Coherence for Skew-Monoidal Categories

Tarmo Uustalu
Institute of Cybernetics at Tallinn University of Technology, Estonia
tarmo@cs.ioc.ee

We motivate a variation (due to K. Szlachányi) of monoidal categories called skew-monoidal categories where the unital and associativity laws are not required to be isomorphisms, only natural transformations. Coherence has to be formulated differently than in the well-known monoidal case. In my (to my knowledge new) version it becomes a statement of uniqueness of normalizing rewrites. We present a proof of this coherence proof and also formalize it fully in the dependently typed programming language Agda.

1 Introduction

Mac Lane's monoidal categories are ubiquitous in category theory, mathematics and computer science. One of their remarkable properties is the coherence theorem stating that in any monoidal category, any two parallel maps that are "formal" (in the sense that they are put together from the identity, composition, the tensor, the two unitors and the associator) are equal. In other words, in the free monoidal category over a given set of objects, any two maps with the same domain and codomain are equal. This is theorem is both beautiful and extremely useful. (There is also an simple existence condition for a map between two given objects in the free monoidal category.)

Szlachányi [16] has recently introduced variation of monoidal categories, called skew-monoidal categories. The important difference from monoidal categories is that the unitors and associator are not required to be isomorphisms. His study was motivated by structures from quantum physics. In my joint work with Altenkirch and Chapman [1], I ran into the same definition when generalizing monads to non-endofunctors.

In a free skew-monoidal category over a set of objects, uniqueness of parallel maps is lost. But it is still only reasonable to enquire, if some kind of coherence theorems are possible like they exist for many types of categories, e.g., Cartesian closed categories etc.

In this paper, I state and prove one such theorem. I obtained it by playing with Beylin and Dybjer's formalization [3] of Mac Lane's coherence theorem. Essentially, I looked at the high-level proof structure and checked what can be kept in the skew-monoidal case and what must necessarily be given up at least if one sticks to the same proof idea. The theorem states that maps to certain objects—"normal forms"—are unique. As a corollary, the same holds also for maps from "reverse-normal forms". For maps with different domains and co-domains no information is given.

I have formalized the results in the dependently typed programming (DTP) language Agda. I found it a very interesting exercise. Of course this is by no means uncommon with DTP projects, but a certainly project like this forces one to think carefully about deep matters in programming with the identity type in intensional type theory.

The structure of this short paper is as follows. I first define skew-monoidal categories, compare them to monoidal categories, and give some examples. Then I present the coherence statement and proof (as formalized in Agda), describe the rewriting intuition behind it and also hint at what everything means category-theoretically.

Submitted to:
MSFP 2014

The accompanying Agda formalization of the whole development and more (approximately 750 lines; self-contained, only propositional equality (the identity type) is taken from the library) is available from `http://cs.ioc.ee/~tarmo/papers/`.

2 Skew-monoidal categories

Skew-monoidal categories of Szlachányi [16] are a variation of monoidal categories, originally due to Mac Lane [14].

A *(right) skew-monoidal category* is a category $\mathscr{C}$ together with a distinguished object I, a functor $\otimes : \mathscr{C} \times \mathscr{C} \to \mathscr{C}$ and three natural transformations λ, ρ, α typed

$$\lambda_A : \mathsf{I} \otimes A \to A$$
$$\rho_A : A \to A \otimes I$$
$$\alpha_{A,B,C} : (A \otimes B) \otimes C \to A \otimes (B \otimes C)$$

satisfying the laws

$$\text{(a)}\quad \begin{array}{ccc} & \mathsf{I}\otimes\mathsf{I} & \\ {\scriptstyle\rho_\mathsf{I}}\nearrow & & \searrow{\scriptstyle\lambda_\mathsf{I}} \\ \mathsf{I} & = & \mathsf{I} \end{array} \qquad \text{(b)}\quad \begin{array}{ccc} (A\otimes\mathsf{I})\otimes B & \xrightarrow{\alpha_{A,\mathsf{I},B}} & A\otimes(\mathsf{I}\otimes B) \\ {\scriptstyle\rho_A\otimes B}\uparrow & & \downarrow{\scriptstyle A\otimes\lambda_B} \\ A\otimes B & = & A\otimes B \end{array}$$

$$\text{(c)}\quad \begin{array}{ccc} (\mathsf{I}\otimes A)\otimes B & \xrightarrow{\alpha_{\mathsf{I},A,B}} & \mathsf{I}\otimes(A\otimes B) \\ {\scriptstyle\lambda_A\otimes B}\searrow & & \swarrow{\scriptstyle\lambda_{A\otimes B}} \\ & A\otimes B & \end{array} \qquad \text{(d)}\quad \begin{array}{ccc} (A\otimes B)\otimes\mathsf{I} & \xrightarrow{\alpha_{A,B,\mathsf{I}}} & A\otimes(B\otimes\mathsf{I}) \\ {\scriptstyle\rho_{A\otimes B}}\nwarrow & & \nearrow{\scriptstyle A\otimes\rho_B} \\ & A\otimes B & \end{array}$$

