

Data structures for topologically sound higher-dimensional diagram rewriting

Amar Hadzihasanovic

Tallinn University of Technology
& Quantinuum, Compositional Intelligence
amar@cs.ioc.ee

Diana Kessler

Tallinn University of Technology
diana-maria.kessler@taltech.ee

We present a computational implementation of diagrammatic sets, a model of higher-dimensional diagram rewriting that is “topologically sound”: diagrams admit a functorial interpretation as homotopies in cell complexes. This has potential applications both in the formalisation of higher algebra and category theory and in computational algebraic topology. We describe data structures for well-formed shapes of diagrams of arbitrary dimensions and provide a solution to their isomorphism problem in time $O(n^3 \log n)$. On top of this, we define a type theory for rewriting in diagrammatic sets and provide a semantic characterisation of its syntactic category. All data structures and algorithms are implemented in the Python library `rewalt`, which also supports various visualisations of diagrams.

Introduction

This article concerns the computational implementation of higher-dimensional diagrams in the sense of higher category theory, and contains some first steps in the computational complexity theory of diagrammatic rewriting in arbitrary dimensions.

Higher-dimensional rewriting, as emergent from the theory of polygraphs [5] – see [12] for a survey – is founded on an interpretation of *rewrites as directed homotopies*. A particular aim of our work is provable *topological soundness*, namely, the existence of a functorial interpretation of rewrite systems as cell complexes, and of rewrites as homotopies. This ensures that our implementation of higher-dimensional rewriting can act as a formal system for homotopical algebra and higher category theory in all generality.

With this aim, we turn to the *diagrammatic set* model [13] developed by the first author as a combinatorial alternative to polygraphs. Diagrammatic sets have a dual nature as higher-dimensional rewrite systems and “combinatorial directed cell complexes”. They support a model of weak higher categories and, unlike polygraphs, are topologically sound.

Beside the formalisation of higher algebra and category theory, potential applications are manifold. *String diagram rewriting*, which is a form of 3-dimensional rewriting, is arguably the characteristic computational mechanism of applied category theory. It has been suggested [4] that even “classical” forms of rewriting are more faithfully represented as diagram rewriting: for example, term rewriting implemented as rewriting in monoidal categories with cartesian structure explicates the “hidden costs” of copying and deleting terms. In these contexts, it is important to have a grasp on the computational complexity of the basic operations of diagram rewriting, to ensure that one’s cost model for a machine operating by diagram rewriting is reasonable.

Via topological soundness, we also envisage applications to computational algebraic topology. Directness of cells gives an *algebraic* grip on their pasting, which lends itself better to computation. Directed cell complexes are also equipped with an orientation on their cells, which makes them naturally suited to the computation of cellular homology.

Structure of the paper. In Section 1, we present some basic data structures from the theory of diagrammatic sets, together with their formal encoding: in particular, *oriented graded posets* which are used to encode shapes of diagrams.

In Section 2, we focus on the implementation of *regular molecules*, the inductive subclass of oriented graded posets corresponding to well-formed shapes of diagrams. To construct regular molecules, we need to decide their isomorphism problem; for general oriented graded posets, this is equivalent to the graph isomorphism problem (Proposition 2.11), not known to be in P. Our main result is a solution to the isomorphism problem for regular molecules in time $O(n^3 \log n)$ (Theorem 2.18), which also gives us a canonical form, hence a unique representation of shapes of diagrams.

In Section 3, we move on to the formalisation of diagrams and diagrammatic sets. We present this in the form of a type theory DiagSet living “on top” of our implementation of shapes of diagrams: the terms, corresponding to diagrams, are “filtered by regular molecules”. This allows us to define formal semantics and give a semantic characterisation of our formal system (Theorem 3.10).

Related work. A number of type theories for higher-categorical structures of arbitrary dimension have been defined in recent years: most notably, Finster and Mimram’s CaTT [8], implementing the Maltsiniotis model of weak higher categories [3], together with its “strictly associative” [10] and “strictly unital” [9] variants; and the *opetopic* type theories by Ho Thanh, Curien, and Mimram [15, 6].

The former are not particularly concerned with diagram rewriting, and focus instead on the implementation of coherent globular composition; the link to our work is tenuous. The latter have some commonality, albeit with a focus on a more restrictive class of shapes. In fact, DiagSet takes some inspiration not from one of the published opetopic type theories, but from a privately communicated variant due to Curien, which similarly rests on a “black-boxed” implementation of opetopic shapes.

Most closely related is the work by Vicary, Bar, Dorn, and others on quasistrict [2] and later associative [7, 17] n -categories, serving as the foundation of the homotopy.io proof assistant. While the aim is nearly the same, we believe that our framework has a number of advantages over associative n -categories.

From a theoretical perspective, it is only conjectural that associative n -categories, in general, are topologically sound or satisfy the homotopy hypothesis. They also currently lack connections with other models of higher categories and a clear functorial viewpoint. On the other hand, diagrammatic sets are topologically sound, satisfy a version of the homotopy hypothesis, and support a model of weak higher categories with concrete functorial ties to well-established models.

From a user perspective, the main point of divergence is that diagrams in associative n -categories have “strict units” but “weak interchange”, while our diagrams have “strict interchange” but need weak units to model “nullary” inputs or outputs. For rewrite systems with many “nullary” generators, associative n -categories may have a practical advantage, while diagrammatic sets are otherwise favoured.

Finally, in associative n -categories, diagram shapes are essentially descriptions of cubical tilings, and by lack of strict interchange, each rewrite gets by default its own “layer” in the tiling. This makes it so a “local” rewrite on a portion of a diagram leads to an inefficient “global” duplication of information. Our “face poset” representation of diagrams, on the other hand, allows local rewrites to stay local, which is more efficient and will be beneficial to the parallelisability of diagram rewriting.

Implementation. All data structures, algorithms, and systems discussed in this article were implemented by the authors as part of a Python library for higher-dimensional rewriting and algebra, called `rewalt`.¹ An example of `rewalt` code is included in Example 3.13. The library also supports various kinds of visualisation for diagrams, optionally in the form of TikZ output. All the Hasse and string diagrams in this article were generated by `rewalt` and included here with no subsequent retouching.

¹Project page: <https://github.com/ahadziha/rewalt>. Documentation: <https://rewalt.readthedocs.io>.

1 Basic data structures

1.1. In the theory of diagrammatic sets, the shape of a pasting diagram is encoded by its *face poset*, recording whether a cell is located in the boundary of another cell, together with *orientation* data which specifies whether an $(n-1)$ -dimensional cell is in the *input* or *output* half of the boundary of an n -dimensional cell. We call the mathematical structure containing these data an *oriented graded poset*. This is essentially the same as what Steiner calls a *directed precomplex* [18] and Forest an ω -*hypergraph* [11].

1.2 (Graded poset). Let P be a finite poset with order relation \leq and let P_\perp be P extended with a least element \perp . We say that P is *graded* if, for all $x \in P$, all directed paths from x to \perp in the Hasse diagram $\mathcal{H}P_\perp$, with edges going from covering to covered elements, have the same length. If this length is $n+1$, we let $\dim(x) := n$ be the *dimension* of x . We write P_n for the subset of n -dimensional elements of P .

1.3 (Oriented graded poset). An *orientation* on a finite poset P is an edge-labelling of its Hasse diagram with values in $\{+, -\}$. An *oriented graded poset* is a finite graded poset with an orientation.

Implementation 1.4. If we linearly order the elements of an oriented graded poset in each dimension, each element x is uniquely identified by a pair of integers (n, k) , where n is the *dimension* of x , and k is the *position* of x in the linear ordering of n -dimensional elements.

