Equational Reasoning with 2-dimensional Diagrams

Yves Lafont

Laboratoire de Mathématiques Discrètes, UPR 9016 du CNRS
163 avenue de Luminy - Case 930 - 13288 Marseille Cedex 9, France
Email: lafont@lmd.univ-mrs.fr

Abstract

The significance of the 2-dimensional calculus, which goes back to Penrose, has already been pointed out by Joyal and Street. Independently, Burroni has introduced a general notion of n-dimensional presentation and he has shown that the equational logic of terms is a special case of 2-dimensional calculus.

Here, we propose a combinatorial definition of 2-dimensional diagrams and a simple method for proving that certain monoidal categories are finitely 2-presentable. In particular, we consider Burroni's presentation of finite maps and we extend it to the case of finite relations.

This paper should serve as a reference for our future work on symbolic computation, including a theory of 2-dimensional rewriting and the design of software for interactive diagrammatic reasoning.

1 From terms to diagrams

Terms are one of the basic ingredients of logic. Consider for instance the theory of commutative rings, expressed by means of function symbols $+$, $\times$ (both of arity 2), $-$ (of arity 1), 0, 1 (both of arity 0) and variables. This theory is called *algebraic* because it consists of equations between terms. It is generated by the following axioms:

$$(x+y)+z = x+(y+z), \qquad 0+x = x, \qquad x+y = y+x, \qquad (-x)+x = 0,$$

$$(x\times y)\times z = x\times(y\times z), \qquad 1\times x = x, \qquad x\times y = y\times x,$$

$$x\times(y+z) = (x\times y)+(x\times z), \qquad x\times 0 = 0.$$

From those axioms, we derive theorems such as

$$-(-x) = x, \qquad (x+y)\times(x+(-y)) = (x\times x)+(-(y\times y)),$$

using the rules of *equational logic*, which are independent of the theory, and can be expressed as follows:

$$\frac{}{u=u} \qquad \frac{u=v \qquad v=w}{u=w} \qquad \frac{u=v}{w[u/x]=w[v/x]} \qquad \frac{u=v}{u[w/x]=v[w/x]}$$

In these inference rules, u, v, w range over terms and $u[v/x]$ is the term u where all occurrences of the variable x have been replaced by v.

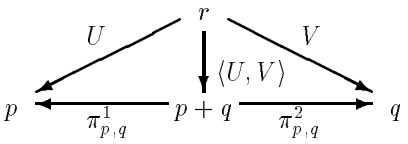
1.1 Lawvere's approach

Fix a set Σ of symbols with arities, and consider the following category $\mathbf{T}(\Sigma)$:

- The objects are natural numbers.
- A morphism $U : p \to q$ is given by a sequence $u_1, \ldots, u_q$ of terms built with symbols in Σ and variables in $\{x_1, \ldots, x_p\}$.
- The composition is defined by substitution. More precisely, if $U : p \to q$ and $V : q \to r$ are morphisms, then $V \circ U : p \to r$ is given by the sequence $w_1, \ldots, w_r$ where w_i is $v_i[u_1/x_1, \ldots, u_q/x_q]$. The identity on p is given by the sequence $x_1, \ldots, x_p$.

This $\mathbf{T}(\Sigma)$ is a *cartesian category* (or *category with finite products*, in the terminology of [Mac71]):

- The *cartesian product* of p by q is $p + q$. The *cartesian unit* (or *terminal object*) is 0.
- The *projections* $\pi^1_{p,q} : p + q \to p$ and $\pi^2_{p,q} : p + q \to q$ are given by the sequences $x_1, \ldots, x_p$ and $x_{p+1}, \ldots, x_{p+q}$.
- For any $U : r \to p$ and $V : r \to q$, there is a unique morphism $W = \langle U, V \rangle : r \to p + q$ such that $\pi^1_{p,q} \circ W = U$ and $\pi^2_{p,q} \circ W = V$:



  It is given by the sequence $u_1, \ldots, u_p, v_1, \ldots, v_q$. Similarly, for any object p, there is a unique morphism from p to 0, which is given by the empty sequence.

In fact, $\mathbf{T}(\Sigma)$ is the *free cartesian category* generated by Σ. Furthermore, if $\mathcal{E}$ is a set of equations on Σ, we can introduce the category $\mathbf{T}(\Sigma; \mathcal{E})$, whose objects are natural numbers and whose morphisms are sequences of terms modulo $\mathcal{E}$. This $\mathbf{T}(\Sigma; \mathcal{E})$ is also a cartesian category, and any model of $\mathcal{E}$ can be seen as a functor Φ from $\mathbf{T}(\Sigma; \mathcal{E})$ to the category of sets which preserves the cartesian structure. This functor is defined as follows:

- $\Phi(p) = M^p$ where M is the underlying set of the model.
- If $U : p \to q$ is given by the sequence $u_1, \ldots, u_q$, then $\Phi(U)$ is the map $f : M^p \to M^q$ which is defined by the formula $f(x_1, \ldots, x_p) = (u_1, \ldots, u_q)$ (interpreted in the model).

This is the basis of the *Categorical Logic* (see [Law63, Poi85]) which replaces the *syntactical* notions of terms, substitution, equational logic, ... by the *algebraic* notion of cartesian category.

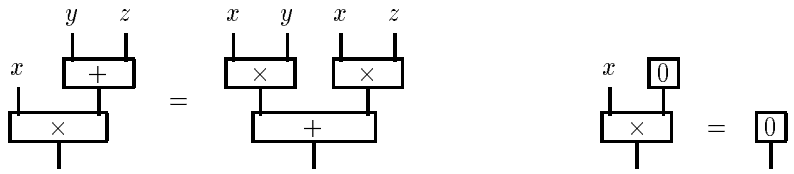
1.2 Burroni's approach

The structure of cartesian category itself is rather sophisticated. Unlike composition, an operation such as $U, V \mapsto \langle U, V \rangle$ has no simple geometric interpretation. However, at the cost of adding a finite number of symbols and equations, cartesian categories can be replaced by strict monoidal categories, where all operations have an obvious geometric interpretation in dimension 2. This suggests an equational logic of planar diagrams, which is more low-level and also more general than the equational logic of terms. In fact, this kind of logic has already been used for a long time, for instance in theoretical physics (see [PeR86]).

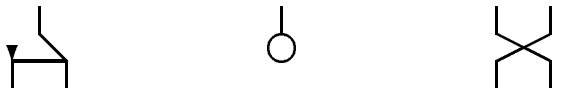
Before giving a formal definition of these diagrams, we can already explain the representation of terms. Each function symbol of arity p is pictured as a cell with p inputs and one output:



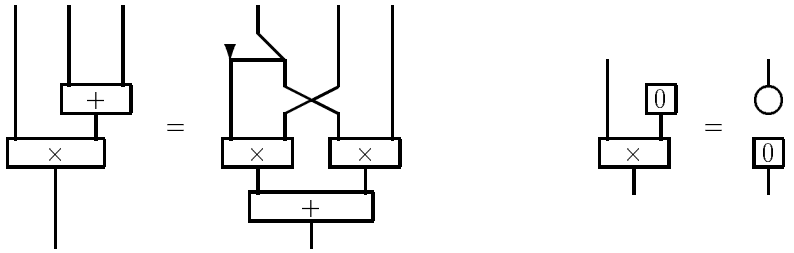
In a first step, equations such as $x \times (y + z) = (x \times y) + (x \times z)$ and $x \times 0 = 0$ can be pictured as follows:



Next, variables are eliminated by means of three extra cells for sharing, discarding and swapping[1]:



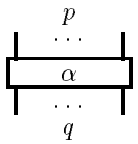
Now, our equations are pictured as follows:



No variable is needed because in each equation, the inputs of the left member correspond to the inputs of the right one, in the same order. Of course, this representation of terms is not unique and some extra equations will be needed, which do not appear at the level of terms. The precise correspondence between terms and diagrams has been established by Burroni (see [Bur93]). We shall return to this in section 5.

2 2-dimensional syntax

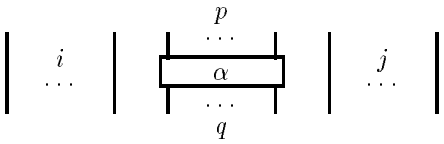
We start with a finite set Σ of *symbols*, where a pair p, q of natural numbers is given for every $\alpha \in \Sigma$. We write $\alpha : p \to q$ and α is pictured as a cell with p inputs and q outputs:



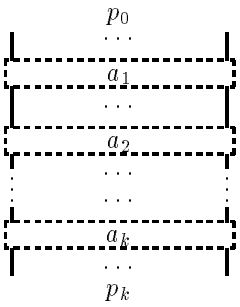
[1] Those cells are related to the so-called *structural rules* of Gentzen's sequent calculus: contraction, weakening and exchange.

Here is a possible way of defining our planar diagrams:

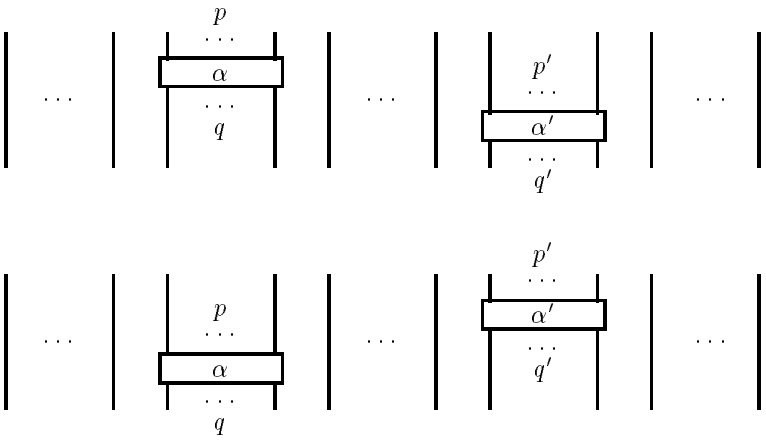
– A *layer* a is given by a symbol $\alpha : p \to q$ and a pair i, j of natural numbers (the *context*). We write $a = i + \alpha + j : i + p + j \to i + q + j$.



– A *diagram* u is given by a sequence of natural numbers $p_0, \ldots, p_k$ and a sequence of layers $a_1 : p_0 \to p_1$, $a_2 : p_1 \to p_2$, $\ldots$, $a_k : p_{k-1} \to p_k$. We write $u = a_k \circ \ldots \circ a_2 \circ a_1 : p_0 \to p_k$.



Now, assume that a diagram contains two consecutive layers $i + \alpha + j : i + p + j \to i + q + j$ and $i' + \alpha' + j' : i' + p' + j' \to i' + q' + j'$ such that $i + q \leq i'$, or equivalently $j \geq p' + j'$, and consider the diagram obtained by interchanging α with α':



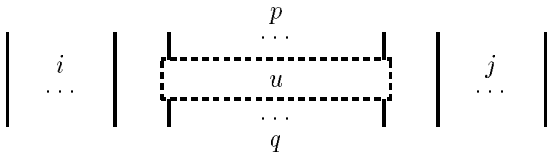
Note that, in this process, the context of α becomes i, $j - p' + q'$ and the one of α' becomes $i' - q + p$, j'. We call *syntactical isotopy* the equivalence relation generated by this transformation. Geometrically, two diagrams $u, v : p \to q$ are isotopic if it is possible to go from u to v by a continuous deformation. It is easy to see that equivalence classes are finite, so that isotopy is decidable. From now on, we shall always work modulo isotopy.

