Transformations*. PhD thesis, University of London, 1998. Currently available at <http://www.purplefinder.com/~potts/>.
- [Set97] A. Setzer. *Wheels (draft)*. Currently available at <http://www-compsci.swan.ac.uk/~csetzer/>, 1997.