We then represent an oriented graded poset as a pair $(\text{face_data}, \text{coface_data})$ of *arrays of arrays of pairs of sets of integers*, where

1. $j \in \text{face_data}[n][k][i]$ if and only if $(n-1, j)$ is covered by (n, k) , and
2. $j \in \text{coface_data}[n][k][i]$ if and only if $(n+1, j)$ covers (n, k)

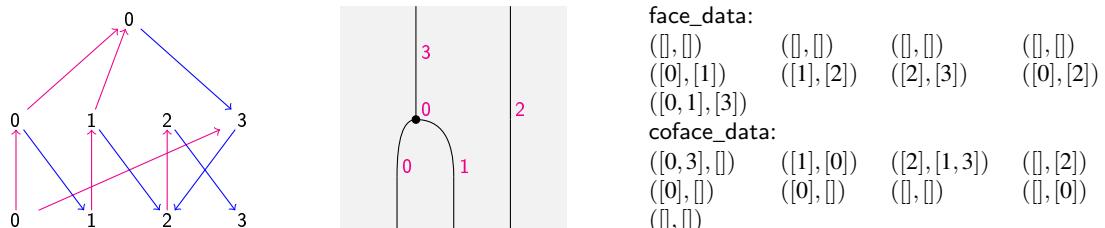
with orientation $-$ ($i = 0$) or $+$ ($i = 1$). We may implement the sets of integers as sorted arrays, or another data type which supports binary search in logarithmic time. This defines a data type *OgPoset*.

This representation is essentially an adjacency list representation of the poset's Hasse diagram, with vertices separated according to their dimension, and incoming and outgoing edges separated according to their label. If E_P is the set of edges of the Hasse diagram of P , the *OgPoset* representation of P takes space $O(|P| + |E_P|)$.

Storing both *face_data* and *coface_data* is redundant since these are uniquely determined by each other. However, most of the computations we need to perform on oriented graded posets require regular access both to faces (covered elements) and cofaces (covering elements) of a given element, so it is advantageous to be able to access them in constant time.

Example 1.5. Consider a diagram formed of one 2-cell with two input 1-cells and a single output 1-cell, whiskered to the right with a single 1-cell. The following are representations of its shape as

- an oriented face poset, pictured as a Hasse diagram with input faces pointing upwards (in magenta) and output faces downwards (in blue);
- a string diagram (0-cells are unlabelled, but correspond to bounded regions of the plane);
- the pair of *face_data* and *coface_data* (rows are outer array indices and columns inner array indices).



Remark 1.6. The representation of an oriented graded poset (up to isomorphism) is not unique: any permutation of the linear order on elements in each dimension leads to an equivalent representation.

1.7. Many important computations are performed on (*downwards*) *closed subsets*, rather than the whole of an oriented graded poset. In particular, the structure of an oriented graded poset supports a purely combinatorial definition of the input and output boundary of a closed subset.

1.8 (Closed subsets). Let P be an oriented graded poset and $U \subseteq P$. We say that U is *closed* if, for all $y \in U$ and $x \in P$, if $x \leq y$ then $x \in U$. The *closure* of U is the subset $\text{cl}U := \{x \in P \mid \exists y \in U \ x \leq y\}$.

We let $\dim(U)$ be the maximum of $\dim(x)$ for $x \in U$, or -1 if U is empty.

1.9 (Input and output boundaries). Let P be an oriented graded poset and $U \subseteq P$ a closed subset. For all $\alpha \in \{+, -\}$ and $n \in \mathbb{N}$, let

- $\Delta_n^\alpha U \subseteq U$ be the subset of elements x such that $\dim(x) = n$ and, if $y \in U$ covers x , then it covers it with orientation α ;
- $\mathcal{M}_n U \subseteq U$ be the subset of elements x such that $\dim(x) = n$ and x is maximal in U (not covered by any other element of U).

The *input* ($\alpha := -$) or *output* ($\alpha := +$) n -*boundary* of U is the closed subset

$$\partial_n^\alpha U := \text{cl}\left(\Delta_n^\alpha U \cup \bigcup_{k < n} \mathcal{M}_k U\right).$$

We let $\partial_n U := \partial_n^+ U \cup \partial_n^- U$ and omit n when $n = \dim(U) - 1$. For all $x \in P$, we let $\partial_n^\alpha x := \partial_n^\alpha \text{cl}\{x\}$.

Remark 1.10. It is convenient to also let $\partial_{-1}^\alpha U = \partial_{-2}^\alpha U := \emptyset$, so that $\partial^\alpha U$ is defined for all $U \subseteq P$.

Example 1.11. Let U be the oriented face poset of Example 1.5. Then

$$\begin{aligned} \partial_1^- U &= \{(0,0), (0,1), (0,2), (0,3), (1,0), (1,1), (1,2)\}, \\ \partial_1^+ U &= \{(0,0), (0,2), (0,3), (1,2), (1,3)\}, \\ \partial_0^- U &= \{(0,0)\}, \quad \partial_0^+ U = \{(0,3)\}. \end{aligned}$$

Implementation 1.12. We represent a set of elements of an OgPoset as an *array of sets of positions, indexed by dimensions*. This allows us to access the subset of elements of a given dimension in constant time. The size of arrays can be fixed to be equal to the dimension of a specific OgPoset, or dynamically adjusted to the dimension of each set of elements. Sets of positions can again be implemented as sorted arrays. This defines a data type **GrSet** (for *graded set*).

1.13 (Map of oriented graded posets). A *map* $f: P \rightarrow Q$ of oriented graded posets is a function of their underlying sets that satisfies $\partial_n^\alpha f(x) = f(\partial_n^\alpha x)$ for all $x \in P$, $n \in \mathbb{N}$, and $\alpha \in \{+, -\}$. We call an injective map an *inclusion*. Oriented graded posets and their maps form a category **ogPos**.

Example 1.14. A closed subset of an oriented graded poset inherits the structure of an oriented graded poset by restriction. Its subset inclusion is an inclusion of oriented graded posets.

Implementation 1.15. We represent a map $f: P \rightarrow Q$ as an *array of arrays of pairs of integers mapping*, together with pointers *source*, *target* to OgPoset representations of P and Q . This defines a data type **OgMap**. As an array of arrays, mapping has the same size of P 's *face_data*, and is defined by

$$\text{mapping}[n][k] = (m, j) \text{ if and only if } f((n, k)) = (m, j).$$

This representation takes space $O(|P|)$.

2 Unique representation of shapes of diagrams

2.1. In the theory of diagrammatic sets, shapes of diagrams form an inductively generated class of oriented graded posets, called regular *molecules* after Steiner [18].

2.2 (Round subset). Let U be a closed subset of an oriented graded poset, $n := \dim(U)$. We say that U is *round* if, for all $k < n$,

$$\partial_k^+ U \cap \partial_k^- U = \partial_{k-1} U.$$

Remark 2.3. Roundness is called “spherical boundary” in [13].

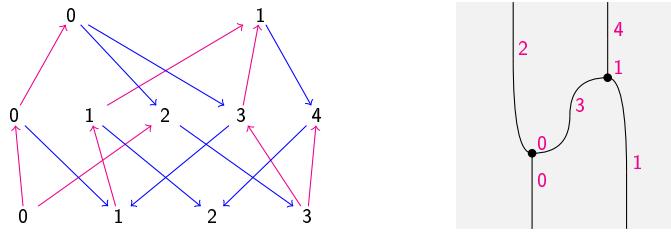
Example 2.4. Shapes of 2-dimensional diagrams, as oriented face posets, are round precisely when

1. their string diagram representation is *connected*, and
2. all nodes of the string diagram have at least one input and one output wire.

For example, the oriented graded poset of Example 1.5 is not round: we have

$$\partial_0 U = \{(0,0), (0,3)\} \subsetneq \partial_1^+ U \cap \partial_1^- U = \{(0,0), (0,2), (0,3)\}.$$

On the other hand, the following oriented graded poset is round:



2.5 (Regular molecules). The class of regular molecules is generated by the following clauses.

- (Point). The terminal oriented graded poset \bullet is a regular molecule.
- (Atom). Let U, V be *round* regular molecules such that $\dim(U) = \dim(V)$ and, for all $\alpha \in \{+, -\}$, $\partial^\alpha U$ is isomorphic to $\partial^\alpha V$. Then $U \Rightarrow V$ is a regular molecule, where $U \Rightarrow V$ is the essentially unique oriented graded poset $U \Rightarrow V$ with the property that
 1. $U \Rightarrow V$ has a greatest element, and
 2. $\partial^-(U \Rightarrow V)$ is isomorphic to U , while $\partial^+(U \Rightarrow V)$ is isomorphic to V .
- (Paste). Let U, V be regular molecules and $k < \min(\dim(U), \dim(V))$, such that $\partial_k^+ U$ is isomorphic to $\partial_k^- V$. Then the pushout $U \#_k V$ of the span $\partial_k^+ U \hookrightarrow U, \partial_k^+ U \xrightarrow{\sim} \partial_k^- V \hookrightarrow V$ is a regular molecule.