For any $u : p \to q$ and $v : q \to r$, the diagram $v \circ u : p \to r$ is defined by concatenation, and for any p, the diagram $\mathrm{id}_p : p \to p$ is defined by an empty sequence of layers:

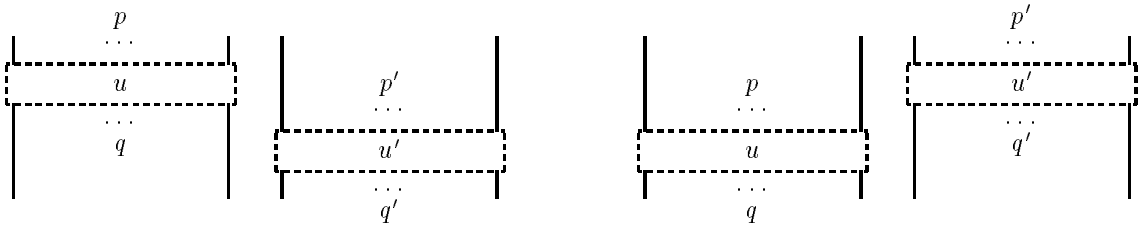


Clearly, the composition is associative with unit id_p.

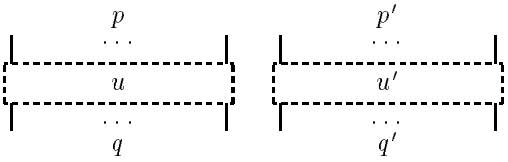
If $u : p \to q$ is a diagram and i, j are natural numbers, the diagram $i + u + j : i + p + j \to i + q + j$ is defined by extending the context:



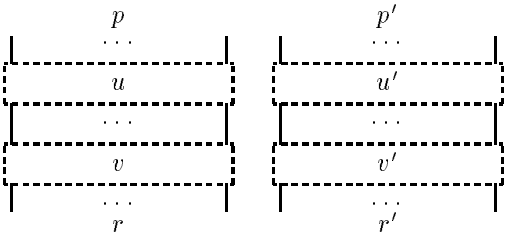
In particular we write $i + u$ for $i + u + 0$ and $u + j$ for $0 + u + j$. Now, if $u : p \to q$ and $u' : p' \to q'$ are diagrams, the sum $u + u' : p + p' \to q + q'$ is by definition $(q + u') \circ (u + p')$ which is isotopic to $(u + q') \circ (p + u')$:



In practice, $u + u'$ will be pictured as follows:

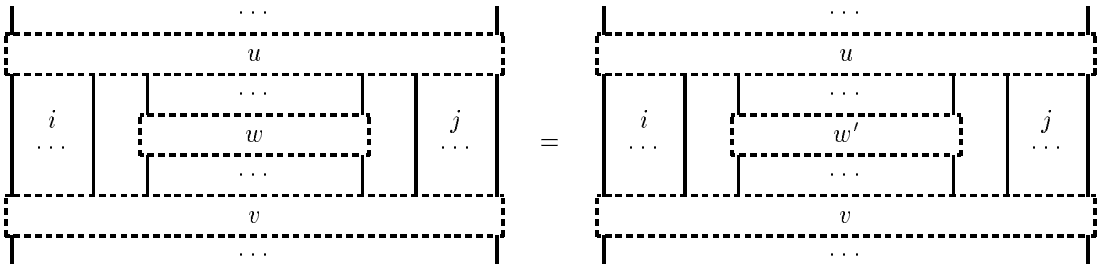


Clearly, this sum is associative with unit id_0 (the *empty diagram*) but not commutative. Furthermore, it is functorial since $(v \circ u) + (v' \circ u')$ is isotopic to $(v + v') \circ (u + u')$. Indeed, both diagrams can be pictured as follows:



Therefore, we have defined a *strict monoidal category* $\mathbf{D}(\Sigma)$ whose objects are natural numbers and whose morphisms are (isotopy classes of) diagrams. In fact, it is not hard to see that $\mathbf{D}(\Sigma)$ is freely generated by Σ.

Now, let $\mathcal{E}$ be a set of pairs of diagrams $w, w' : p \to q$. Each pair will be called an *equation*. If a diagram is of the form $v \circ (i + w + j) \circ u$ where w is the left member of an equation, we can replace w by the right member w' of this equation:



Again, it is easy to see that the equivalence generated by this transformation (and isotopy) is a congruence: it is compatible with composition and sum. The quotient of $\mathbf{D}(\Sigma)$ by this congruence is a strict monoidal category $\mathbf{D}(\Sigma; \mathcal{E})$. In the subsequent sections, we shall see examples of strict monoidal categories which can be presented in this way, with finite Σ and $\mathcal{E}$. Of course, $\mathbf{D}(\Sigma; \mathcal{E})$ satisfies a universal property: if C is a strict monoidal category with an object A and a morphism $\varphi_\alpha : A^p \to A^q$ for each $\alpha : p \to q$ in Σ, such that all equations in $\mathcal{E}$ are satisfied, then there is a unique strict monoidal functor $\Phi : \mathbf{D}(\Sigma; \mathcal{E}) \to \mathcal{C}$ such that $\Phi(\alpha) = \varphi_\alpha$ for all $\alpha \in \Sigma$.

The most classical monoidal categories, for instance sets with cartesian product, or vector spaces with tensor product, are not strict. However, this is not a real problem since any monoidal category C is in fact equivalent to a strict one, whose objects are sequences of objects in C. Therefore we can forget about this bureaucratic distinction between strict and non-strict monoidal categories.

3 Presentation of the finite maps

Here we give the first example of a finitely presented monoidal category, which will play an essential role in the interpretation of algebraic theories. We consider the category $\mathbf{F}$ of finite maps:

- The objects are natural numbers.
- A morphism $\varphi : p \to q$ is a map from $\{1, \ldots, p\}$ to $\{1, \ldots, q\}$.

If $\varphi : p \to q$ and $\varphi' : p' \to q'$ are finite maps, then $\psi = \varphi + \varphi' : p + p' \to q + q'$ is defined by

$$\begin{array}{lll} \psi(i) & = \varphi(i) & \text{for } i = 1, \ldots, p, \\ \psi(p + i) & = q + \varphi'(i) & \text{for } i = 1, \ldots, p'. \end{array}$$

This non-commutative addition defines a (strict) monoidal structure on $\mathbf{F}$. We shall write $\varphi + q$ for $\varphi + \mathrm{id}_q$ and $p + \psi$ for $\mathrm{id}_p + \psi$.

3.1 monotone maps

Any monotone map $\omega : m \to n$ is a product of maps of the form $i + \mu + j$ and $i + \eta + j$, where μ is the unique map from 2 to 1 and η is the unique map from 0 to 1. In other words, μ and η generate the monoidal category of monotone maps.

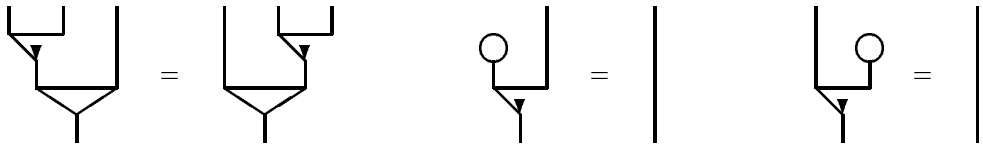
These generators satisfy the following equations:

$$\mu \circ (\mu + 1) = \mu \circ (1 + \mu), \qquad \mu \circ (\eta + 1) = \mathrm{id}_1, \qquad \mu \circ (1 + \eta) = \mathrm{id}_1.$$

Let us introduce the following cells for μ and η:



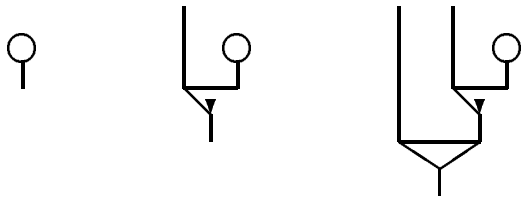
The equations are pictured as follows:



(A_1) (A_2) (A_3)

If two diagrams in $\mathbf{D}(\mu, \eta)$ are equivalent modulo (A_1), (A_2) and (A_3), they represent the same map. We shall now prove the converse.

Let α_n be the unique map from n to 1. It can be represented by a diagram a_n with $n + 1$ layers. For instance, here are a_0, a_1 and a_2:



Lemma 1. *(decomposition of a monotone map) Any monotone map $\omega : m \to n$ can be uniquely written as $\alpha_{p_1} + \ldots + \alpha_{p_n}$.*

This is obvious (p_i is the number of antecedents of i by ω), and this defines a canonical diagram $\widehat{\omega}$ for every monotone map ω. In particular, this shows that μ and η generate the monotone maps. Here is a syntactical analogue of lemma 1:

Lemma 2. *Any diagram $u : m \to n$ in $\mathbf{D}(\mu, \eta)$ is equivalent to a diagram of the form $a_{p_1} + \ldots + a_{p_n}$ modulo (A_1), (A_2) and (A_3).*

Proof. By induction on the length of u. If u is id_n, then u is equivalent to $a_1 + \ldots + a_1$ by n applications of (A_3). Otherwise u is of the form $(i + \mu + j) \circ v$ or $(i + \eta + j) \circ v$. In the first case, we use the fact that $\mu \circ (a_p + a_q)$ is equivalent to a_{p+q} (generalized associativity) by p applications of (A_1) and one application of (A_2). The second case is obvious. □

Using this lemma and the uniqueness of the decomposition in lemma 1, we see that any diagram representing a monotone map ω is equivalent to $\widehat{\omega}$. Therefore, if two diagrams represent the same map, they are equivalent. In other words, the monoidal category of monotone maps is finitely presented by μ, η and the equations (A_1), (A_2), (A_3). The category of monotone maps is also called *simplicial category* because of its geometric interpretation (see [Mac71]).