$$\text{(e)}\quad \begin{array}{ccccc} (A\otimes(B\otimes C))\otimes D & & \xrightarrow{\alpha_{A,B\otimes C,D}} & & A\otimes((B\otimes C)\otimes D) \\ {\scriptstyle\alpha_{A,B,C}\otimes D}\uparrow & & & & \downarrow{\scriptstyle A\otimes\alpha_{B,C,D}} \\ ((A\otimes B)\otimes C)\otimes D & \xrightarrow{\alpha_{A\otimes B,C,D}} & (A\otimes B)\otimes(C\otimes D) & \xrightarrow{\alpha_{A,B,C\otimes D}} & A\otimes(B\otimes(C\otimes D)) \end{array}$$

A monoidal category is obtained, if λ, ρ, α are additionally required to be isomorphisms. Here this requirement is not made.

When dropping the requirement of isomorphisms from the definition of monoidal categories, the key question is of course which half of each of the three isomorphisms should be kept and how the laws (coherence conditions) should be stated. In a right skew-monoidal category λ "reduces", ρ "expands" and α "associates to the right". With this decision, the monoidal category laws (c), (d), (e) can be stated in only one way. But for (a) there are two choices ($\lambda_\mathsf{I} \circ \rho_\mathsf{I} = \mathsf{id}_\mathsf{I}$ and $\rho_\mathsf{I} \circ \lambda_\mathsf{I} = \mathsf{id}_{\mathsf{I}\otimes\mathsf{I}}$) and for (b) even three. The "correct" options turn to be those we have chosen.

Notice also that (a-e) are directed versions of the original Mac Lane axioms [14]. Later Kelly [9] discovered that (a), (c), (d) can be derived from (b) and (e). For skew-monoidal categories, this is not the case!

There is also an analogous notion of a *left skew-monoidal category*. It is important to realize that the opposite category $\mathscr{C}^{\mathrm{op}}$ of a right skew-monoidal category $\mathscr{C}$ is left skew-monoidal, not right-skew monoidal. But the "reverse" version $(\mathscr{C}^{\mathrm{op}})^{\mathrm{rev}}$ of $\mathscr{C}^{\mathrm{op}}$ (obtained by swapping the arguments of $\otimes$ and also λ and ρ) is right skew-monoidal.

In the rest of this text, my focus is on right skew-monoidal categories and speaking of a skew-monoidal category without specifying its skew, I mean the right skew.

Here are some examples of skew-monoidal categories.

Example 1 *We skew the monoid of addition of non-negative rational numbers. Consider the partial order of non-negative rationals as a thin category. This is made skew-monoidal with the unitors and the associator not isomorphisms by taking* $\mathsf{I} = 0$, $x \otimes y = x/2 + 2y$. *We have* $\lambda_x : x < 2x = 0/2 + 2x$, $\rho_x : x/2 + 2 \cdot 0 = x/2 < x$, $\alpha_{x,y,z} : (x/2+2y)/2 + 2z = x/4 + y + 2z = x/2 + y + 4z < x/2 + 2(y/2 + 2z)$. *Clearly* λ, ρ, α *are not equalities.*

Example 2 *The category of pointed sets and point-preserving functions has the following skew-monoidal structure. Take* $\mathsf{I} = (1, *)$ *and* $(X, p) \otimes (Y, q) = (X + Y, \mathsf{inl}\ p)$ *(notice the "skew" in choosing the point). We define* $\lambda_X : (1, *) \otimes (X, p) = (1 + X, \mathsf{inl} *) \to (X, p)$ *by* $\lambda_X\ (\mathsf{inl}\ *) = *$, $\lambda_X\ (\mathsf{inr}\ x) = x$ *(this is not injective). We let* $\rho_X : (X, p) \to (X + 1, p) = (X, p) \otimes (1, *)$ *by* $\rho_X\ x = \mathsf{inl}\ x$ *(this is not surjective). Finally we let* $\alpha_{X,Y,Z} : (X, p) \otimes (Y, p) \otimes (Z, r) = ((X + Y) + Z, \mathsf{inl}\ (\mathsf{inl}\ p)) \to (X + (Y + Z), \mathsf{inl}\ p) = (X, p) \otimes ((Y, p) \otimes (Z, r))$ *to be the obvious isomorphism.*