A regular molecule is an *atom* if it has a greatest element; these are precisely the molecules whose final generating clause is (Point) or (Atom).

The *submolecule* relation $U \sqsubseteq V$ is the preorder generated by $U, V \sqsubseteq U \Rightarrow V$ and $U, V \sqsubseteq U \#_k V$.

Comment 2.6. The properties of regular molecules are explored in [13, Sections 1, 2]. Importantly, the following results ensure that §2.5 is a valid definition:

1. the category **ogPos** has pushouts of inclusions;
2. if U and V are isomorphic regular molecules, they are isomorphic *in a unique way*;
3. input and output boundaries of regular molecules are regular molecules;

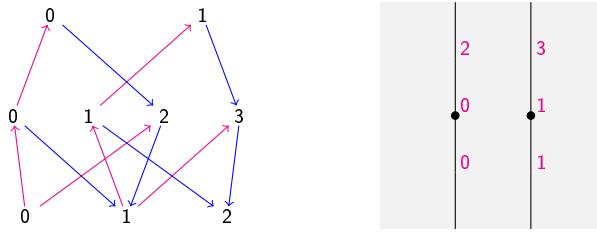
4. if U and V are round, then a pair of isomorphisms between $\partial^\alpha U$ and $\partial^\alpha V$ for $\alpha \in \{+, -\}$ extends uniquely to an isomorphism between ∂U and ∂V .

The first three imply that $U \#_k V$ is well-defined and does not depend on a choice of isomorphism between $\partial_k^+ U$ and $\partial_k^- V$. The fourth implies that $U \Rightarrow V$ can be uniquely constructed by extending the isomorphisms $\partial^\alpha U \xrightarrow{\sim} \partial^\alpha V$ to an isomorphism $\partial U \xrightarrow{\sim} \partial V$, then gluing U and V along this isomorphism, and finally adding a greatest element with the appropriate orientation.

Example 2.7. Let $\text{arrow} := (\bullet \Rightarrow \bullet)$ and $\text{binary} := ((\text{arrow} \#_0 \text{arrow}) \Rightarrow \text{arrow})$. The shape of the diagram of Example 1.5 is generated as $\text{binary} \#_0 \text{arrow}$, while the oriented graded poset of Example 2.4 is generated as $(\text{cobinary} \#_0 \text{arrow}) \#_1 (\text{arrow} \#_0 \text{binary})$, where $\text{cobinary} := (\text{arrow} \Rightarrow (\text{arrow} \#_0 \text{arrow}))$.

Remark 2.8. As discussed in [13, §2.1], the pasting constructions $- \#_k -$ satisfy the equations of composition in strict ω -categories *up to unique isomorphism*. It follows that the “same” regular molecule may be constructed in different ways. For example, letting $\text{globe} := (\text{arrow} \Rightarrow \text{arrow})$, we have

$$(\text{globe} \#_0 \text{arrow}) \#_1 (\text{arrow} \#_0 \text{globe}) \simeq \text{globe} \#_0 \text{globe} \simeq (\text{arrow} \#_0 \text{globe}) \#_1 (\text{globe} \#_0 \text{arrow}).$$



Implementation 2.9. We want to implement regular molecules as a subtype `Shape` of `OgPoset` with a nullary constructor `point` and partial binary constructors `atom`($-$, $-$) and `pastek`($-$, $-$) for $k \in \mathbb{N}$. In order to implement the constructors, we need to be able to perform the following operations:

1. compute input and output k -boundaries;
2. check if a closed subset is round;
3. determine if two regular molecules are isomorphic;
4. compute the pushout of a span of inclusions.

The first, second, and fourth of these admit straightforward algorithms of low-degree polynomial time complexity, that do not rely on any special properties of regular molecules. The third problem, however, is non-trivial. Indeed, the isomorphism problem generalised to all oriented graded posets is equivalent to the *graph isomorphism* (GI) problem, which is not known to be in P; the best known algorithm, due to Babai, runs in quasipolynomial time [1].

Remark 2.10. As customary in this context, a graph is a *simple* graph (no loops or multiple edges).

Proposition 2.11 — *The isomorphism problem for oriented graded posets is GI-complete.*

Proof. Deciding isomorphism of oriented graded posets is equivalent to deciding isomorphism of their Hasse diagrams with $\{+, -\}$ -labelled edges. The isomorphism problem for edge-labelled finite graphs is an instance of the isomorphism problem for finite relational structures, which is GI-complete [16].

Conversely, a directed graph can be represented by its “oriented incidence poset”: the 0-dimensional elements are the vertices, the 1-dimensional elements are the edges, the only input face of an edge is its source, and the only output face of an edge is its target. Two directed graphs are isomorphic if and only if their oriented incidence posets are isomorphic. Since GI reduces to the isomorphism problem for directed graphs, it reduces to the isomorphism problem for 1-dimensional oriented graded posets. ■

Nevertheless, in the special case of regular molecules, we can do much better. Our strategy is to describe a deterministic *traversal algorithm*, where the traversal order depends only on the intrinsic structure of a regular molecule as an oriented graded poset and not on its representation.

Given $U, V : \text{OgPoset}$ representing regular molecules, we traverse both U and V , and then *reorder* their elements in each dimension according to their traversal order. If $U', V' : \text{OgPoset}$ are the reordered versions of U, V , we then have

$$U \simeq V \text{ if and only if } U' \equiv V'.$$

We will show that, with this strategy, we can solve the isomorphism problem for regular molecules in time $O(n^3 \log n)$. A more precise upper bound is given in Theorem 2.18 below.

In addition to solving the isomorphism problem for regular molecules, the traversal order gives us a canonical form for regular molecules in OgPoset form. If we implement the constructors of Shape in such a way that they always produce an OgPoset in traversal order, we obtain that

$$\text{for all } U, V : \text{Shape}, U \simeq V \text{ if and only if } U \equiv V,$$

that is, we have a *unique representation* for shapes of diagrams.

```

procedure TRAVERSE( $U : \text{regular molecule}$ )
  marked  $\leftarrow []$ 
  stack  $\leftarrow [U]$ 
  while stack is not empty do
    5:   focus  $\leftarrow$  top of stack
    dim  $\leftarrow \text{dim}(\text{focus})$ 
    if focus  $\subseteq$  marked then
      pop focus from top of stack
    else
    10:  if  $\partial^-$ focus  $\not\subseteq$  marked then
        push  $\partial^-$ focus to top of stack
    else
    15:  if focus =  $\text{cl}\{x\}$  for some  $x$  then
        append  $x$  to marked
        pop focus from top of stack
        if  $\partial^+$ focus  $\not\subseteq$  marked then
          push  $\partial^+$ focus on top of stack
    else
    20:   $y \leftarrow$  first item of dimension  $\text{dim} - 1$  in marked such that
         $y$  has an unmarked input coface in focus
         $x \leftarrow$  unique input coface of  $y$  in focus
        push  $\text{cl}\{x\}$  on top of stack
  return marked

```

At each iteration of the main loop (line 4), the current state is fully described by the *stack* – including its top element, the *focus* – and by the list of *marked* elements.

Theorem 2.12 — *The traversal algorithm is correct: given a regular molecule U , it terminates returning a unique linear ordering of the elements of U .*

The proof of Theorem 2.12 uses the following preliminary facts.

Lemma 2.13 — *Let V be an item on the stack. Then V is a regular molecule. If W is below V on the stack, then V is a proper subset of W .*

Remark 2.14. In fact, any V that appears on the stack is either $\partial_k^- U$, which we call “ U -linked”, or it is $\text{cl}\{x\}$ or $\partial_k^\alpha x$, which we call “ x -linked”, for some $x \in U$. In the latter case, V is *round*, which implies that it is also *pure* [13, Lemma 1.35]: its maximal elements all have the same dimension.

Lemma 2.15 — Any subset V of U can be pushed onto the stack at most once.

Lemma 2.16 — Let V be the focus, $n := \dim(V)$. Then either V is fully marked, or there exists an n -dimensional element of V which is unmarked.