3.2 permutations

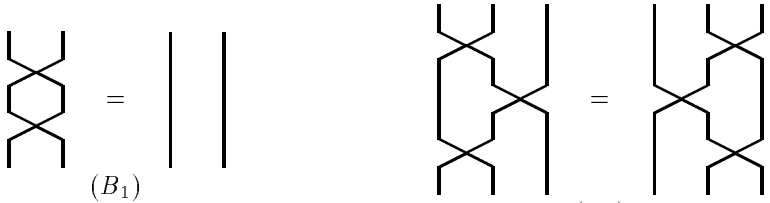
It is well known that any permutation $\pi : n \to n$ is a product of transpositions $i + \tau + j$, where $\tau : 2 \to 2$ is defined by $\tau(1) = 2$ and $\tau(2) = 1$. In other words, τ generates the monoidal category of finite permutations.
This generator satisfies the following equation:

$$\tau \circ \tau = \mathrm{id}_2, \qquad (\tau + 1) \circ (1 + \tau) \circ (\tau + 1) = (1 + \tau) \circ (\tau + 1) \circ (1 + \tau).$$

Let us introduce the following cell for τ:



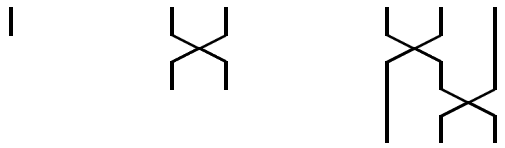
The equations can be pictured as follows:



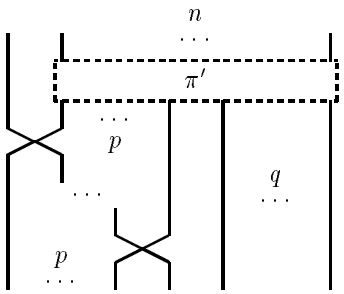
(B_1) (B_2)

We shall now prove that we have indeed a 2-presentation of the permutations.

Let $\beta_n : n + 1 \to n + 1$ be the circular permutation mapping 1 to $n + 1$. It can be represented by a diagram b_n with n layers. For instance, here are b_0, b_1 and b_2:



Lemma 3. *(decomposition of a permutation) Any permutation $\pi : n + 1 \to n + 1$ can be uniquely written as $(\beta_p + q) \circ (1 + \pi')$ with $\pi' : n \to n$.*



This is obvious (take $p = \pi(1) - 1$ and $q = n - p$). Applying this decomposition inductively, we get a canonical diagram $\widehat{\pi}$ for every permutation $\pi : n \to n$ (if $n = 0$, take $\widehat{\pi} = \mathrm{id}_0$). In particular, this shows that τ generates the finite permutations.

Lemma 4. *Any diagram $u : n+1 \to n+1$ in $\mathbf{D}(\tau)$ is equivalent to a diagram of the form $(b_p + q) \circ (1 + u')$ modulo (B_1) and (B_2).*

Proof. By induction on the length of u. If u is id_{n+1}, it is of the form $(b_0 + n) \circ (1 + \mathrm{id}_n)$. Otherwise u is of the form $(i + \tau + j) \circ v$ and by induction hypothesis, v is equivalent to $(b_p + q) \circ (1 + v')$. Now there are four cases (figure 1):

- If $i < p - 1$, then u is equivalent to $(b_p + q) \circ (1 + u')$ where u' is $(i + \tau + (j - 1)) \circ v'$.
- If $i = p - 1$, then u is equivalent to $(b_{p-1} + (q + 1)) \circ (1 + v')$.
- If $i = p$, then u is $(b_{p+1} + (q - 1)) \circ (1 + v')$.
- If $i > p$, then u is $(b_p + q) \circ (1 + u')$ where u' is $((i - 1) + \tau + j) \circ v'$.

We have used (B_2) in the first case and (B_1) in the second case. □

As in the case of monotone maps, we can therefore conclude that the monoidal category of finite permutations is finitely presented by τ and the equations $(B_1), (B_2)$. This 2-dimensional presentation can be compared with the traditional 1-dimensional presentation of the symmetric group S_n by means of $n - 1$ generators $\tau_0, \ldots, \tau_{n-2}$ and $n(n-1)/2$ equations:

$$\begin{aligned} \tau_i^2 &= 1 && \text{for } i = 0, \ldots, n-2, \\ \tau_i \tau_{i+1} \tau_i &= \tau_{i+1} \tau_i \tau_{i+1} && \text{for } i = 0, \ldots, n-3, \\ \tau_i \tau_j &= \tau_j \tau_i && \text{for } i = 0, \ldots, n-4 \text{ and } j = i+2, \ldots, n-2. \end{aligned}$$

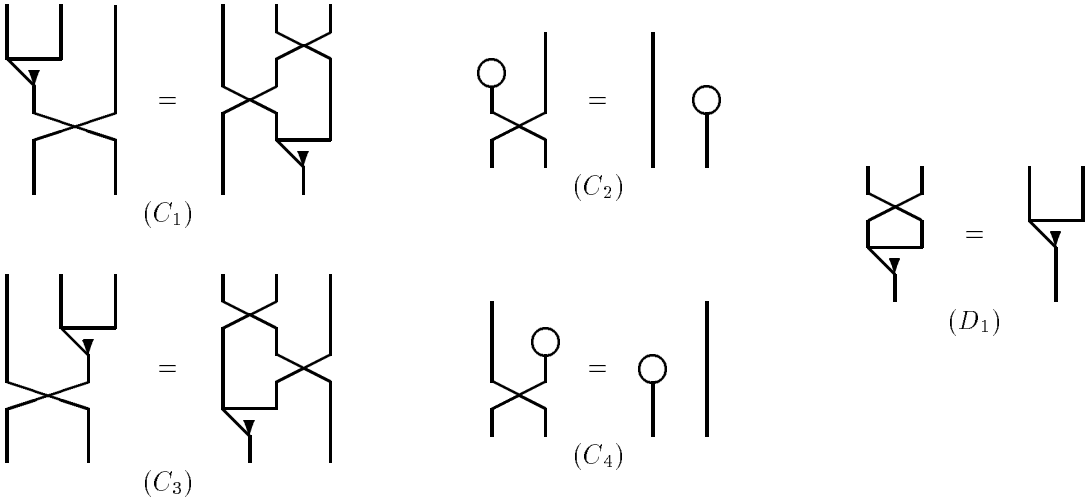
Here, the generator τ_i represents the map $i + \tau + (n - i - 2)$. Therefore, the two first families of equations correspond to (B_1) and (B_2). The third one corresponds to isotopy. Of course, the 2-dimensional presentation is much more economical and deals with all S_n at the same time.

3.3 maps

Any map $\varphi : p \to q$ can be written as $\omega \circ \pi$ where ω is a monotone map and π is a permutation. Therefore μ, η and τ generate the monoidal category of finite maps. Of course, the above equations are satisfied, but there are some extra equations relating μ and η with τ:

$$\begin{aligned} \tau \circ (\mu + 1) &= (1 + \mu) \circ (\tau + 1) \circ (1 + \tau), & \tau \circ (\eta + 1) &= 1 + \eta, \\ \tau \circ (1 + \mu) &= (\mu + 1) \circ (1 + \tau) \circ (\tau + 1), & \tau \circ (1 + \eta) &= \eta + 1, \\ & \mu \circ \tau = \mu. \end{aligned}$$

These equations are pictured as follows:



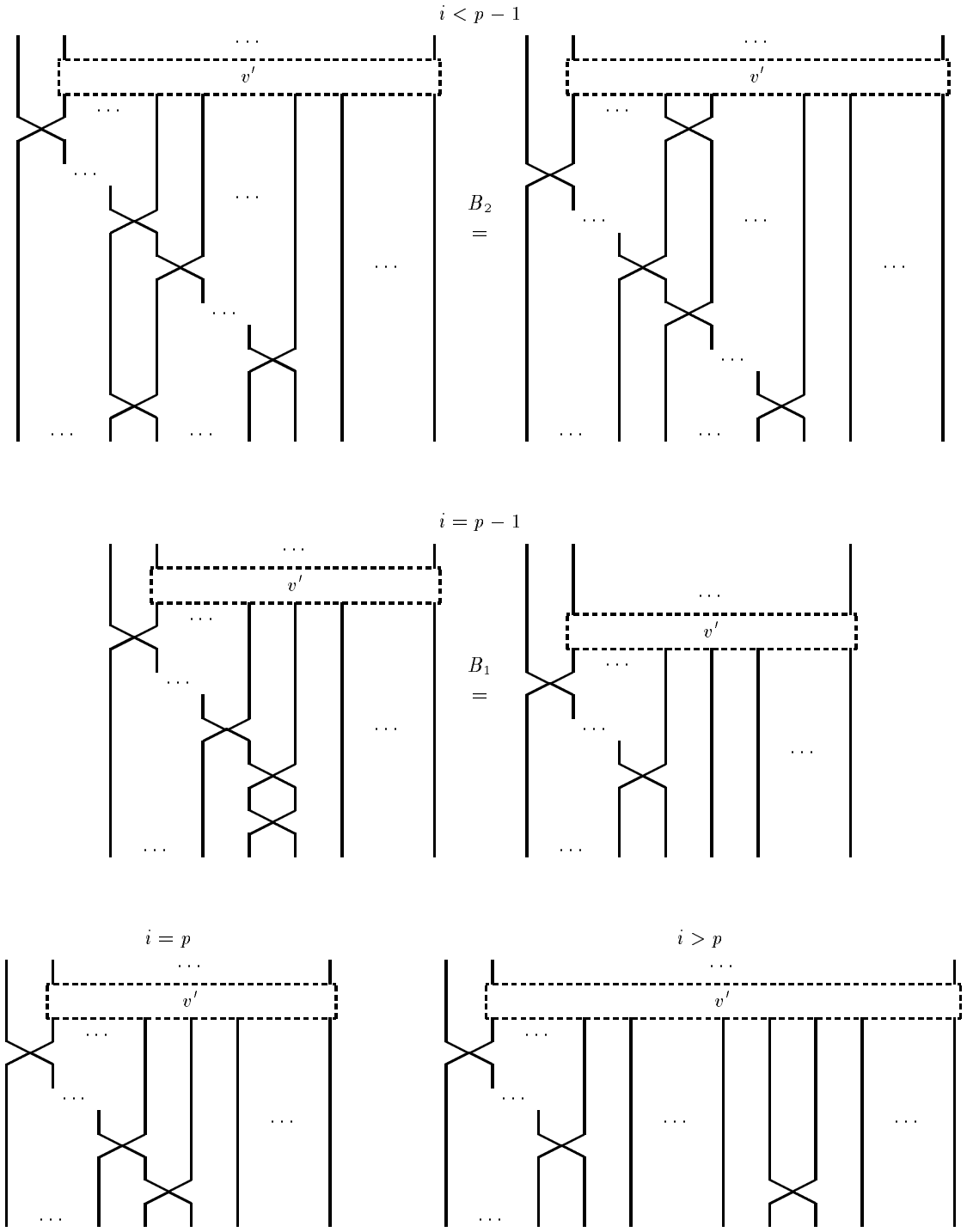
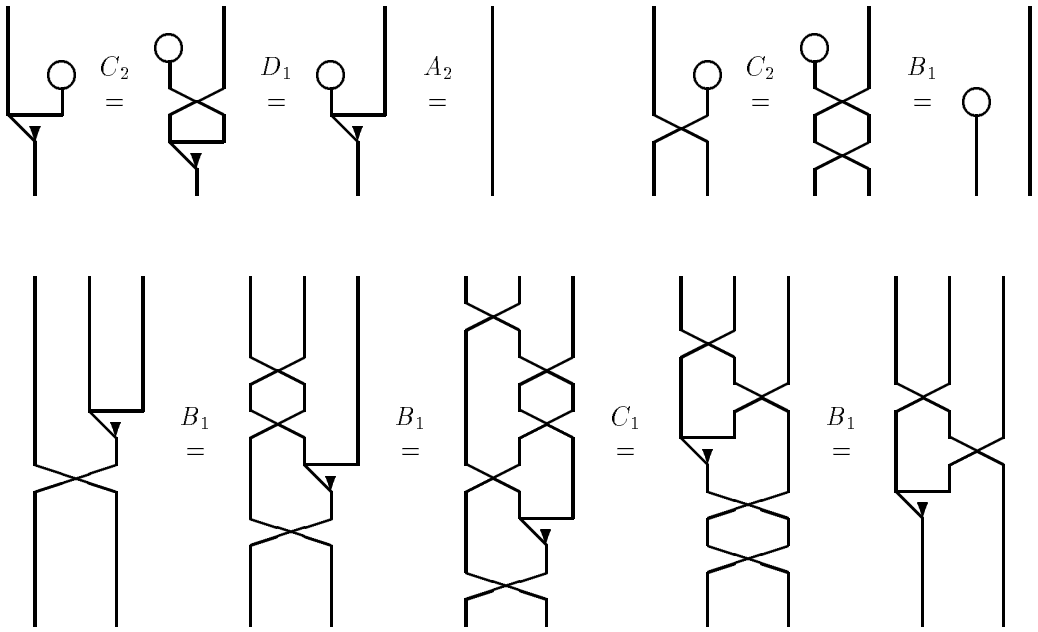
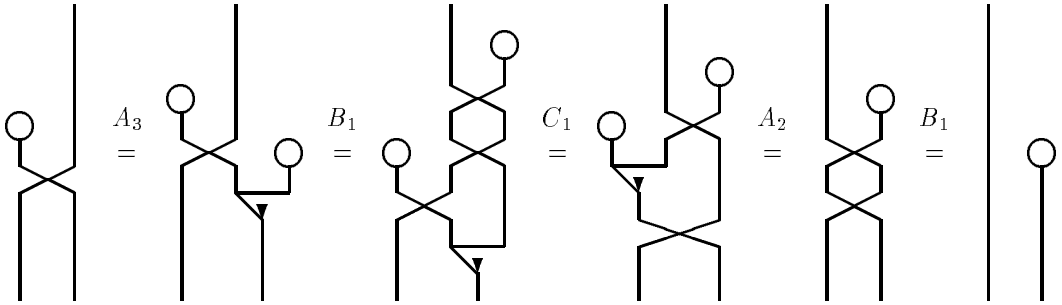