Example 3 *Given a monoidal category* $(\mathscr{C}, \mathsf{I}, \otimes)$. *Given also a lax monoidal comonad* $(D, \mathsf{e}, \mathsf{m})$ *on* $\mathscr{C}$. *The category* $\mathscr{C}$ *has a skew-monoidal structure given by* $\mathsf{I}^D = \mathsf{I}$, $A \otimes^D B = A \otimes D\ B$. *The unitors and associator are the following:*

$$\lambda^D_A = \ 1 \otimes D\,A \xrightarrow{\lambda_{DA}} D\,A \xrightarrow{\varepsilon_A} A$$
$$\rho^D_A = \ A \xrightarrow{\rho} A \otimes 1 \xrightarrow{A \otimes \mathsf{e}} A \otimes D\ 1$$
$$\alpha^D_{A,B,C} = \ (A \otimes D\ B) \otimes D\ C \xrightarrow{(A \otimes DB) \otimes \delta_C} (A \otimes D\ B) \otimes D\ (D\ C) \xrightarrow{\alpha_{A,DB,DC}} A \otimes (D\ B \otimes D\ (D\ C)) \xrightarrow{A \otimes \mathsf{m}_{B,C}} A \otimes D\ (B \otimes D\ C)$$

A similar right-skew monoidal category is also obtained with an oplax monoidal monad.

Example 4 *Consider two categories* $\mathscr{J}$ *and* $\mathscr{C}$ *and a functor* $J : \mathscr{J} \to \mathscr{C}$. *The functor category* $[\mathscr{J}, \mathscr{C}]$ *has a skew-monoidal structure given by* $\mathsf{I} = J$, $F \otimes G = \mathrm{Lan}_J F \cdot G$ *(assuming that the left Kan extension* $\mathrm{Lan}_J : \mathscr{C} \to \mathscr{C}$ *exists for every* $F : \mathscr{J} \to \mathscr{C}$*). The unitors and associator are the canonical natural transformations* $\lambda_F : \mathrm{Lan}_J\ J \cdot F \to F$, $\rho_F : F \to \mathrm{Lan}_J\ F \cdot J$, $\alpha_{F,G,H} : \mathrm{Lan}_J\ (\mathrm{Lan}_J\ F \cdot G) \cdot H \to \mathrm{Lan}_J\ F \cdot \mathrm{Lan}_J\ G \cdot H$. *This category becomes properly monoidal under certain conditions on* J. ρ *is an isomorphism, if* J *is fully-faithful.* λ *is an isomorphism, if* J *is dense. (This is the example from our relative monads work [1].)*

3 The coherence theorem

I now give a sufficient criterion for equality of two parallel maps in the free skew-monoidal category.

I first present the minimal technical development leading to a statement and proof the result, not commenting at all on what everything means category-theoretically. (This development follows the Agda formalization.) Then I give a rewriting "interpretation" of the story. Finally I explain the categorical meaning of the result.

The objects of the free symmetric monoidal category over a set Var of objects are given by the set of "object expressions" Tm defined inductively as follows:

$$\frac{X : \mathsf{Tm}}{`X : \mathsf{Tm}} \qquad \frac{}{\mathsf{I} : \mathsf{Tm}} \qquad \frac{A : \mathsf{Tm} \quad B : \mathsf{Tm}}{A \otimes B : \mathsf{Tm}}$$

The maps between two objects A and B are given by the set $A \Rightarrow B$ of "map expressions" quotiented by the relation $\doteq$ of "derivable equality". The former is defined inductively by the rules

$$\frac{}{\mathsf{id} : A \Rightarrow A} \qquad \frac{f : B \Rightarrow C \quad g : A \Rightarrow C}{f \circ g : A \Rightarrow C}$$

$$\frac{f : A \Rightarrow C \quad g : B \Rightarrow D}{f \otimes g : A \otimes B \Rightarrow C \otimes D}$$

$$\frac{}{\lambda : I \otimes A \Rightarrow A} \qquad \frac{}{\rho : A \Rightarrow A \otimes I} \qquad \frac{}{\alpha : (A \otimes B) \otimes C \Rightarrow A \otimes (B \otimes C)}$$

while the latter is defined inductively by the rules

$$\frac{}{f \doteq f} \qquad \frac{f \doteq g}{g \doteq f} \qquad \frac{f \doteq g \quad g \doteq h}{f \doteq h} \qquad \frac{f \doteq g \quad h \doteq k}{f \circ h \doteq g \circ k} \qquad \frac{f \doteq g \quad h \doteq k}{f \otimes h \doteq g \otimes k}$$

$$\frac{}{\mathsf{id} \circ f \doteq f} \qquad \frac{}{f \doteq f \circ \mathsf{id}} \qquad \frac{}{(f \circ g) \circ h \doteq f \circ (g \circ h)}$$

$$\frac{}{\mathsf{id} \otimes \mathsf{id} \doteq \mathsf{id}} \qquad \frac{}{(h \circ f) \otimes (k \circ g) \doteq h \otimes k \circ f \otimes g}$$

$$\frac{}{\lambda \circ \mathsf{id} \otimes f \doteq f \circ \lambda} \qquad \frac{}{\rho \circ f \doteq f \otimes \mathsf{id} \circ \rho} \qquad \frac{}{\alpha \circ (f \otimes g) \otimes h \doteq f \otimes (g \otimes h) \circ \alpha}$$

$$\frac{}{\lambda \circ \rho \doteq \mathsf{id}} \qquad \frac{}{\mathsf{id} \doteq \mathsf{id} \otimes \lambda \circ \alpha \circ \rho \otimes \mathsf{id}}$$

$$\frac{}{\lambda \circ \alpha \doteq \lambda \otimes \mathsf{id}} \qquad \frac{}{\alpha \circ \rho \doteq \mathsf{id} \otimes \rho} \qquad \frac{}{\alpha \circ \alpha \doteq \mathsf{id} \otimes \alpha \circ \alpha \circ \alpha \otimes \mathsf{id}}$$