2.17. In what follows, for a fixed regular molecule U , we let $|E_n|$ be the number of edges between n and $(n-1)$ -dimensional elements in the Hasse diagram of U , and we let

$$|U_{\max}| := \max_n |U_n|, \quad |E_{\max}| := \max_n |E_n|.$$

Theorem 2.18 — The traversal algorithm admits an implementation running in time

$$O(|U|^2(|E_{\max}| \cdot \log |E_{\max}| + |U_{\max}| \cdot \log |U_{\max}|)).$$

Proof. First of all, we represent any closed set on the stack with its graded set of maximal elements. To initialise the algorithm, we only need to compute the maximal elements of U . This can be done in time $O(|U|)$ by going through the elements of U and checking if their set of cofaces is empty.

Next, let us find an upper bound for the number of iterations of the main loop (line 4). Let V be a set on the stack, $n := \dim(V)$. Then V can become the focus

- at most once before pushing $\partial^- V$ onto the stack (line 11),
- at most once before pushing $\text{cl}\{y\}$ onto the stack for each $y \in V_n$ (line 22), and
- at most once to be popped from the stack (line 8),

after which, by Lemma 2.15, it can never appear again. Thus, the number of loop iterations with V as focus is bounded by $|V_n| + 2$.

By Remark 2.14, every set V on the stack is either “ U -linked” or “ x -linked” for some $x \in U$. There are $(\dim(U) + 1)$ many U -linked focusses and $(2\dim(x) + 1)$ many x -linked focusses. Then

- the number of loop iterations with U -linked focusses is bounded by $|U| + 2\dim(U) + 2$, and
- for each x , the number of loop iterations with x -linked focusses is bounded by $|\text{cl}\{x\}| + 4\dim(x) + 2$.

Since there are $|U|$ elements, $|\text{cl}\{x\}| \leq |U|$, and $\dim(x) \leq \dim(U)$, we have a coarse upper bound of $(|U| + 1)(|U| + 4\dim(U) + 2)$ on the total number of iterations, which is $O(|U|^2)$.

Next, in our implementation, we split the list of marked elements into three objects: a list order (for the total traversal order), an array of lists grorder (for the traversal order split by dimension), and a graded set marked (for the set of marked elements).

Consider a single loop iteration with focus V , $n := \dim(V)$.

(Line 7). By Lemma 2.16, to check if V is fully marked, it suffices to check whether $V_n \subseteq \text{marked}_n$. Since both are sorted arrays of integers, they can be compared in time linear in $|V_n| + |\text{marked}_n|$, which is $O(|U_n|)$. At this stage, we may also record the unmarked n -dimensional elements of V in a sorted array unmarked without affecting the complexity.

(Line 10). To compute the maximal elements of $\partial^- V$ and $\partial^+ V$, we may use different strategies depending on whether V is “ U -linked” or not.

If $V = \partial_n^- U$, we compute the $(n-1)$ -dimensional elements of $\partial^- V = \partial_{n-1}^- U$ simply by going through the elements of U_{n-1} and checking which ones have empty sets of output cofaces, in time $O(|U_{n-1}|)$. Lower-dimensional maximal elements are shared between V and $\partial^- V$, so we may then point from the latter to the former, at no extra cost.

If V is not U -linked, V and its boundaries are pure, so the set of maximal elements of $\partial^\alpha V$ is equal to $\Delta^\alpha V$, and each of its elements is covered by an element of V_n . To compute it, we add all the input and output faces of all $x \in V_n$ to sets `in_faces` and `out_faces`, respectively, then use the relations $\Delta^- V = \text{in_faces} \setminus \text{out_faces}$ and $\Delta^+ V = \text{out_faces} \setminus \text{in_faces}$.

There are $O(|E_n|)$ faces of elements of V_n , and we can sort `in_faces` and `out_faces`, remove duplicates, and compute their difference in time $O(|E_n| \cdot \log |E_n|)$.

At this stage, we also create an associative array `candidates` as follows: whenever $x \in V_n$ is in unmarked, and y is an input face of x , we add the position of x as a value to `candidates`, indexed by the position of y . We then sort the indices of `candidates`. This also takes time $O(|E_n| \cdot \log |E_n|)$ so it does not affect the overall complexity.

(Lines 10, 16). By the same reasoning applied to line 7, checking if $\partial^- V$ and $\partial^+ V$ are fully marked takes time $O(|U_{n-1}|)$.

(Line 14). If V_n has a single element that we mark, adding it to `order` and `grorder` takes constant time with an appropriate implementation of lists. Adding it to `marked` takes $O(|U_n|)$.

(Lines 19—21). To select the next focus we traverse `grordern-1` starting from the first item and search for each item in the indices of `candidates` until we find a hit y . This takes time $O(|U_{n-1}| \cdot \log |U_{n-1}|)$ in the worst case. The next focus will be $\text{cl}\{x\}$, where x is the value corresponding to index y .

Overall, the worst-case complexity is $O(|U_n| + |E_n| \cdot \log |E_n| + |U_{n-1}| \cdot \log |U_{n-1}|)$. Using the bounds $|U_n|, |U_{n-1}| \leq |U_{\max}|$ and $|E_n| \leq |E_{\max}|$, and multiplying by our bound on the number of iterations, we conclude. ■

3 A type theory for higher-dimensional rewriting

3.1. We rapidly go through the definitions of diagrammatic sets and some related notions. For a thorough treatment, we refer to [13, Section 4 and onwards], and to [14, Section V] for diagrammatic complexes as presentations of higher-dimensional theories.

3.2 (Diagrammatic set). Let \odot (to be read *atom*) be a skeleton of the full subcategory of **ogPos** on the atoms of every dimension. A *diagrammatic set* is a presheaf on \odot . Diagrammatic sets and their morphisms of presheaves form a category $\odot\text{Set}$.

3.3. We identify \odot with a full subcategory $\odot \hookrightarrow \odot\text{Set}$ via the Yoneda embedding. With this identification, we use morphisms in $\odot\text{Set}$ as our notation for both elements and structural operations of a diagrammatic set X :

- $x \in X(U)$ becomes $x: U \rightarrow X$, and
- for each map $f: V \rightarrow U$ in \odot , $X(f)(x) \in X(V)$ becomes $f;x: V \rightarrow X$.

The embedding $\odot \hookrightarrow \odot\text{Set}$ extends along pushouts of inclusions to the full subcategory of **ogPos** on the regular molecules.

3.4 (Diagrams and cells). Let X be a diagrammatic set and U a regular molecule. A *diagram of shape U in X* is a morphism $x: U \rightarrow X$. A diagram is a *cell* if U is an atom. For all $n \in \mathbb{N}$, we say that x is an *n-diagram* or an *n-cell* when $\dim(U) = n$.

If U decomposes as $U_1 \#_k U_2$, we write $x = x_1 \#_k x_2$ for $x_i := \iota_i; x$, where ι_i is the inclusion $U_i \hookrightarrow U$ for $i \in \{1, 2\}$. Let $\iota_k^\alpha: \partial_k^\alpha U \hookrightarrow U$ be the inclusions of the k -boundaries of U . The *input k-boundary* of x is the diagram $\partial_k^- x := \iota_k^-; x$ and the *output k-boundary* of x is the diagram $\partial_k^+ x := \iota_k^+; x$. We write $x: y^- \Rightarrow y^+$ to express that $\partial_k^\alpha x = y^\alpha$ for each $\alpha \in \{+, -\}$.

3.5 (Diagrammatic complex). For each $n \in \mathbb{N}$, let \odot_n be the full subcategory of \odot on the atoms of dimension $\leq n$, and let \odot_{-1} be the empty subcategory. The restriction functor $\odot\mathbf{Set} \rightarrow \mathbf{PSh}(\odot_n)$ has a left adjoint; let $\sigma_{\leq n}$ be the comonad induced by this adjunction. The n -skeleton of a diagrammatic set X is the counit $\sigma_{\leq n}X \rightarrow X$. For all $k \leq n$, the k -skeleton factors uniquely through the n -skeleton of X .