Fig. 1. Reducing a diagram of permutation to its canonical form (induction step)

Notice that (A_3), (C_3) and (C_4) are in fact superfluous because they are consequences of the other equations:



Alternatively, we could keep (A_3), and (C_2) would be derivable:



To show that we have a 2-presentation of the finite maps, we could start from the decomposition $\varphi = \omega \circ \pi$, as in [Bur93]. However, the permutation π is not uniquely determined by φ, the simplest counterexample being given by the equation (D_1). In fact, π is uniquely determined if we add the extra requirement that it does not permute any pair x, y such that $\varphi(x) = \varphi(y)$. Here, we shall avoid this complication by using a direct argument, without going through the 2-presentations of monotone maps and permutations.

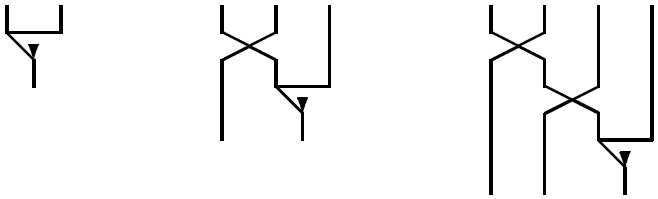
Let $\eta_n = \eta + \ldots + \eta$ be the unique map from 0 to n. It can be represented by the following diagram h_n:



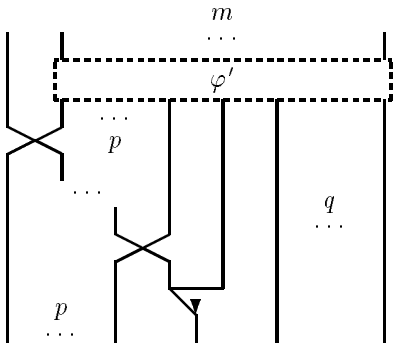
Lemma 5. *For any $u : p \to q$ in $\mathbf{D}(\mu, \eta, \tau)$, the diagram $u \circ \eta_p$ is equivalent to η_q modulo (A_2) and (C_2).*

This is easily proved by induction on the length of u. As a special case, since η_0 is the empty diagram id_0, we get the fact that any diagram $u : 0 \to n$ is equivalent to η_n.

Now, let $\gamma_n : n+2 \to n+1$ be the map defined by $\gamma_n(1) = n+1$ and $\gamma_n(i) = i-1$ for $i = 2, \ldots, n+2$. It can be represented by a diagram c_n with $n+1$ layers. For instance, here are c_0, c_1 and c_2:



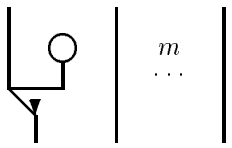
Lemma 6. *(decomposition of a map) Any map $\varphi : m+1 \to n$ can be uniquely written as $(\gamma_p + q) \circ (1 + \varphi')$ with $\varphi' : m \to n$*



This is obvious (take $p = \varphi(1) - 1$ and $q = n - p - 1$). Applying this decomposition inductively, we get a canonical diagram $\widehat{\varphi}$ for every map $\varphi : m \to n$ (if $m = 0$, take $\widehat{\varphi} = \eta_n$). In particular, this shows that μ, η and τ generate the finite maps.

Lemma 7. *Any diagram $u : m+1 \to n$ in $\mathbf{D}(\mu, \eta, \tau)$ is equivalent to a diagram of the form $(c_p + q) \circ (1 + u')$ modulo (A_1), (A_3), (B_1), (B_2), (C_1), (C_3), (C_4) and (D_1).*

Proof. By induction on the length of u. If u is id_{m+1}, then u is equivalent to $(c_0 + n) \circ (1 + \eta + n)$ by (A_3):



Otherwise, u is of the form $(i + \tau + j) \circ v$, or $(i + \mu + j) \circ v$, or $(i + \eta + j) \circ v$, and by induction hypothesis, v is equivalent to $(c_p + q) \circ (1 + v')$. If $i > p$, then u is already in the required form



Fig. 2. Two derivable equations

(modulo isotopy). Otherwise, there are several cases to consider, as in the proof of lemma 4. We need (B_2), (C_3), (C_4), (C_1), (A_1) and two other equations which are derivable (figure 2). □

We conclude that $\mathbf{F}$ is finitely presented by μ, η, τ and the equations of figure 3. The dual $\mathbf{F}^{op}$ of $\mathbf{F}$ is the category $\mathbf{T}(\emptyset)$ of section 1, which is also the free cartesian category generated by a single object. The fact that $\mathbf{F}^{op}$ is finitely presentable is the first step towards the reduction of the cartesian structure to a monoidal one.

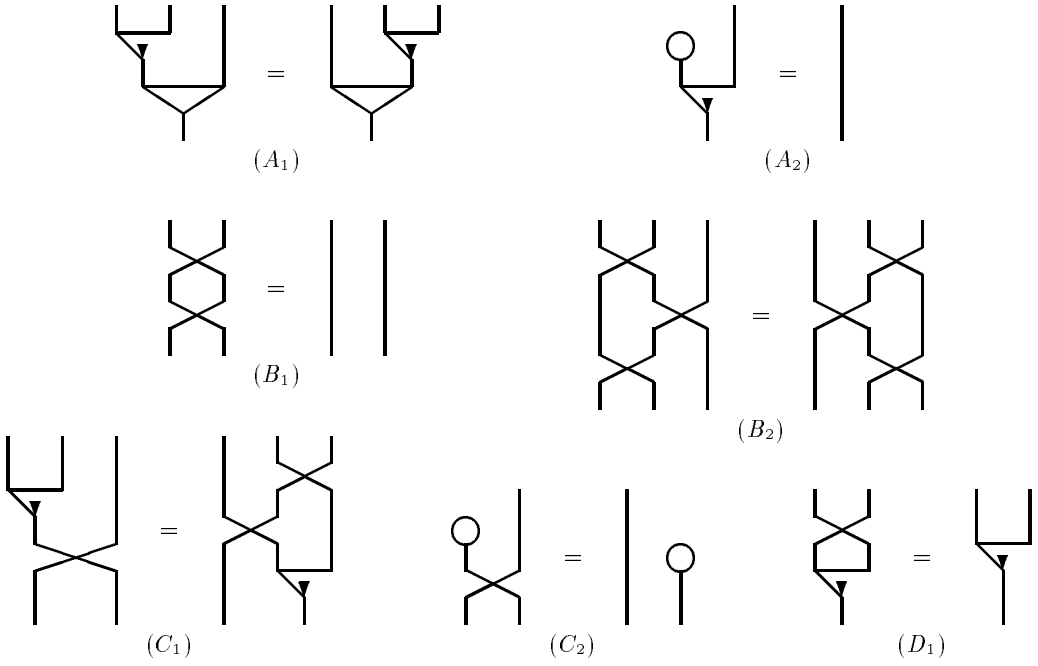


Fig. 3. 2-presentation of the finite maps

4 Presentation of the finite relations

We consider now the (strict) monoidal category of finite relations:

- The objects are natural numbers.
- A morphism $\rho : p \to q$ is a relation between $\{1, \ldots, p\}$ and $\{1, \ldots, q\}$.

Addition for relations is defined in the same way as for finite maps. Furthermore, we have a contravariant involution which maps a relation $\rho : p \to q$ to its dual $\rho^* : q \to p$. We shall draw upside down cells for μ^* and η^* (τ being its own dual):



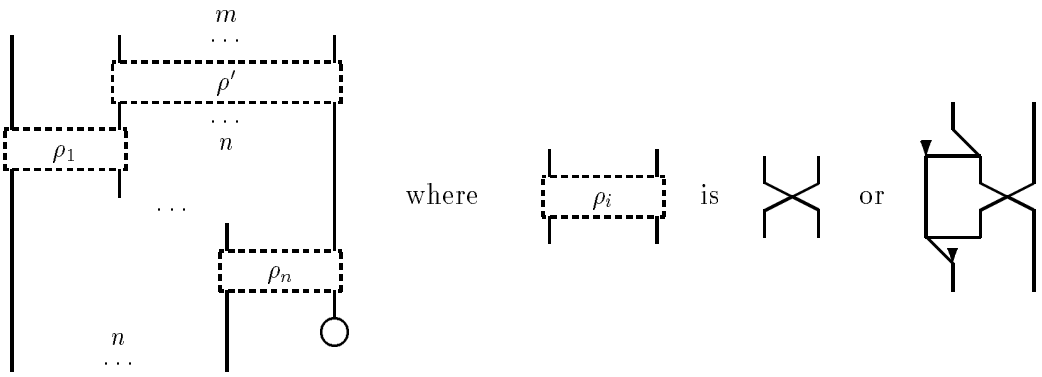
Any relation $\rho : p \to q$ can be written as $\psi \circ \varphi^*$ where $\varphi : r \to p$, $\psi : r \to q$ are finite maps, and r is the cardinality of (the graph of) ρ. Therefore μ, η, μ^*, η^* and τ generate finite relations.