We define "normal forms" of object expressions as the set Nf defined inductively by

$$\frac{}{\mathsf{J} : \mathsf{Nf}} \qquad \frac{N : \mathsf{Nf}}{X \,\text{‘}\!\otimes N : \mathsf{Nf}}$$

Normal forms embed into object expressions via the function $\mathsf{emb} : \mathsf{Nf} \to \mathsf{Tm}$ defined recursively by

$$\begin{aligned} &\mathsf{emb}\ \mathsf{J} = \mathsf{I} \\ &\mathsf{emb}\ (X \,\text{‘}\!\otimes N) = \text{‘}X \otimes N \end{aligned}$$

Let $[\![-]\!] : \mathsf{Tm} \to \mathsf{Nf} \to \mathsf{Nf}$ be the function defined recursively by the element of Tm by

$$\begin{aligned} &[\![\text{‘}X]\!]\ N = X \,\text{‘}\!\otimes N \\ &[\![\mathsf{I}]\!]\ N = N \\ &[\![A \otimes B]\!]\ N = [\![A]\!]\ ([\![B]\!]\ N) \end{aligned}$$

Every object expression is assigned a normal form with the normalization function $\mathsf{nf} : \mathsf{Tm} \to \mathsf{Nf}$ defined by

$$\mathsf{nf}\ A = [\![A]\!]\ \mathsf{J}$$

We can make some first important observations.

Lemma 1

1. *For any $f : A \Rightarrow B$ and $N : \mathsf{Nf}$, $[\![A]\!]\ N = [\![B]\!]\ N$.*
2. *For any $f : A \Rightarrow B$, $\mathsf{nf}\ A = \mathsf{nf}\ B$.*

Proof:

1. By induction on f.
2. Immediate from (1). □

Lemma 2 *For any* $N : \mathsf{Nf}$, $\mathsf{nf}\ (\mathsf{emb}\ N) = N$.

Proof: By induction on N. □

Lemma 3 *For any* $f : A \Rightarrow \mathsf{emb}\ N$, $\mathsf{nf}\ A = N$.

Proof: An immediate combination of Lemmata 1(2) and 2. □

Let now $\langle\!\langle - \rangle\!\rangle : \Pi A : \mathsf{Tm}.\ \Pi N : \mathsf{Nf}.\ A \otimes \mathsf{emb}\ N \Rightarrow \mathsf{emb}\ ([\![A]\!]\ N)$ be the function defined by

$$\begin{aligned}
&\langle\!\langle `X \rangle\!\rangle\ N = \mathsf{id}\\
&\langle\!\langle \mathsf{I} \rangle\!\rangle\ N = \lambda\\
&\langle\!\langle A \otimes B \rangle\!\rangle\ N = \langle\!\langle A \rangle\!\rangle\ ([\![B]\!]\ N) \circ \mathsf{id} \otimes \langle\!\langle B \rangle\!\rangle\ N \circ \alpha
\end{aligned}$$

To every object expression we assign a “normalizing” map expression the function $\mathsf{nm} : \Pi A : \mathsf{Tm}.\ A \Rightarrow \mathsf{emb}\ (\mathsf{nf}\ A)$ defined by

$$\mathsf{nm}\ A = \langle\!\langle A \rangle\!\rangle\ \mathsf{J} \circ \rho$$

We are ready to state our result.

Lemma 4 (Main lemma)

1. *For any* $f : A \Rightarrow B$ *and* $N : \mathsf{Nf}$, $\langle\!\langle A \rangle\!\rangle\ N \doteq \langle\!\langle B \rangle\!\rangle\ N \circ (f \otimes \mathsf{id})$. *(This statement is well-formed as* $[\![A]\!]\ N = [\![B]\!]\ N$ *by Lemma 1(1).)*
2. *For any* $f : A \Rightarrow B$, $\mathsf{nm}\ A \doteq \mathsf{nm}\ B \circ f$. *(This statement is well-formed as* $\mathsf{nf}\ A = \mathsf{nf}\ B$ *by Lemma 1(2).*

Proof:

1. By induction on f. This is a tedious but simple proof with six cases, some are tricky for formalization! (Read the Agda development.)