A *diagrammatic complex* is a diagrammatic set X together with a set $\mathcal{X} = \sum_{n \in \mathbb{N}} \mathcal{X}_n$ of *generating cells* such that, for all $n \in \mathbb{N}$,

$$\begin{array}{ccc} \bigsqcup_{x \in \mathcal{X}_n} \partial U(x) & \longrightarrow & \bigsqcup_{x \in \mathcal{X}_n} U(x) \\ \downarrow (\partial x)_{x \in \mathcal{X}_n} & & \downarrow (x)_{x \in \mathcal{X}_n} \\ \sigma_{\leq n-1}X & \longrightarrow & \sigma_{\leq n}X \end{array}$$

is a pushout in $\odot\mathbf{Set}$, where $U(x)$ denotes the shape of x . A diagrammatic complex is *finite* if \mathcal{X} is finite.

3.6 (Support-based diagrammatic complex). Each cell in a diagrammatic complex (X, \mathcal{X}) is uniquely of the form $(p: U \twoheadrightarrow V, x: V \rightarrow X)$, where p is a surjective map of atoms and $x \in \mathcal{X}$. We let $\text{supp}(p, x) := x$, the *support* of (p, x) .

A *support-based diagrammatic complex* is the quotient of a diagrammatic complex by the relations

$$x \sim y \text{ if and only if } \text{supp}(\iota; x) = \text{supp}(\iota; y) \quad (1)$$

for all cells $x, y: U \rightarrow X$ and all inclusions of atoms $\iota: V \hookrightarrow U$. We let $\odot\mathbf{Cpx}_{fsb}$ denote the full subcategory of $\odot\mathbf{Set}$ on the finite, support-based diagrammatic complexes.

3.7. We define a dependent type theory for diagrammatic sets – more precisely, for finite, support-based diagrammatic complexes – that relies on an underlying unique representation of regular molecules and their maps, treated as a “black box”. Of course, in the previous section we have provided such an implementation and proved that it is computationally feasible. Nevertheless, it is useful to separate its abstract properties from the implementation details.

3.8 (DiagSet). Let \mathbb{V} be an infinite set of variables. We define a type theory DiagSet as follows.

Terms. A term t is a pair of a regular molecule U , the *shape* of t , and a function $t: U \rightarrow \mathbb{V}$. We write t/U to express that t is a term of shape U . Maps $p: U \rightarrow V$ act on terms by precomposition: if t/V is a term, then $p^*t := (p; t)/U$. In particular, we let $\partial_k^\alpha t := (\iota_k^\alpha; t)/\partial_k^\alpha V$ for all $k \in \mathbb{N}$ and $\alpha \in \{+, -\}$.

Types. A type A is either \emptyset or an expression $t \Rightarrow s$ where t, s are terms. We may annotate a term t of shape U with the type $A := \emptyset$ if $U \equiv \bullet$, and $A := \partial^- t \Rightarrow \partial^+ t$ otherwise.

Contexts. A context Γ is a list $x_1 : A_1, \dots, x_n : A_n$ of typed variables. We consider two contexts to be equal if they are equal up to a permutation. If $x : A$ is a typed variable, we say that x has *shape* \bullet if $A \equiv \emptyset$, and $U \Rightarrow V$ if $A \equiv t/U \Rightarrow s/V$. We write $x/U : A$ to express that $x : A$ has shape U .

Substitutions. A substitution σ is a list $x_1 \mapsto t_1, \dots, x_n \mapsto t_n$ of assignments of terms to variables. We consider two substitutions to be equal if they are equal up to a permutation.

Judgments. We consider three kinds of judgments:

- $\Gamma \vdash$ meaning that Γ is a well-formed context,
- $\Gamma \vdash t$ meaning that t is a well-formed term in context Γ , and
- $\Delta \vdash \sigma : \Gamma$ meaning that σ is a well-formed substitution from context Δ to context Γ .

The inference rules of DiagSet are the following. We use $\langle \rangle$ to indicate the empty list.

Rules for contexts.	
$\frac{}{\langle \rangle \vdash} \text{init}$	$\frac{\Gamma \vdash}{\Gamma, x : \emptyset \vdash} \text{pt}$
	$\frac{\Gamma \vdash t/U : r^- \Rightarrow r^+ \quad \Gamma \vdash s/V : r^- \Rightarrow r^+ \quad U, V \text{ round}}{\Gamma, x : t \Rightarrow s \vdash} \text{gen}$
(where $x \in \mathbb{V}$ is fresh)	
Rules for terms.	
$\frac{\Gamma \vdash \quad (x/V : A) \in \Gamma \quad U \text{ atom} \quad p : U \rightarrow V \text{ surjective}}{\Gamma \vdash p^* \hat{x}/U} \text{cell}$	
$\frac{\Gamma \vdash t/U \quad \Gamma \vdash s/V \quad \partial_k^+ t \equiv \partial_k^- s}{\Gamma \vdash (t \#_k s)/(U \#_k V)} \text{paste}_k, \quad k < \min(\dim(U), \dim(V))$	
Rules for substitutions.	
$\frac{\Gamma \vdash}{\Gamma \vdash \langle \rangle : \Gamma} \text{id}$	$\frac{\Delta \vdash \sigma : \Gamma \quad \Gamma, x : s/U \Rightarrow r/V \vdash \quad \Delta \vdash t/U \Rightarrow V : s[\sigma] \Rightarrow r[\sigma]}{\Delta \vdash \langle \sigma, x \mapsto t \rangle : (\Gamma, x : s \Rightarrow r)} \text{ext}$

In the rules cell and paste, the terms \hat{x} and $t \#_k s$ are defined as follows:

- \hat{x} is the unique term of shape V which sends the greatest element of V to x , and, if $A \equiv t \Rightarrow s$, is equal to t on $\partial^- V$ and to s on $\partial^+ V$;
- $t \#_k s$ is the unique term of shape $U \#_k V$ which is equal to t on $U \hookrightarrow (U \#_k V)$ and to s on $V \hookrightarrow (U \#_k V)$.

The side conditions for gen and paste ensure that this is well-defined.

To define the action $t[\sigma]$ of a well-formed substitution σ on a term t , we extend σ to a function $\mathbb{V} \rightarrow \mathbb{V}$ as follows: for all $x \in \mathbb{V}$, if $(x \mapsto t/U) \in \sigma$, we let $\sigma(x) := t(\top)$, where \top is the greatest element of U ; otherwise, $\sigma(x) := x$. Then $t[\sigma]$ is the composite of $t : U \rightarrow \mathbb{V}$ and $\sigma : \mathbb{V} \rightarrow \mathbb{V}$. Note that this is well-defined because a well-formed substitution assigns to each variable a term whose shape is an atom.

3.9 (Syntactic category). The syntactic category $\mathbf{Ctx}[\mathbf{DiagSet}]$ has

- well-formed contexts Γ as objects, and
- well-formed substitutions as morphisms from Δ to Γ ,

with the obvious composition of substitutions, and empty substitutions as identities.

Theorem 3.10 — *The category $\mathbf{Ctx}[\mathbf{DiagSet}]^{\text{op}}$ is equivalent to $\odot \mathbf{Cpx}_{\text{fsb}}$.*

Sketch of proof. We define an encoding enc of finite support-based diagrammatic complexes, diagrams, and morphisms as contexts, terms, and substitutions. Given (X, \mathcal{X}) , we pick an injective function $\text{name} : \mathcal{X} \rightarrow \mathbb{V}$, assigning unique variable names to the generating cells of X .

For all diagrams $d : U \rightarrow X$, we define a term $\text{enc}(d)$ as follows: for all $x \in U$, we let $\text{enc}(d)(x)$ be equal to $\text{name}(\text{supp}(d|_{\text{cl}\{x\}}))$. Since (X, \mathcal{X}) is support-based, $\text{enc}(d) \equiv \text{enc}(d')$ implies $d = d'$.