We have all equations for finite maps, those obtained by duality and some extra ones relating μ and η with μ^* and η^* (figure 4). To show that we have a 2-presentation of the finite relations, we follow the same method as for finite maps.

Lemma 8. *For any $u : p \to q$ in $\mathbf{D}(\mu, \eta, \mu^*, \eta^*, \tau)$, the diagram $u \circ \eta_p$ is equivalent to η_q modulo (A_2), (C_2), (E_2^*) and (E_3).*

In particular, any diagram $u : 0 \to n$ is equivalent to η_n. Now, let $\kappa : 2 \to 2$ be the relation consisting of the pairs $(1, 1)$, $(1, 2)$ and $(2, 1)$.

Lemma 9. *(decomposition of a relation) Any relation $\rho : m + 1 \to n$ can be uniquely written as $(n + \eta^*) \circ ((n-1) + \rho_n) \circ \cdots \circ (\rho_1 + (n-1)) \circ (1 + \rho')$ with $\rho' : m \to n$ and where each $\rho_i : 2 \to 2$ is τ or κ.*



In this decomposition, ρ_i is κ when the pair $(1, i)$ belongs to (the graph of) ρ. Applying this decomposition inductively, we get a canonical diagram $\widehat{\rho}$ for every relation $\rho : m \to n$. In particular, this shows that μ, η, μ^*, η^* and τ generate finite relations.

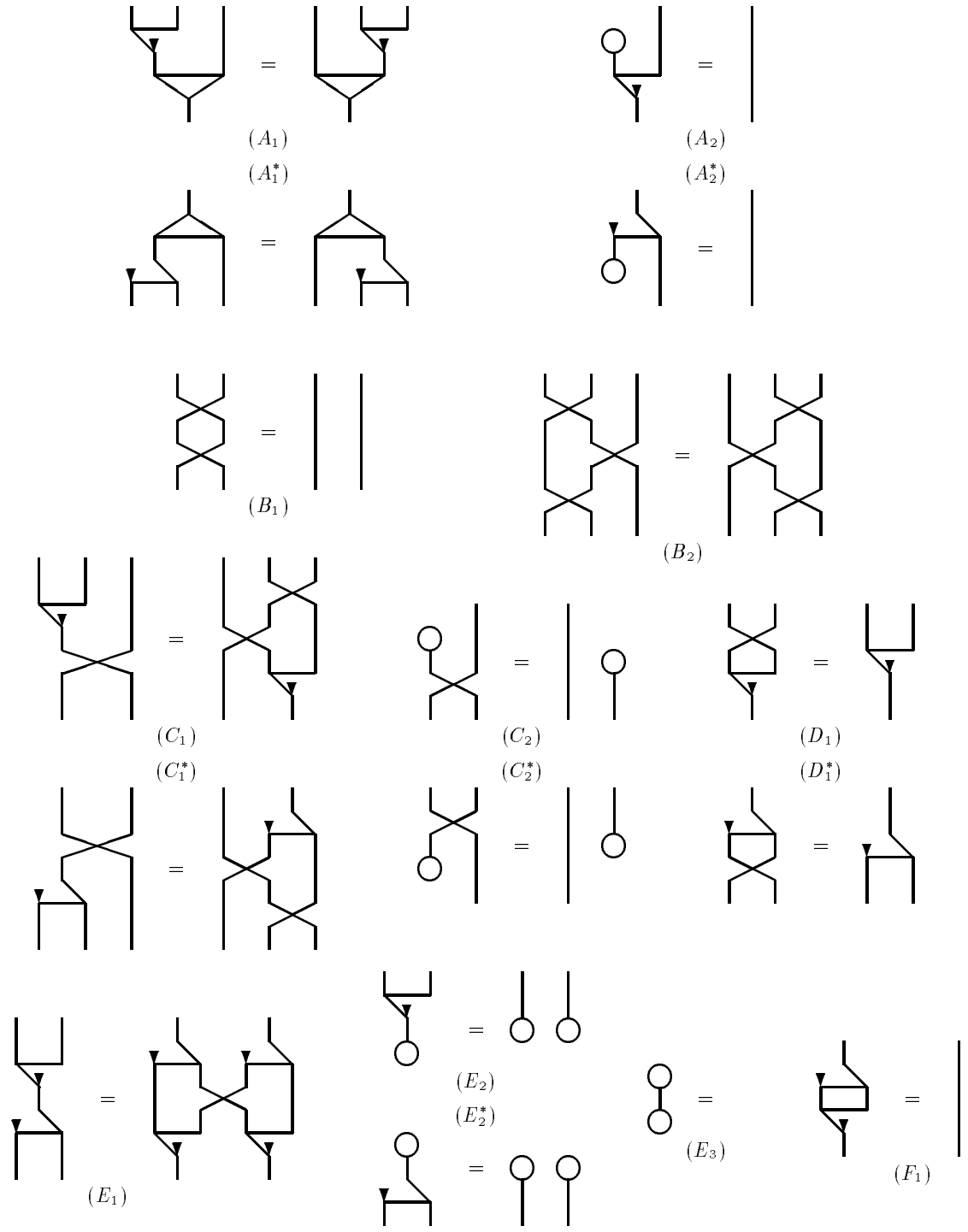


Fig. 4. 2-presentation of the finite relations

Lemma 10. *Any diagram $u : m+1 \to n$ in $\mathbf{D}(\mu, \eta, \mu^*, \eta^*, \tau)$ is equivalent to a diagram of the above form modulo the equations of figure 4.*

Proof. By induction on the length of u. If u is id_{m+1}, we use the derivable equations (A_3), (A_3^*) and (C_3^*). Otherwise, we proceed as in the proof of lemma 7, using (B_2), (C_3), (C_4), (C_1^*), (C_2^*) and eight other equations which are derivable. □

Using the same argument as for finite maps, we conclude that the monoidal category of finite relations is finitely presented by μ, η, μ^*, η^*, τ and the equations of figure 4. It is possible to show that, if we remove the equation (F_1) from the list of figure 4, we get a 2-presentation of the finite multirelations[2].

5 Algebraic theories

Here we consider a set Σ of symbols with arities, and a set $\mathcal{E}$ of equations on Σ.

5.1 syntax

Like any (strict) cartesian category, the category $\mathbf{T}(\Sigma)$ of section 1 has a (strict) monoidal structure:

- The monoidal product of p by q is $p+q$.
- If $U : p \to q$ and $U' : p' \to q'$ are morphisms, then $U + U' : p + p' \to q + q'$ is defined by the sequence $u_1, \ldots, u_q, v_1, \ldots, v_{q'}$ where v_i is $u'_i[x_{p+1}/x_1, \ldots, x_{p+p'}/x_{p'}]$.

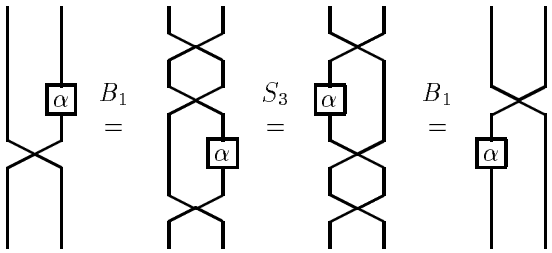
In the special case where Σ is empty, we get the dual of the category of finite maps. Therefore, it is natural to consider three generators $\delta : 1 \to 2$, $\varepsilon : 0 \to 1$ and $\sigma : 2 \to 2$, respectively given by x_1, x_1, the empty sequence and x_2, x_1. They correspond to μ, η and τ by duality.

In the general case, an extra generator $\alpha : p \to 1$, given by the term $\alpha(x_1, \ldots, x_p)$, is needed for each symbol α of arity p. Beside the seven equations obtained by duality from the 2-presentation of the finite maps (figure 5), there are equations relating δ, ε and σ with each symbol of Σ:

$$\delta \circ \alpha = (\alpha + \alpha) \circ \delta_n, \qquad \varepsilon \circ \alpha = \varepsilon_n,$$

$$\sigma \circ (\alpha + 1) = (1 + \alpha) \circ \sigma_{n,1}, \qquad \sigma \circ (1 + \alpha) = (\alpha + 1) \circ \sigma_{1,n},$$

where $\delta_n : n \to 2n$, $\varepsilon_n : n \to 0$ and $\sigma_{p,q} : p+q \to p+q$ are respectively given by $x_1, \ldots, x_n, x_1, \ldots, x_n$, the empty sequence and $x_{p+1}, \ldots, x_{p+q}, x_1, \ldots, x_p$. Those equations are pictured in figure 6 for symbols of arity ≤ 2. Notice that (S_4) is in fact superfluous, since it is a consequence of (S_3) and (B_1). In the unary case, we have indeed the following derivation:



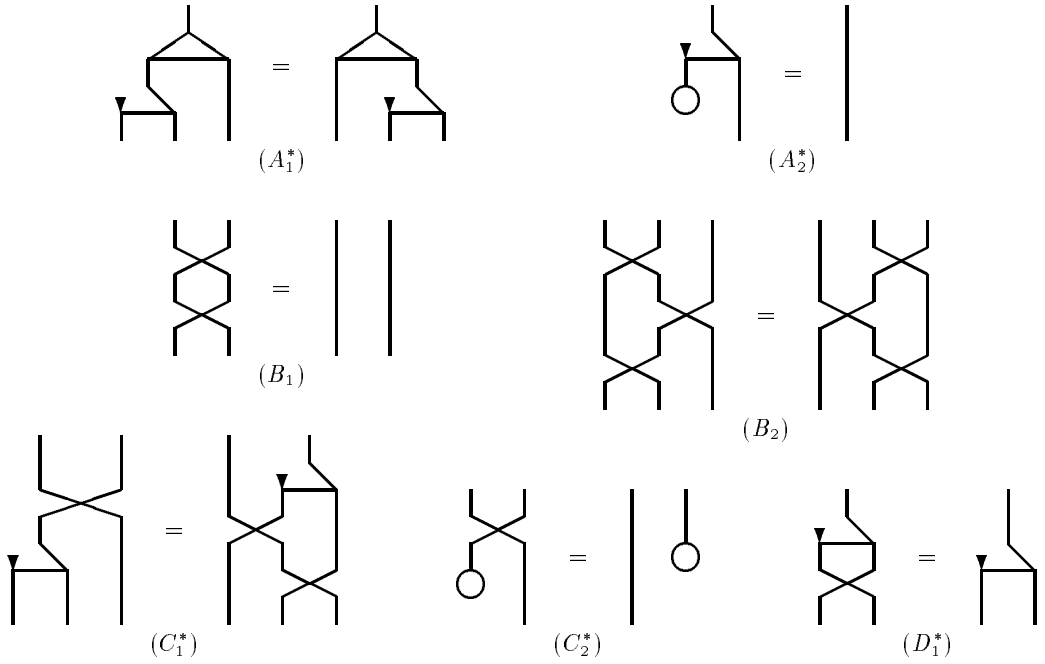


Fig. 5. Equations for δ, ε and σ

The other cases have been handled in 3.3.