   Of course the proof relies on the equality of map expressions being induced by the five coherence conditions. All of them are needed and exactly in the versions chosen (remember that for conditions (a), (b) there were multiple inequivalent options).
2. Follows from (1). □

Lemma 5 *For any* $N : \mathsf{Nf}$, $\mathsf{nm}\ (\mathsf{emb}\ N) \doteq \mathsf{id}$. *(This statement is well-formed as* $\mathsf{nf}\ (\mathsf{emb}\ N) = N$ *by Lemma 2.)*

Proof: By induction on N. □

Lemma 6 (Main theorem) *For any* $f : A \Rightarrow \mathsf{emb}\ N$, $\mathsf{nm}\ A \doteq f$. *(This statement is well-formed, because* $\mathsf{nf}\ A = N$ *by Lemma 3).*

Proof: By combining Lemmata 4(2) and 5. □

Of course nothing prevents us from playing the reverse game. We can define a set Nf^{r} and functions $\mathsf{emb}^{\mathsf{r}} : \mathsf{Nf}^{\mathsf{r}} \to \mathsf{Tm}$, $[\![-]\!]^{\mathsf{r}} : \mathsf{Tm} \to \mathsf{Nf}^{\mathsf{r}} \to \mathsf{Nf}^{\mathsf{r}}$ and $\mathsf{nf}^{\mathsf{r}} : \mathsf{Tm} \to \mathsf{Nf}^{\mathsf{r}}$:

$$\frac{}{\mathsf{J}^{\mathsf{r}} : \mathsf{Nf}} \qquad \frac{R : \mathsf{Nf}^{\mathsf{r}}}{R \, `\!\otimes^{\mathsf{r}} X : \mathsf{Nf}^{\mathsf{r}}}$$

$$\begin{aligned} &\mathsf{emb}^{\mathsf{r}}\ \mathsf{J}^{\mathsf{r}} = \mathsf{I} \\ &\mathsf{emb}^{\mathsf{r}}\ (R \, `\!\otimes^{\mathsf{r}} X) = R \otimes `X \end{aligned}$$

$$\begin{aligned} &[\![`X]\!]^{\mathsf{r}}\ R = R \, `\!\otimes^{\mathsf{r}} X \\ &[\![\mathsf{I}]\!]^{\mathsf{r}}\ R = R \\ &[\![A \otimes B]\!]^{\mathsf{r}}\ R = [\![B]\!]^{\mathsf{r}}\ ([\![A]\!]^{\mathsf{r}}\ R) \end{aligned}$$

$$\mathsf{nf}^{\mathsf{r}}\ A = [\![A]\!]^{\mathsf{r}}\ \mathsf{J}^{\mathsf{r}}$$

Further we can define functions $\langle\!\langle - \rangle\!\rangle^{\mathsf{r}} : \Pi A : \mathsf{Tm}.\ \Pi N : \mathsf{Nf}.\ \mathsf{emb}^{\mathsf{r}}\ ([\![A]\!]^{\mathsf{r}}\ R) \Rightarrow \mathsf{emb}^{\mathsf{r}}\ R \otimes A$ and $\mathsf{nm}^{\mathsf{r}} : \Pi A : \mathsf{Tm}.\ \mathsf{emb}^{\mathsf{r}}\ (\mathsf{nf}^{\mathsf{r}}\ A) \Rightarrow A$ and theorems as above hold hold for them. Furthermore, $\mathsf{nf}\ A = \mathsf{nf}\ B$ if and only if $\mathsf{nf}^{\mathsf{r}}\ A = \mathsf{nf}^{\mathsf{r}}\ B$.

Thus we see that for two object expressions A and B to have exactly one map expression to them (up to $\doteq$), it suffices to have $A = \mathsf{emb}^{\mathsf{r}}\ R$ or $B = \mathsf{emb}\ N$ for some R or N (i.e., A in reverse normal form or B in normal form).

It is important to notice that this is merely a sufficient condition for a unique map expression between two object expressions. It is perfectly possible have a map expression between A and B, even if A is not a reverse normal form, B is not a normal form. The simplest example is $A = B = `X$, since we have the map $\mathsf{id} : `X \Rightarrow `X$.

At the same time it is easy to find pairs of object expressions A and B with $\mathsf{nf}\ A = \mathsf{nf}\ B$ (which is the same as $\mathsf{nf}^{\mathsf{r}}\ A = \mathsf{nf}^{\mathsf{r}}\ B$) with no or several map expressions between them.

A typical example of non-existence of a map between A and B with the same normal form is given by $A = `X \otimes ((`Y \otimes `Z) \otimes \mathsf{I})$ and $B = (`X \otimes `Y) \otimes (`Z \otimes \mathsf{I})$ (both have $`X \otimes (`Y \otimes (`Z \otimes \mathsf{I}))$ as their normal form).

Some examples of multiple maps:

- $\mathsf{id} \neq \rho \circ \lambda : \mathsf{I} \otimes I \Rightarrow \mathsf{I} \otimes I$,
- $\mathsf{id} \neq \rho \otimes \mathsf{id} \circ \mathsf{id} \otimes \lambda \circ \alpha : (`X \otimes I) \otimes `Y \Rightarrow (`X \otimes I) \otimes `Y$,
- $\lambda \neq \mathsf{id} \otimes \lambda : \mathsf{I} \otimes (\mathsf{I} \otimes `X) \Rightarrow (\mathsf{I} \otimes `X)$.