Let n be the greatest dimension in which \mathcal{X}_n is non-empty, and pick a linear ordering x_1, \dots, x_{m_k} of \mathcal{X}_k for all $k \leq n$. We let $\text{enc}(X, \mathcal{X}) := \Gamma_0, \dots, \Gamma_n$, where

$$\Gamma_k := \text{name}(x_1) : \text{enc}(\partial^- x_1) \Rightarrow \text{enc}(\partial^+ x_1), \dots, \text{name}(x_{m_k}) : \text{enc}(\partial^- x_{m_k}) \Rightarrow \text{enc}(\partial^+ x_{m_k}).$$

By the construction of X as a colimit of its generating cells, any map $X \rightarrow Y$ is uniquely determined by what it does on \mathcal{X} . Given a map $f : (X, \mathcal{X}) \rightarrow (Y, \mathcal{Y})$ in $\odot \mathbf{Cpx}_{\text{fsb}}$, we let $\text{enc}(f)$ be the substitution

$$\langle \text{name}_X(x) \mapsto \text{enc}_Y(f(x)) \rangle_{x \in \mathcal{X}}.$$

Conversely, we define an interpretation $\llbracket - \rrbracket$ of well-formed contexts, terms, and substitutions by induction on inference rules of DiagSet . At each step the interpretation $\llbracket \Gamma \rrbracket$ of a well-formed context is a support-based diagrammatic complex with one generator $\llbracket \hat{x} \rrbracket$ of shape U for each variable x/U in Γ .

- (init) The interpretation of the empty context is the initial diagrammatic set.
- (pt) Suppose $\llbracket \Gamma \rrbracket$ is defined. The interpretation of $\Gamma, x : \emptyset$ is the coproduct $\llbracket \Gamma \rrbracket + \bullet$. The interpretation of \hat{x} is the inclusion $\bullet \hookrightarrow \llbracket \Gamma \rrbracket + \bullet$.
- (gen) Suppose $\llbracket \Gamma \rrbracket$ and $\llbracket t/U \rrbracket, \llbracket s/V \rrbracket$ are defined. The interpretation of $\Gamma, x : t \Rightarrow s$ is the pushout of $\partial \llbracket \hat{x} \rrbracket : \partial(U \Rightarrow V) \rightarrow \llbracket \Gamma \rrbracket$ and $\partial(U \Rightarrow V) \hookrightarrow (U \Rightarrow V)$, quotiented by the equations (1), where $\partial \llbracket \hat{x} \rrbracket$ is equal to $\llbracket t \rrbracket$ on $\partial^-(U \Rightarrow V)$ and to $\llbracket s \rrbracket$ on $\partial^+(U \Rightarrow V)$.
- (cell) Suppose $\llbracket \Gamma \rrbracket$ is defined and has a generating cell $\llbracket \hat{x} \rrbracket$. The interpretation of $p^* \hat{x}$ is $p; \llbracket \hat{x} \rrbracket$.
- (paste_k) Suppose $\llbracket \Gamma \rrbracket$ and $\llbracket t \rrbracket, \llbracket s \rrbracket$ are defined with $\partial_k^+ \llbracket t \rrbracket = \llbracket \partial_k^+ t \rrbracket = \llbracket \partial_k^- s \rrbracket = \partial_k^- \llbracket s \rrbracket$. The interpretation of $t \#_k s$ is the diagram $\llbracket t \rrbracket \#_k \llbracket s \rrbracket$.
- (id) The interpretation of the empty substitution in context Γ is the identity of $\llbracket \Gamma \rrbracket$.
- (ext) Suppose $\llbracket \sigma \rrbracket$ and $\llbracket t \rrbracket$ are defined, where $\llbracket \hat{x} \rrbracket$ and $\llbracket t \rrbracket$ both have the same shape U . By the construction of $\llbracket \Gamma, x \rrbracket$ as a colimit of $\llbracket \Gamma \rrbracket$ and U , the pair of $\llbracket \sigma \rrbracket : \llbracket \Gamma \rrbracket \rightarrow \llbracket \Delta \rrbracket$ and $\llbracket t \rrbracket : U \rightarrow \llbracket \Gamma \rrbracket$ induces a unique morphism $\llbracket \sigma, x \mapsto t \rrbracket : \llbracket \Gamma, x \rrbracket \rightarrow \llbracket \Delta \rrbracket$.

It is routine to check that enc and $\llbracket - \rrbracket$ define contravariant functors between OCpx_{fsb} and $\text{Ctx}[\text{DiagSet}]$, and that they are each other's inverse up to natural isomorphism. ■

Remark 3.11. The proof of Theorem 3.10 gives a semantic characterisation of well-formed terms as diagrams in a diagrammatic set. An immediate consequence is that the following rule is admissible:

$$\frac{\Gamma \vdash t/V \quad p: U \rightarrow V \text{ map}}{\Gamma \vdash p^* t/U} \text{pb}$$

where p is an arbitrary map of regular molecules.

Comment 3.12. A sticking point in our type theory is the fact that cell is parametrised by an arbitrary surjective map of atoms p . This is necessary to access the “weak units” and degenerate cells which in our framework are needed, among other things, to model nullary operations in an algebraic theory.

In practice, however, this is the one point in which the underlying implementation of regular molecules and their maps has to be explicitly accessed in order to define p and its domain. To avoid this, in a practical implementation, we want to include explicitly some extra admissible rules, corresponding to the application of useful maps that are parametric in their codomain.

In particular, we want to explicitly include

- the trivial case $p \equiv \text{id}_U$:
$$\frac{\Gamma \vdash (x/U : A) \in \Gamma}{\Gamma \vdash \hat{x}/U} \text{cell}^{'},$$
- *unit* rules, modelling [13, §4.16]:
$$\frac{\Gamma \vdash t/U}{\Gamma \vdash \text{unit}(t) := \tau^*(t) : t \Rightarrow t} \text{unit},$$
- *left and right unitor rules*, modelling [*ibid.*, §4.17]:
$$\frac{\Gamma \vdash t/U \quad V \sqsubseteq \partial^- U \text{ round}}{\Gamma \vdash \text{lunitor}_V(t) := (\ell_{V \hookrightarrow U}^-)^* t} \text{lunitor}$$

$$\frac{\Gamma \vdash t/U \quad V \sqsubseteq \partial^+ U \text{ round}}{\Gamma \vdash \text{runitor}_V(t) := (r_{V \hookrightarrow U}^+)^* t} \text{runitor}$$

where V can be specified, for example, by the set of positions of its maximal elements.

We may also have extra rules for *simplex and cube degeneracy* maps and for *cube connection* maps, in the case where U is an oriented simplex or cube as in [*ibid.*, §3.33]. All of these are implemented as diagram methods in `rewalt`.

Example 3.13. As an example, we give a presentation in DiagSet of the theory of a left-unital binary operation, together with its implementation in `rewalt`. In the framework of diagrammatic sets, a many-sorted “monoidal theory” is presented by a diagrammatic complex with a single 0-cell; this is analogous to the way a monoidal category is a bicategory with a single 0-cell. The sorts are generating 1-cells, the basic operations are generating 2-cells, and “oriented equations” are generating 3-cells.

First, we add a single 0-cell x and a single sort a .

$$\begin{array}{c}
 \frac{}{\langle \rangle \vdash} \text{init} \\
 \frac{}{x : \emptyset \vdash} \text{pt} \\
 \frac{}{x : \emptyset \vdash \widehat{x}} \text{cell}
 \end{array}
 \quad
 \begin{array}{c}
 \frac{x : \emptyset \vdash \widehat{x} \quad x : \emptyset \vdash \widehat{x}}{x : \emptyset, a : \widehat{x} \Rightarrow \widehat{x} \vdash} \text{gen} \\
 \frac{x : \emptyset, a : \widehat{x} \Rightarrow \widehat{x} \vdash}{x : \emptyset, a : \widehat{x} \Rightarrow \widehat{x} \vdash \widehat{a}} \text{cell'}
 \end{array}
 \quad
 \begin{array}{l}
 1 \text{ import rewalt} \\
 2 \text{ Lun} = \text{rewalt.DiagSet}() \\
 3 \text{ x} = \text{Lun.add('x')} \\
 4 \text{ a} = \text{Lun.add('a', x, x)}
 \end{array}$$

Let $\Gamma := x : \emptyset, a : \widehat{x} \Rightarrow \widehat{x}$. We add a binary operation m .

$$\frac{\Gamma \vdash \widehat{a} \quad \Gamma \vdash \widehat{a}}{\Gamma \vdash \widehat{a} \#_0 \widehat{a}} \text{ paste}_0 \quad
 \frac{\Gamma \vdash \widehat{a}}{\Gamma, m : \widehat{a} \#_0 \widehat{a} \Rightarrow \widehat{a} \vdash} \text{gen} \quad
 \begin{array}{l}
 5 \text{ m} = \text{Lun.add('m', a.paste(a), a)} \\
 \hline
 \frac{\Gamma, m : \widehat{a} \#_0 \widehat{a} \Rightarrow \widehat{a} \vdash}{\Gamma, m : \widehat{a} \#_0 \widehat{a} \Rightarrow \widehat{a} \vdash \widehat{m}} \text{cell'}
 \end{array}$$

Let $\Gamma' := \Gamma, m : \widehat{a} \#_0 \widehat{a} \Rightarrow \widehat{a}$. We produce a weak unit on x and add a nullary operation u .