We shall say that a morphism $U : m \longrightarrow n$ is *strongly linear* if the variables $x_1, \ldots, x_m$ appear exactly once and in order. Clearly, such a morphism is represented by a diagram built only with symbols of Σ, and this diagram is unique. Furthermore, we have the following decomposition:

Lemma 11. *Any morphism U in $\mathbf{T}(\Sigma)$ can be uniquely written as $W \circ V$ where V is (the dual of) a finite map and W is strongly linear.*

Now, let Γ be the alphabet consisting of δ, ε and σ.

Lemma 12. *Modulo (S_1), (S_2) and (S_3) for all $\alpha \in \Sigma$, any diagram u in $\mathbf{D}(\Gamma \cup \Sigma)$ is equivalent to a diagram of the form $w \circ v$ where v is in $\mathbf{D}(\Gamma)$ and w is in $\mathbf{D}(\Sigma)$.*

This is proved by induction on the length of u. Using those lemmas and the 2-presentation of the finite maps in section 3, we conclude that $\mathbf{T}(\Sigma)$ is finitely presented by $\Gamma \cup \Sigma$ and the equations of figures 5 and 6. Similarly, $\mathbf{T}(\Sigma; \mathcal{E})$ is a strict monoidal category which can be finitely presented by means of the same generators and equations as $\mathbf{T}(\Sigma)$, plus the equations of $\mathcal{E}$ expressed with diagrams. Therefore, if there are n functions symbols in Σ and m equations in $\mathcal{E}$, the 2-presentation has $n + 3$ generators and $m + 3n + 7$ equations.

[2] A multirelation between sets X and Y is a relation with multiplicities, *i.e.* a map from $X \times Y$ to $\mathbb{N}$.

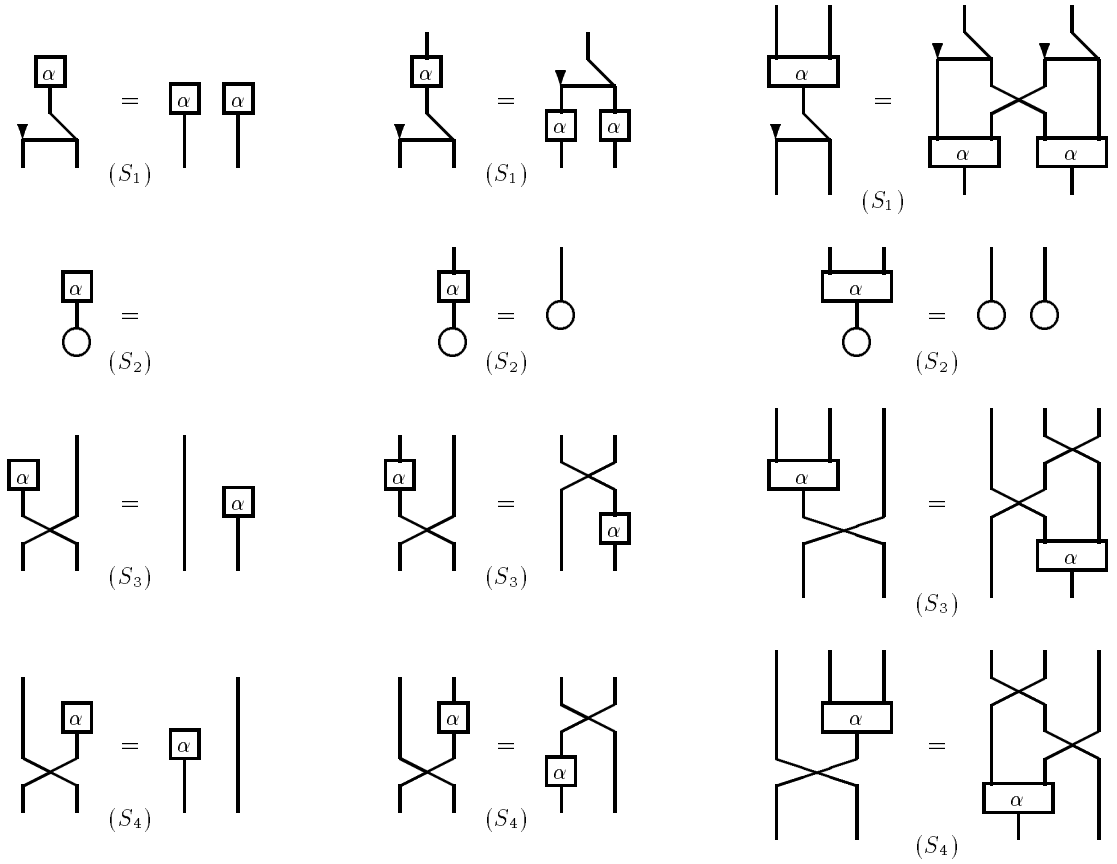


Fig. 6. Extra equations for symbols of arity 0, 1 and 2

5.2 semantics

In section 1, we have seen models of $\mathcal{E}$ as functors from $\mathbf{T}(\Sigma;\mathcal{E})$ to the category of sets which preserve the cartesian structure. Here, it would be more natural to consider functors preserving the monoidal structure. In fact, we shall see that the two notions coincide.

Let C be a cartesian category, for instance the category of sets. We can always assume that C is strict (see our remark at the end of section 2). Now, if A is an object of C, the *diagonal* $\delta = \langle \mathrm{id}_A, \mathrm{id}_A \rangle : A \to A \times A$, the *canonical morphism* $\varepsilon : A \to 1$ and the *symmetry* $\sigma = \langle \pi^2_{A,A}, \pi^1_{A,A} \rangle : A \times A \to A \times A$ satisfy all the equations of figure 5. The converse holds:

Lemma 13. *If $\delta : A \to A \times A$, $\varepsilon : A \to 1$ and $\sigma : A \times A \to A \times A$ are three morphisms satisfying the four equations (A^*_2), (A^*_3), (C^*_2) and (C^*_4), then δ is the diagonal, ε is the canonical morphism and σ is the symmetry.*

Proof. Clearly, we have $\mathrm{id}_A \times \varepsilon = \pi^1_{A,A}$ and $\varepsilon \times \mathrm{id}_A = \pi^2_{A,A}$. Therefore, the four equations give $\pi^2_{A,A} \circ \delta = \mathrm{id}_A$, $\pi^1_{A,A} \circ \delta = \mathrm{id}_A$, $\pi^2_{A,A} \circ \sigma = \pi^1_{A,A}$ and $\pi^1_{A,A} \circ \sigma = \pi^2_{A,A}$. □

This means that, in the case of a cartesian category, there is no choice for interpreting the three generators δ, ε and σ. As a consequence, any functor from $\mathbf{T}(\Sigma;\mathcal{E})$ to C which preserves the monoidal structure also preserves the cartesian structure. In particular, a model of $\mathcal{E}$ can be seen as a monoidal functor from $\mathbf{T}(\Sigma;\mathcal{E})$ to the category of sets.

If we replace the category of sets by another monoidal category, for instance a category of vector spaces, we get a nonstandard notion of model. In that case, the interpretation of δ, ε and σ is not forced by the equations.

6 Monoids, groups and Hopf algebras

Now we shall illustrate the previous section with some fundamental examples of algebraic theories.

6.1 monoids

The theory of monoids can be expressed by means of two function symbols $\cdot$ (of arity 2) and e (of arity 0). It is generated by the following equations:

$$(x \cdot y) \cdot z = x \cdot (y \cdot z), \qquad e \cdot x = x, \qquad x \cdot e = x.$$

If we picture $\cdot$ and e in the same way as μ and η in section 3, we notice that those equations correspond exactly to the 2-presentation of the monotone maps. Furthermore, writing (S_1), (S_2) and (S_3) for both function symbols, we get the equations (E_1), (E_2), (C_1), (E_2^*), (E_3) and (C_2) of figure 4. But here, we have (A_3), so that (C_2) is superfluous (see 3.3) and we obtain the 2-presentation of figure 7 for the theory of monoids.

For the theory of commutative monoids, we add the commutativity axiom $x \cdot y = y \cdot x$ which is interpreted as (D_1), and we obtain a 2-presentation of the finite multirelations. This is not surprising, since an element of the free commutative monoid generated by a finite set $\{a_1, \ldots, a_k\}$ can be uniquely written as $\Sigma_{i=1}^{k} n_i a_i$ where the coefficients $n_1, \ldots, n_k$ are natural numbers. If furthermore, we add the idempotency axiom $x \cdot x = x$ which is interpreted as (F_1), we obtain a 2-presentation of the finite relations.

6.2 groups

For groups, we need another function symbol $^{-1}$ of arity 1, which will be pictured as follows:



Beside the axioms for monoids, we have the following equations:

$$x^{-1} \cdot x = e, \qquad x \cdot x^{-1} = e,$$

which are pictured as follows:



(G_1) (G_2)