Rewriting interpretation Let us see what we have established from the rewriting perspective.

The elements of Tm can be thought of as *terms* made of I and $\otimes$ over Var. The elements of $A \Rightarrow B$ should be thought of as *rewrites* of A into B: λ, ρ, α are rewrite rules, $\otimes$ allows applying rewrite rules inside a term, id is the nil rewrite, $\circ$ is sequential composition of two rewrites. The relation $\doteq$ provides a congruence between two rewrites of one term to another.

With Nf we have carved out from the set of all terms Tm some terms that we have decided to consider in *normal form*. emb is the inclusion of this set Nf into Tm.

With $\mathsf{nf}\ A$ we have assigned to every term a particular normal form, which we define to be its (unique) normal form. Notice that this is an entirely rewriting-independent definition of normalization.

Lemma 1 says one term can only be written into another, if their normal forms are the same. Lemma 2 says that a normal form's normal form is itself. Lemma 3 is the obvious conclusion that, if a term rewrites to a normal form, it is that term's normal form.

With $\mathsf{nm}\ A$ we have at least one (canonical) rewrite of any term A to its normal form.

Lemma 4 says that the canonical normalizing rewrite $\mathsf{nm}\ A$ of a term A factors through all others rewrites of it. Lemma 5 tells says that the canonical normalizing rewrite of a normal form (into itself) is the nil rewrite.

Lemma 6 tells us that any normalizing rewrite of a term A is equal to the canonical normalizing rewrite $\mathsf{nm}\ A$. Thus all normalizing rewrites of A are equal.

Categorical meaning Categorically speaking, what we have proved is a relationship between two categories **Tm**, which is the free skew-monoidal category over Var, and **Nf**, which is the free strictly monoidal category over Var.

The category **Tm** has Tm as the set of objects, $(A \Rightarrow B)/\doteq$ the set of maps between objects A, B, id the identity, $\circ$ as composition, I the unit, $\otimes$ the tensor, λ, ρ and α the unitors and associator. The category **Nf** is discrete and has Nf (the set of lists over Var) as the set of objects.

emb is clearly a functor from **Nf** to **Tm** (with a trivial map mapping).

With $\mathsf{nf} : \mathsf{Tm} \to \mathsf{Nf}$ we provide an object mapping for a functor from **Tm** to **Nf**. Lemma 1, stating that, for any $f : A \to B$, we have $\mathsf{nf}\ A = \mathsf{nf}\ B$, tells us that nf with the constant identity map mapping is a functor from **Tm** to **Nf**.

Lemma 2, stating that $\mathsf{nf}\ (\mathsf{emb}\ N) = N$, establishes that identity has the correct type to be a counit, if nf were a left adjoint of emb ("$\varepsilon : L\ (R\ X) \to X$").

Lemma 3 concludes from Lemmata 1 and 2 that, if $f : A \Rightarrow \mathsf{emb}\ N$, then $\mathsf{nf}\ A = N$. This shows that constant identity is a good candidate for the left transpose operation of an adjunction. (The proof is nothing but the standard definition of left transpose from the counit ("$f^\dagger = \varepsilon \circ Lf$").

The polymorphic function $\mathsf{nm}\ \{A\} : \mathsf{Tm}.\ A \Rightarrow \mathsf{emb}\ (\mathsf{nf}\ A)$ is, by its typing, a candidate for the unit of an adjunction ("$\eta : X \to R\ (L\ X)$"). The main lemma (Lemma 4), stating that, if $f : A \Rightarrow B$, then $\mathsf{nm}\ A \doteq \mathsf{nm}\ B \circ f$, establishes that nm is a natural transformation ("$R\ (L\ f) \circ \eta = \eta \circ f$").

Lemma 5 stating $\mathsf{nm}\ (\mathsf{emb}\ N) = \mathsf{id}$, establishes one of the adjunction laws ("$R\ \varepsilon \circ \eta = \mathsf{id}$").

Lemma 6, the conclusion from Lemmata 4 and 5, stating that $f : A \Rightarrow \mathsf{emb}\ N$ implies $\mathsf{nm}\ A = f$, is a proof of the equivalent adjunction law in terms of the left transpose ("$Rf^\dagger \circ \eta = f$").

The other adjunction laws are trivial as **Nf** is a discrete category. Hence our coherence result is really establishing that nf and emb provide an adjunction between **Tm** and **Nf**.