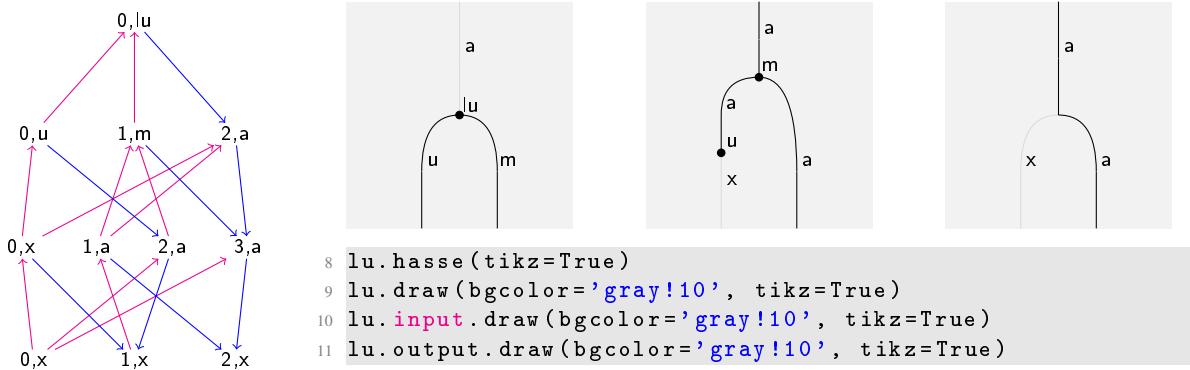
$$\frac{\Gamma' \vdash \widehat{x}}{\Gamma' \vdash \text{unit}(\widehat{x})} \text{unit} \quad
 \frac{\Gamma' \vdash \widehat{a}}{\Gamma', u : \text{unit}(\widehat{x}) \Rightarrow \widehat{a} \vdash} \text{gen} \quad
 \begin{array}{l}
 6 \text{ u} = \text{Lun.add('u', x.unit(), a)} \\
 \hline
 \frac{\Gamma', u : \text{unit}(\widehat{x}) \Rightarrow \widehat{a} \vdash}{\Gamma', u : \text{unit}(\widehat{x}) \Rightarrow \widehat{a} \vdash \widehat{u}} \text{cell'}
 \end{array}$$

Let $\Gamma'' := \Gamma', u : \text{unit}(\widehat{x}) \Rightarrow \widehat{a}$. We produce a left unit on a , and add an “oriented equation” exhibiting the fact that u is a left unit for m .

$$\frac{\Gamma'' \vdash \widehat{u} \quad \Gamma'' \vdash \widehat{a}}{\Gamma'' \vdash \widehat{u} \#_0 \widehat{a}} \text{ paste}_0 \quad
 \frac{\Gamma'' \vdash \widehat{m}}{\Gamma'' \vdash \widehat{m}} \text{ paste}_1 \quad
 \frac{\Gamma'' \vdash \widehat{a}/\text{arrow}}{\Gamma'' \vdash \text{lunitor}_{\partial-\text{arrow}}(\widehat{a})} \text{lunitor} \quad
 \frac{}{\Gamma'', lu : ((\widehat{u} \#_0 \widehat{a}) \#_1 \widehat{m}) \Rightarrow \text{lunitor}_{\partial-\text{arrow}}(\widehat{a}) \vdash} \text{gen}$$

$$7 \text{ lu} = \text{Lun.add('lu', u.paste(a).paste(m), a.lunitor())}$$

The following is a representation of lu as a term of DiagSet, that is, an oriented graded poset labelled with names, together with string diagram representations of lu , its input boundary, and its output boundary, and the `rewalt` code that generated them.



Comment 3.14. Provided we have a unique underlying representation of shapes, as described in Section 2, every term of DiagSet also has a unique representation. In this sense, terms of DiagSet are “noncomputational”: all the computation, which consists exclusively of computing and matching shapes, happens under the hood before a term is even created, so the equality theory of terms is trivial.

This is intended. Rather than a computational theory in itself, DiagSet is intended as a *substrate for computational theories* according to the paradigm of higher-dimensional rewriting. A term $t : r^- \Rightarrow r^+$ can be seen as a rewrite of the “lower-dimensional” term r^- to the term r^+ , and the extension of t via the paste_k rules establishes how the rewrite can happen in a wider context. In this sense, every well-formed context in DiagSet contains its own internal computational theory on terms of each dimension.

Remark 3.15. While “rewrites in context” can be built with the paste_k rules, this is quite impractical. In practice, one wants to start from a diagram and apply a generating rewrite directly to a subdiagram. This is modelled by *pasting along a subdiagram* [13, §4.12] in the theory of diagrammatic sets.

Pasting along a subdiagram is implemented in `rewalt` with methods `to_inputs` and `to_outputs`. These invoke a procedure for recognising subdiagrams, which currently uses a quite naive algorithm. The issue of recognising subdiagrams deserves further study, so we leave it to future work.

Conclusions and outlook

We have provided a formal implementation of “plain” diagrammatic sets. An obvious next step is the formalisation of *weakly invertible* cells, and then of diagrammatic sets with weak composites, a model of weak higher categories [13, Sections 5, 6]. This is in fact part of `rewalt`, but still lacks a formal analysis.

In addition, we still have a limited range of high-level methods for handling weak units. We may want, for example, flexible higher-dimensional versions of “Mac Lane triangle” rules for shuffling weak units around. Development of these methods, and others tailored to specific applications, will likely go hand in hand with practical experience in the use of `rewalt` as a proof assistant.

To conclude, we have only scratched the surface of the algorithm and complexity theory of diagram rewriting in higher dimensions. In particular, we have not yet studied the problem of searching for a subdiagram within another diagram, whose solution is essential to any form of fully automated or assisted diagram rewriting. We plan to tackle this problem in future work.

Acknowledgements. This work was supported by the ESF funded Estonian IT Academy research measure (project 2014-2020.4.05.19-0001) and by the Estonian Research Council grant PSG764.