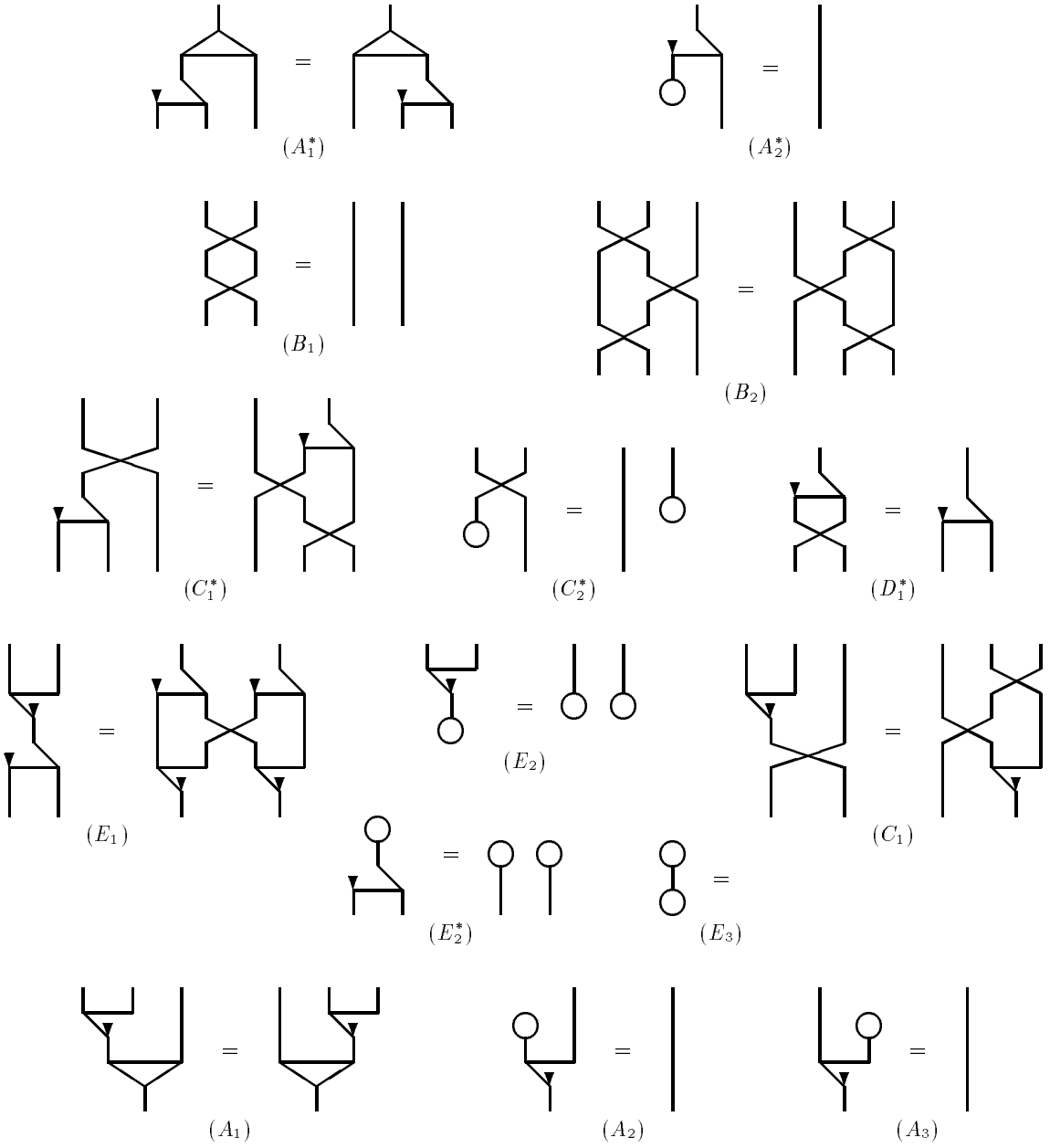
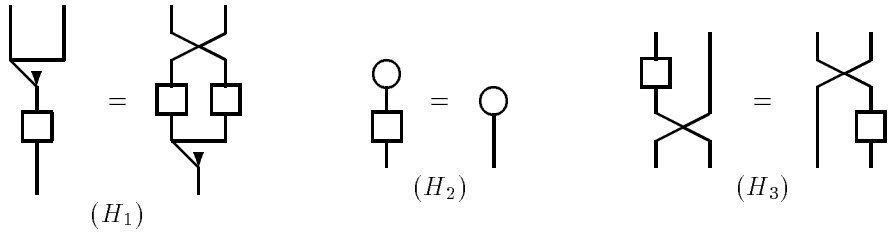


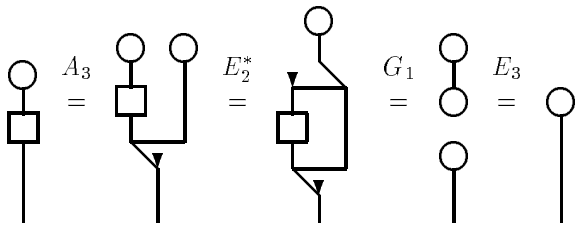
Fig. 7. Equations for monoids

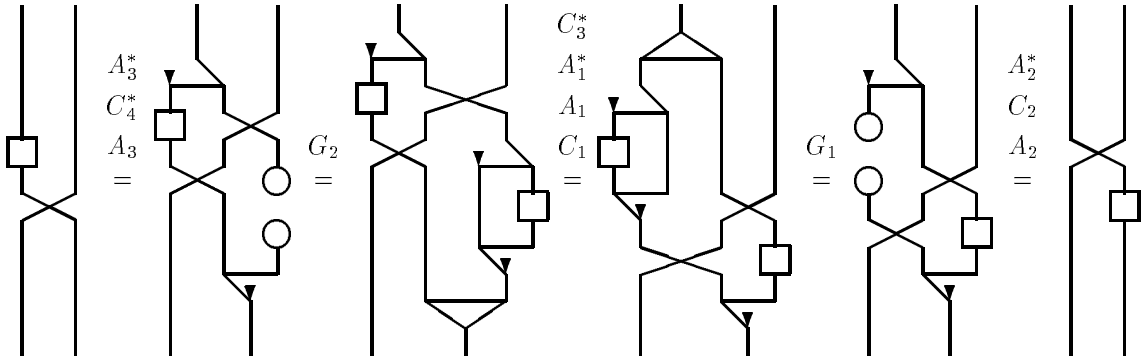
Of course, we should add (S_1), (S_2) and (S_3) for $^{-1}$, but we shall see that in fact, all three are derivable. Consider indeed the three following equations:



(H_1) (H_2) (H_3)

(H_1) and (H_2) are diagrammatic versions of the equations $(x \cdot y)^{-1} = y^{-1} \cdot x^{-1}$ and $e^{-1} = e$, whereas (H_3) is just (S_3) for $^{-1}$. Here are derivations for (H_2) and (H_3), using the equations (A_3^*), (C_3^*) and (C_4^*) which are themselves derivable:





It would be tedious to derive (H_1) directly. Instead, let us see how $(x \cdot y)^{-1} = y^{-1} \cdot x^{-1}$ is derived from the five axioms of the theory of groups. First we show that $(y^{-1} \cdot x^{-1}) \cdot (x \cdot y) = e$ as follows:

$$(y^{-1} \cdot x^{-1}) \cdot (x \cdot y) = y^{-1} \cdot ((x^{-1} \cdot x) \cdot y) = y^{-1} \cdot (e \cdot y) = y^{-1} \cdot y = e.$$

Then we use the fact that $u \cdot v = e$ implies $u = v^{-1}$, which is proved as follows:

$$u = u \cdot e = u \cdot (v \cdot v^{-1}) = (u \cdot v) \cdot v^{-1} = e \cdot v^{-1} = v^{-1}.$$

The first part of this reasoning is pictured in figure 8. For the second part, we assume that $u, v : 2 \to 1$ are arbitrary diagrams satisfying the three equations (1), (2) and (3) of figure 9. The equation (1) is just the diagrammatic version of $u \cdot v = e$ in the case of terms with two variables. If u and v are the diagrams corresponding to $y^{-1} \cdot x^{-1}$ and $x \cdot y$, all three hypotheses are satisfied, (2) and (3) being respectively (E_1) and (E_2). We conclude that (H_1) is derivable.

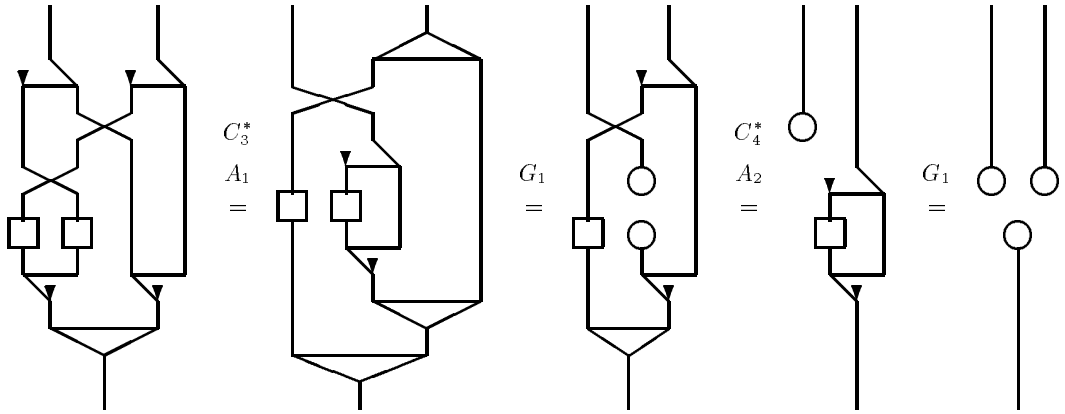


Fig. 8. Diagrammatic proof for $(y^{-1} \cdot x^{-1}) \cdot (x \cdot y) = e$

So far, we have not used (D_1^*), which means that all our equations can be turned upside down. Therefore, the dual (H_1^*) of (H_1) is also derivable, and from it we derive (S_1) for $^{-1}$, using (D_1^*), (H_3) and (B_1). Similarly we derive (H_2^*), that is (S_2) for $^{-1}$.

Alternatively, it is possible to show that (G_2) is derivable from (G_1), (S_1), (S_2) and the equations for monoids. This corresponds to the well-known fact that, in the traditional formulation, the equation $x \cdot x^{-1} = e$ is derivable from the other ones.

6.3 Hopf algebras

The theory of groups is almost self dual. There are two ways of making it perfectly symmetrical: adding (D_1), in which case we get the theory of commutative groups, or removing (D_1^*), in which case we get the theory of *Hopf algebras*. In the latter case, we do not get an algebraic theory in the sense of section 5. It is clear indeed that (D_1^*) is not derivable, because otherwise its dual (D_1) would be derivable by symmetry, and this would imply that all groups are commutative. The equation (S_1) for $^{-1}$ is not derivable either, but (H_1), (H_2), (H_3) and their duals are still derivable.

By lemma 13, a Hopf algebra in the category of sets with cartesian product (or in any cartesian category) is just a group. So we must look for models in monoidal categories which are not cartesian. Traditionally, Hopf algebras leave in a symmetric monoidal category, for instance a category of vector spaces, and σ is interpreted by the symmetry of the monoidal product. In that case (B_1), (B_2), (C_1) and (C_1^*) are automatically satisfied.