In fact, much more can be proved. As already mentioned above, the category **Nf** is strictly monoidal, with J (the empty list) the unit and $\boxtimes$ (concatenation of lists) as the tensor. More, it is the free strictly monoidal category over Var. The functors $\mathsf{nf} : \mathbf{Tm} \to \mathbf{Nf}$ and $\mathsf{emb} : \mathbf{Nf} \to \mathbf{Tm}$ are lax skew-monoidal. Further, the unit (and trivially the counit) are lax skew-monoidal too, so the adjunction between **Tm** and Nf is a lax skew-monoidal adjuction. (The Agda development has the full proofs.)

Also, I have downplayed here a further fact (which the proof however relies on implicitly) that the adjunction between **Tm** and **Nf** factors through the discrete functor category $[\mathbf{Nf}, \mathbf{Nf}]$.

Equality of normal forms in Agda My development makes heavy use of equality reasoning on normal forms. Notice in particular that the statements of Lemmata 4–6 are only well-formed since Lemmata 1-3 (stating equalities of normal forms) hold.

In the Agda formalization, based on intensional type theory, I model equality of normal forms with the propositional equality (the identity type) on Nf. In the sense of the Agda development, the maps of **Nf** are exactly proofs of propositional equalities $N \equiv N'$. In particular, identity is refl and composition is trans. Discreteness is the uniqueness of identity proofs principle.

A consequence of this is that, if N and N' are equal only propositionally with a proof p, and not definitionally, then we cannot have a proof $f \doteq g$ for two map expressions $f : A \Rightarrow \mathsf{emb}\, N$ and $g : A \Rightarrow \mathsf{emb}\, N'$. What can hold is $\mathsf{subst}\ (A \Rightarrow -)\ p f \doteq g$. For example, one Agda formulation of Lemma 6 could be:

$$\Pi f : A \Rightarrow \mathsf{emb}\, N.\ \mathsf{subst}\ (A \Rightarrow -)\ (\mathsf{fnfemb}\ f)\ (\mathsf{nm}\, A) \doteq f$$

where fnfemb is the proof of Lemma 3, i.e., of $\Pi f : A \Rightarrow \mathsf{emb}\, N.\ \mathsf{nf}\, A \equiv N$, the left transpose operation of the adjunction between nf and emb.

Working with subst (or alternatives, like `with` and pattern-matching on propositional equality proofs, or `rewrite`) is tedious.

I was therefore relieved to find that, for this project, there is a neat alternative. Substitution by the codomain of a map expression based on $p : N \equiv N'$ can be replaced by postcomposition it with $\mathsf{femb}\ p : \mathsf{emb}\, N \to \mathsf{emb}\, N'$ defined by

$$\mathsf{femb}\ p = \mathsf{subst}\ (\mathsf{emb}\, N \Rightarrow -)\ p\ (\mathsf{id}\ \{N\})$$

(identity map expression with adjusted codomain).

For any $f : A \Rightarrow \mathsf{emb}\, N$ and $p : N \equiv N'$, it is the case that $\mathsf{femb}\ p \circ f \doteq \mathsf{subst}\ (A \Rightarrow \mathsf{emb}\ -)\ p\ f$.

In particular, Lemma 6 can say,

$$\Pi f : A \Rightarrow \mathsf{emb}\, N.\ \mathsf{femb}\ (\mathsf{fnfemb}\ f) \circ \mathsf{nm}\, A \doteq f$$

An additional advantage is that the map mapping part of the functor emb from **Nf** to **Tm**, which is otherwise obscured in the Agda formalization, becomes manifest. For example, Lemma 6 becomes explicitly of the form "$Rf^{\dagger} \circ \eta = f$", which we recognize as one of the adjunction laws.

4 Related work

Skew-monoidal categories were first studied as such by Szlachányi [16] in the context of structures for quantum computing. They immediately attracted the interest of Lack, Street, Buckley and Garner [10, 11, 5]. Lack and Street [11] proved a coherence theorem, which is different from the one here: they give a necessary and sufficient condition for equality two parallel maps of the free skew-monoidal category (but no decision algorithm; I do not know whether one would be easy to construct).

I first met a skew-monoidal category in my work with Altenkirch and Chapman [1] on relative monads: we noticed and made use of the skew-monoidal structure (non-endo)-functor categories. The context categories of Blute, Cockett and Seely [4] have the skew-monoidal category data and laws as part of the structure.

Some other weakened versions of monoidal categories with the unitors and associator not isomorphisms are the pseudocategories of Burroni [6] (like right-skew categories but with λ "expanding") and Grandis d-lax 2-categories [8] (α "associating to the left").

Laplaza [12] studied coherence for a version of semimonoidal categories with associativity not an isomorphism in the 1970s.

Coherence is generally related to equational reasoning and through that to term rewriting. Beke [2] considered replacing the question of uniqueness of equality proofs (equality of parallel maps in categories with isomorphisms only) with uniqueness of (normalizing) rewrites of maps (equality of parallel maps in a more general setting). He also asked whether coherence could be proved for structures like skew-monoidal categories.