References

- [1] L. Babai (2016): *Graph isomorphism in quasipolynomial time [extended abstract]*. In: *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, ACM, pp. 684–697, doi:10.1145/2897518.2897542. Available at <https://doi.org/10.1145/2897518.2897542>.
- [2] K. Bar & J. Vicary (2017): *Data structures for quasistrict higher categories*. In: *2017 32nd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, IEEE, doi:10.1109/lics.2017.8005147. Available at <https://doi.org/10.1109/lics.2017.8005147>.
- [3] T. Benjamin, E. Finster & S. Mimram (2021): *Globular weak ω -categories as models of a type theory*. *arXiv preprint arXiv:2106.04475*.
- [4] G. Bonfante & Y. Guiraud (2009): *Polygraphic programs and polynomial-time functions*. *Logical Methods in Computer Science* 5(2), doi:10.2168/lmcs-5(2:14)2009. Available at <https://doi.org/10.2168/lmcs-5%282%3A14%292009>.
- [5] A. Burroni (1993): *Higher-dimensional word problems with applications to equational logic*. *Theoretical Computer Science* 115(1), pp. 43–62, doi:10.1016/0304-3975(93)90054-w. Available at [https://doi.org/10.1016/0304-3975\(93\)90054-w](https://doi.org/10.1016/0304-3975(93)90054-w).
- [6] P.L. Curien, C. Ho Thanh & S. Mimram (2019): *Syntactic approaches to opetopes*. *arXiv preprint arXiv:1903.05848*.
- [7] C. Dorn (2018): *Associative n -categories*. Ph.D. thesis, University of Oxford.
- [8] E. Finster & S. Mimram (2017): *A type-theoretical definition of weak ω -categories*. In: *2017 32nd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, IEEE, doi:10.1109/lics.2017.8005124. Available at <https://doi.org/10.1109/lics.2017.8005124>.
- [9] E. Finster, D. Reutter & J. Vicary (2020): *A type theory for strictly unital ∞ -categories*. *arXiv preprint arXiv:2007.08307*.
- [10] E. Finster, A. Rice & J. Vicary (2021): *A type theory for strictly associative infinity categories*. *arXiv preprint arXiv:2109.01513*.
- [11] S. Forest (2019): *Unifying notions of pasting diagrams*. *arXiv preprint arXiv:1903.00282*.
- [12] Y. Guiraud (2019): *Rewriting methods in higher algebra*. Thèse d’habilitation à diriger des recherches, Université Paris 7.
- [13] A. Hadzihasanovic (2020): *Diagrammatic sets and rewriting in weak higher categories*. *arXiv preprint arXiv:2007.14505*.
- [14] A. Hadzihasanovic (2021): *The smash product of monoidal theories*. In: *2021 36th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, IEEE, doi:10.1109/lics52264.2021.9470575. Available at <https://doi.org/10.1109/lics52264.2021.9470575>.
- [15] C. Ho Thanh, P.L. Curien & S. Mimram (2019): *A Sequent Calculus for Opétopes*. In: *2019 34th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, IEEE, doi:10.1109/lics.2019.8785667. Available at <https://doi.org/10.1109/lics.2019.8785667>.
- [16] G.L. Miller (1979): *Graph isomorphism, general remarks*. *Journal of Computer and System Sciences* 18(2), pp. 128–142, doi:10.1016/0022-0000(79)90043-6. Available at [https://doi.org/10.1016/0022-0000\(79\)90043-6](https://doi.org/10.1016/0022-0000(79)90043-6).
- [17] D. Reutter & J. Vicary (2019): *High-level methods for homotopy construction in associative n -categories*. In: *2019 34th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, IEEE, doi:10.1109/lics.2019.8785895. Available at <https://doi.org/10.1109/lics.2019.8785895>.
- [18] R. Steiner (1993): *The algebra of directed complexes*. *Applied Categorical Structures* 1(3), pp. 247–284, doi:10.1007/bf00873990. Available at <https://doi.org/10.1007/bf00873990>.

A Correctness of traversal algorithm

Proof of Lemma 2.13. Initially, the stack only contains U , which is a regular molecule by assumption. Assume, inductively, that the statement is true at the beginning of the current iteration with focus V , and that a set V' is pushed onto the stack at the end. Then either

1. $V' = \partial^\alpha V$ for some $\alpha \in \{+, -\}$, or
2. $V' = \text{cl}\{x\}$ for some $x \in V$.

In both cases, V' is a regular molecule and a proper subset of V (hence also of each item below V), under the assumption that V is a regular molecule. ■

Lemma A.1 — *Suppose V is on the stack. Then all elements of V must be marked before any item below V is accessed, or before any proper superset of V becomes the focus.*

Proof. By Lemma 2.13, as long as V is on the stack, only V and its proper subsets can be on top. It follows that, for a proper superset of V to be the focus, V must be popped from the stack at the end of an iteration where V is the focus. There are only two ways this can happen:

- V was already fully marked before the current loop iteration, or
- $\partial^- V$ was fully marked and $V = \text{cl}\{x\}$ for some x which is marked at the current loop iteration.

In both cases, $\partial^- V$ was already fully marked before the current loop iteration. In the latter case, if $\partial^+ V$ is already fully marked, then $V = \{x\} \cup \partial^- V \cup \partial^+ V$ is also fully marked. Otherwise, $\partial^+ V \subsetneq V$ gets pushed onto the stack to replace V , and must be popped before any superset of V becomes the focus. By the same case distinction, whenever $\partial^+ V$ is popped, either

- it was fully marked, in which case V was fully marked, or
- it is of the form $\text{cl}\{y\}$ for some y which is marked at the current loop iteration.

Either way, since all regular molecules satisfy the *globularity* property $\partial^\alpha(\partial^+ V) = \partial^\alpha(\partial^- V) \subseteq \partial^- V$, we know that $\partial^+ V$, hence V , is fully marked at the end of the iteration, and nothing is added to the stack. ■

Proof of Lemma 2.15. Suppose V is pushed onto the stack. As long as V is on the stack, any subsequent addition to the stack must be a proper subset of V , so it cannot be equal to V .

If V is popped from the stack, by Lemma A.1, it must be fully marked before any item below it is accessed. Since the algorithm checks if a set is fully marked before pushing it onto the stack, V can never appear again. ■

Proof of Lemma 2.16. First, we prove a weaker result: either V is fully marked, or there exists a *maximal* element of V which is unmarked.

Let $x \in V$ be marked. At some prior iteration, $\text{cl}\{x\}$ must have been the focus, and by Lemma A.1, in order for V to become the focus, $\text{cl}\{x\}$ must have been fully marked as well. Because

$$V = \bigcup_{k \leq n} \text{cl}\mathcal{M}_k V = \bigcup_{k \leq n} \bigcup_{x \in \mathcal{M}_k V} \text{cl}\{x\},$$

it follows that V is fully marked if and only if its maximal elements are all marked.

Now, V has one of the two forms in Remark 2.14. If V is of the second form, its maximal elements all have the top dimension, so we only need to consider the case $V = \partial_k^- U$.

At the start of the algorithm, $U, \dots, \partial_0^- U$ are all consecutively added to the stack. So $\partial_k^- U$ becomes the focus either at this stage, in which case *all* its elements are unmarked, or after $\partial_{k-1}^- U$ is fully marked. In the latter case, any maximal element of $\partial_k^- U$ of dimension strictly smaller than k also belongs to $\partial_{k-1}^- U$. ■

Proof of Theorem 2.12. As a particular case of Lemma A.1, U must be fully marked before the stack is emptied. Therefore, the algorithm either terminates after all elements have been traversed, or it does not terminate.

To prove that the algorithm does always terminate, it suffices to show that, unless all elements are already marked, it always finds an element to mark. First of all, observe that, from any state, the algorithm first goes through the following sequence of steps:

1. popping all fully marked subsets from the top of the stack;
2. once it reaches a subset which is not fully marked, successively pushing its lower-dimensional input boundaries that are not fully marked onto the stack.

At the end of this sequence, we always reach a state in which the focus V is not fully marked, but $\partial^- V$ is fully marked. Let us call such a V a *proper* focus.

We proceed by induction on dimension and proper subsets of a proper focus. If $\dim(V) = 0$, since a 0-molecule always consists of a single element, $V = \{x\}$, and x gets marked at the current iteration.

Let $n := \dim(V)$. By Lemma 2.16, there is an unmarked $x \in V_n$. If $V = \text{cl}\{x\}$, then x is marked at the current iteration, and we are done. Otherwise, we prove that there always exists a pair (y, x) where $x \in V_n$ is unmarked, and y is a marked input face of x . By [13, Lemma 1.16] applied to V , the coface x is unique given y , so among such pairs we can pick the one where y comes *earliest* in the list of marked elements, and this selects a unique x .

Let $x \in V_n$ be unmarked. By a dual version of [*ibid.*, Lemma 1.37], there exists a sequence

$$y_0 \rightarrow x_0 \rightarrow \dots \rightarrow y_m \rightarrow x_m = x$$

where $y_0 \in \Delta_{n-1}^- V$, $x_i \in V_n$, y_i is an input face of x_i , and y_{i+1} is an output face of x_i . Since V is a proper focus, y_0 is marked. Let k be the smallest index such that x_k is unmarked; because x_m is unmarked, such a k exists. Then x_i is marked for all $i < k$, hence $\text{cl}\{x_i\}$ is also marked. It follows that $y_k \in \partial^+ x_{k-1}$ is marked, and the pair (y_k, x_k) satisfies our requirement.

Thus, the algorithm will find a unique $x \in V_n$ and push $\text{cl}\{x\}$ onto the stack. The next proper focus will necessarily be a proper subset of V , and we conclude by the inductive hypothesis. ■