7 More examples

7.1 braids and knots

A braid is like a permutation, but where you distinguish between overcrossing and undercrossing:



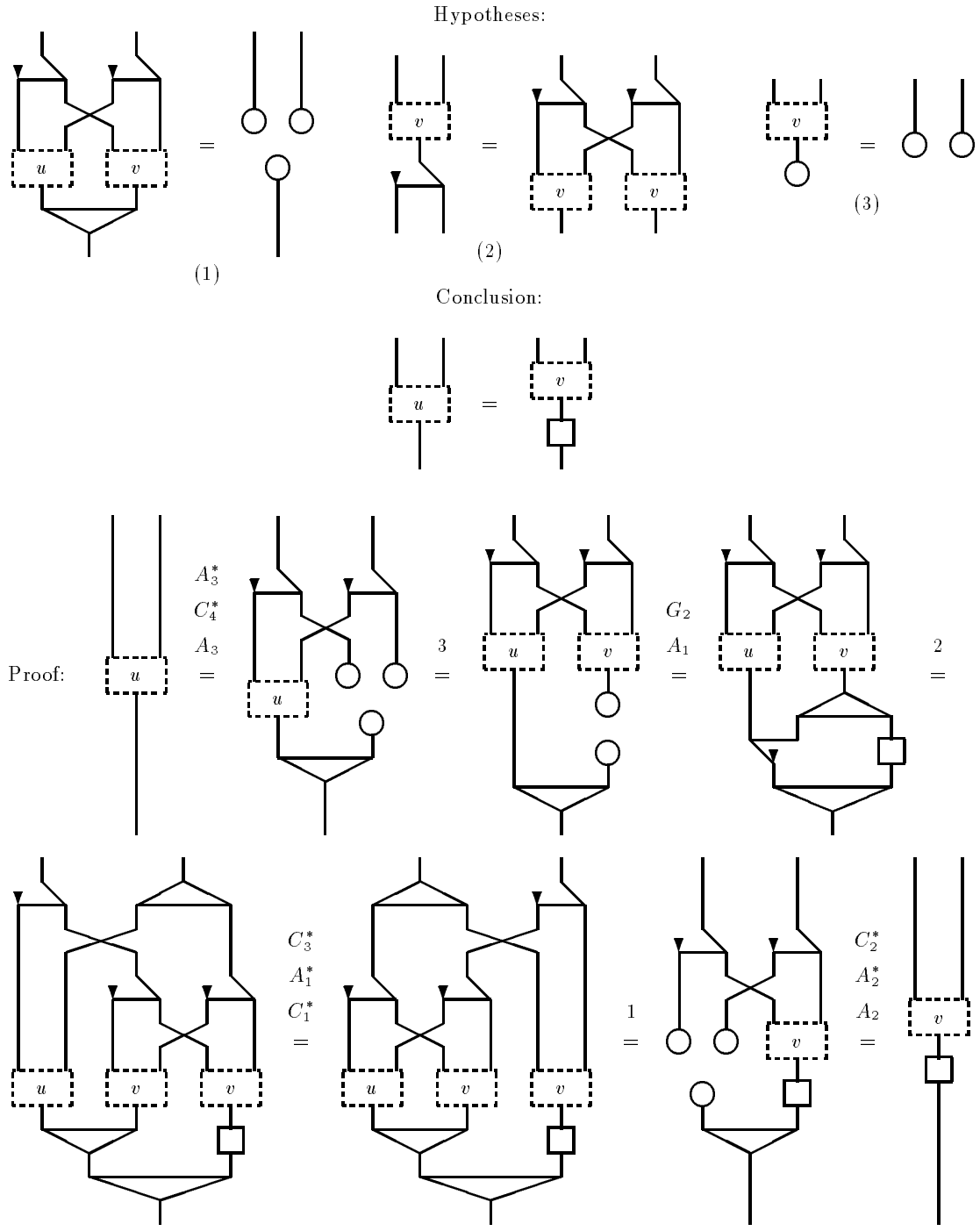
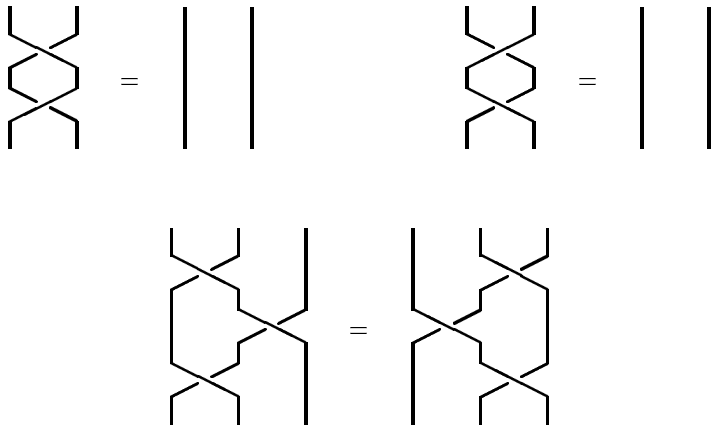


Fig. 9. Diagrammatic proof for $u \cdot v = e$ implies $u = v^{-1}$

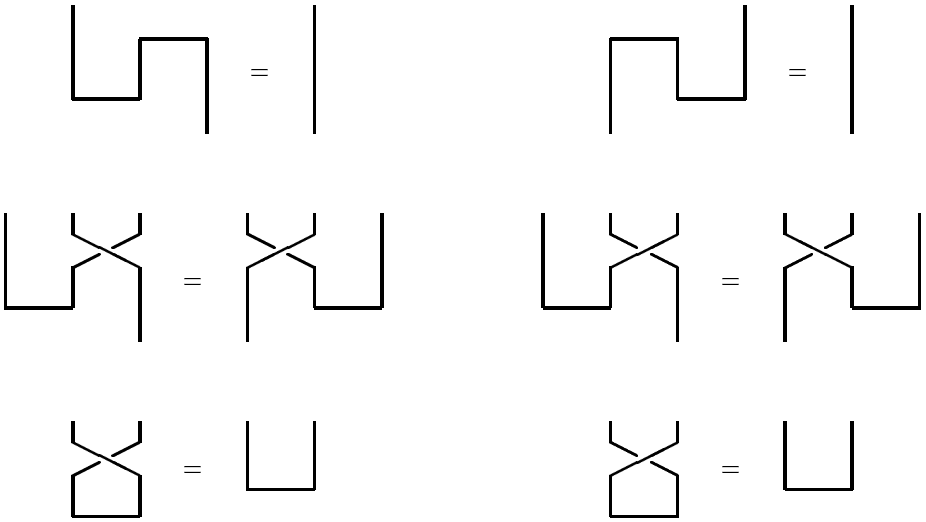
A 2-presentation of braids is obtained from the 1-presentation of the braid group B_n just as a 2-presentation of permutations is obtained from the 1-presentation of the symmetric group S_n:



For knots, two extra generators are needed:



A diagram built with these two generators and those for braids represents a *tangle*. A knot is a special kind of tangle. The extra equations for tangles are the following:



This 2-presentation is given in [FrY89], where several variants are also considered such as oriented tangles, tangles up to regular isotopy and framed tangles. Algebraic invariants for knots can be seen as models of those theories.

7.2 2-categorical reasoning

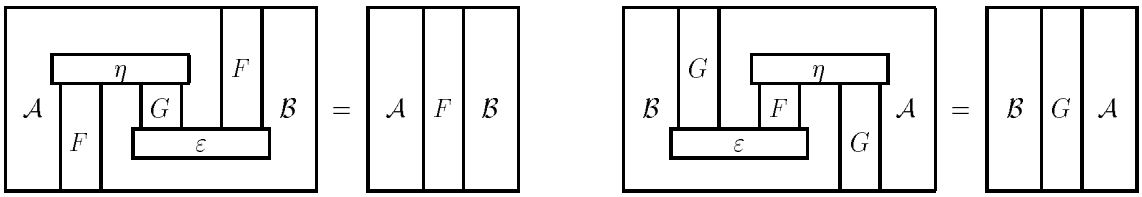
In all previous examples, we have considered monoidal categories generated by a single object. In the case of a monoidal category generated by several objects, wires in diagrams must be labeled or colored.

The next step consists in replacing monoidal categories by 2-categories. A strict monoidal category is indeed a 2-category with a single 0-cell. The algebra group of Santiago de Compostela has made an intensive use of 2-dimensional calculus in the 2-category of categories with *Rodeja diagrams*, but a nice feature of Penrose diagrams is the representation of 0-cells.

For instance, if $\mathcal{A}$ and $\mathcal{B}$ are categories, an *adjunction* between two functors $F : \mathcal{A} \to \mathcal{B}$ and $G : \mathcal{B} \to \mathcal{A}$ consists of two natural transformations $\eta : \mathrm{Id}_{\mathcal{A}} \to GF$ and $\varepsilon : FG \to \mathrm{Id}_{\mathcal{B}}$ satisfying the following equations (see [Mac71]):

$$\varepsilon F \circ F\eta = \mathrm{id}_F, \qquad G\varepsilon \circ \eta G = \mathrm{id}_G.$$

Those equations can be pictured as follows:



These diagrams are interpreted in the 2-category of categories. The wires has been thickened in order to be labeled by F or G, and each connected component of the background is labeled by $\mathcal{A}$ or $\mathcal{B}$. In bigger diagrams, the shape of the background may be quite complicated, and so, colors would be more suitable. The representation of 0-cells is not so nice in the other kinds of diagrams (figure 10).

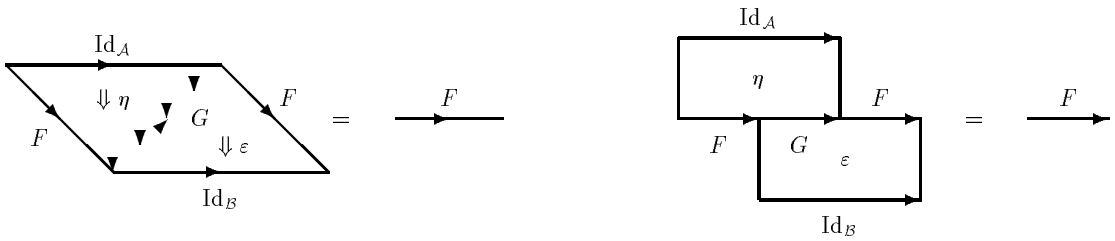


Fig. 10. Pasting diagrams and Rodeja diagrams for an adjunction

Conclusion

For many kinds of computations, the 2-dimensional calculus is very helpful, but drawing diagrams by hand becomes quickly tedious. We shall need computers for two purposes:

- manipulating diagrams without taking care of the details of the drawing, and with the assurance that our computations are correct (interactive reasoning),
- computing normal forms (automated reasoning).

In the second case, a theory of 2-dimensional rewriting is needed, with a generalization of the Knuth-Bendix completion. This approach is sketched in [Laf92].

References

[Bur93] **A. Burroni**, *Higher Dimensional Word Problem*, Theoretical Computer Science 115, 1993, pp. 43–62.

[EW67] **S. Eilenberg & J. B. Wright**, *Automata in general algebras*, Information and Control 115, 1967, pp. 452–470.

[FrY89] **P.J. Freyd & D.N. Yetter**, *Braided Compact Closed Categories with Application to Low Dimensional Topology*, Advances in Mathematics 77, 1989, pp. 156–182.

[JoS91] **A. Joyal & R. Street**, *The Geometry of Tensor Calculus*, Advances in Mathematics 88, 1991, pp. 55–112.

[Laf92] **Y. Lafont**, *Penrose diagrams and 2-dimensional rewriting*, Applications of Categories in Computer Science (ed. M.P. Fourman, P.T. Johnstone & A.M. Pitts), LMSLNS 177, Cambridge University Press, 1992, pp. 191–201.

[Law63] **F.L. Lawvere**, *Functorial Semantics of Algebraic Theories*, Proc. Nat. Acad. Sci. USA, 1963.

[Mac71] **S. Mac Lane**, *Categories for the Working Mathematician*, GTM 5, Springer-Verlag, 1971.

[PeR86] **R. Penrose & W. Rindler**, *Spinors and space-time, Vol. 1: Two-spinor calculus and relativistic fields*, Cambridge University Press, 1986.

[Poi85] **A. Poigné**, *Algebra Categorically*, Category Theory and Computer Programming (ed. D. Pitt, S. Abramsky, A. Poigné & D. Rydeheard), LNCS 240, Springer-Verlag, 1985, pp. 76–102.

This article was processed using the LaTeX macro package with LLNCS style