5 Conclusion and future work

My main conclusion is the same as Beke's [2]: for skew-structured categories (with natural transformations instead of natural isomorphisms), coherence is not about uniqueness of equality proofs in an equational theory, but about uniqueness of rewriting (typically normalization) proofs in a rewrite system. A skew coherence theorem can give a new insight into the proof of the corresponding non-skew theorem. In my case, I realized we could heavily build on the proof of Beylin and Dybjer [3] of coherence for monoidal categories. It literally felt that all of my proof was already present in theirs, only the theorem was missing!

My next goal is to formulate and prove a similar coherence theorem for skew-closed categories of Street [15], a skew version of Eilenberg and Kelly's (non-monoidal) closed categories [7]. Like in this paper, I aim at a proof formalized in Agda. Coherence proofs tend to have a conceptually interesting high-level structure, but underneath they involve many tedious uninspiring case distinctions; it is more than easy to make mistakes.

Acknowledgements This research was supported by the ERDF financed CoE project EXCS, the Estonian Science Foundation grant no. 9475 and the Estonian Ministry of Education and Research target-financed research theme no. 0140007s12.

References

[1] T. Altenkirch, J. Chapman & T. Uustalu (2010): *Monads need not be endofunctors*. In L. Ong, ed.: *Proc. of 13th Int. Conf. on Foundations of Software Science and Computation Structures, FoSSaCS 2010 (Paphos, March 2010)*, *Lect. Notes in Comput. Sci.* 6014, Springer, 2010, pp. 297–311. doi: 10.1007/978-3-642-12032-9_21

[2] T. Beke (2011): *Categorification, term rewriting and the Knuth-Bendix procedure*. *J. of Pure and Appl. Alg.* 215(5), pp. 728–740. doi: 10.1016/j.jpaa.2010.06.019

[3] I. Beylin & P. Dybjer (1996): *Extracting a proof of coherence for monoidal categories from a proof of normalization for monoids*. In S. Berardi & M. Coppo, eds.: *Selected Papers from Int. Wksh. on Types for Proofs and Programs, TYPES '95 (Torino, June 1995)*, *Lect. Notes in Comput. Sci.* 1158, Springer, Berlin, pp. 47–61. doi: 10.1007/3-540-61780-9_61

[4] R. F. Blute, J. R. G. B. Cockett & R. A. G. Seely (1997): *Categories for computation in context and unified logic*. *J. of Pure and Appl. Alg.* 116(1–3), pp. 49–98. doi: 10.1016/s0022-4049(96)00162-4

[5] M. Buckley, R. Garner, S. Lack & R. Street (2013): *Skew-monoidal categories and the Catalan simplicial set*. arXiv preprint arXiv:1307.0265. Available online at `http://arxiv.org/abs/1307.0265`

[6] A. Burroni (1971): *T-catégories (catégories dans un triple)*. *Cahiers de topologie et géométrie différentielle catégoriques* 12, pp. 215–321.

[7] S. Eilenberg & M. Kelly (1966): *Closed categories*. In S. Eilenberg, D. K. Harrison, S. Mac Lane & H. Röhrl, eds.: *Proc. of Conf. on Categorical Algebra (La Jolla, 1965)*, *Lect. Notes in Math.* , pp. 421–562. doi: 10.1007/978-3-642-99902-4_22

[8] M. Grandis (2006): *Lax 2-categories and directed homotopy*. Cahiers de topologie et géométrie différentielle catégoriques 47(2), pp. 107–128.

[9] G. M. Kelly (1964): *On MacLane's conditions for coherence of natural associativities, commutativities, etc.* J. of Alg. 1(4), pp. 397–402. doi: 10.1016/0021-8693(64)90018-3

[10] S. Lack & R. Street (2012): *Skew monoidales, skew warpings and quantum categories*. Theory and Appl. of Categ. 26, pp. 385–402. Available at `http://www.tac.mta.ca/tac/volumes/26/15/26-15abs.html`

[11] S. Lack & R. Street (2013): *Triangulations, orientals, and skew monoidal categories*. arXiv preprint arXiv:1302.4488. Available at `http://arxiv.org/abs/1302.4488`

[12] M. L. Laplaza (1972): *Coherence for associativity not an isomorphism*. J. of Pure and Appl. Alg. 2, pp. 107–120. doi: 10.1016/0022-4049(72)90016-3

[13] M. L. Laplaza (1977): *Coherence in nonmonoidal closed categories*. Trans. of Amer. Math. Soc. 230, pp. 293–311. doi: 10.1090/s0002-9947-1977-0444740-9

[14] S. Mac Lane (1963): *Natural associativity and commutativity*. Rice Univ. Stud. 49, pp. 28–46.

[15] R. Street (2013): *Skew-closed categories*. J. of Pure and Appl. Alg. 217(6), pp. 973–988. doi: 10.1016/j.jpaa.2012.09.020

[16] K. Szlachányi (2012): *Skew-monoidal categories and bialgebroids*. Adv. in Math. 231(3–4), pp. 1694–1730. doi: 10.1016/j.aim.2012.06.027