McMaster Centre for Software Certification

Sketches, Queries, Views, and Kleisli Composition: Towards Universal Algebra of Diagrammatic Operations with Pre- and Post-Conditions

Zinovy Diskin and Patrick Stünkel
Sketches, Queries, Views, and Kleisli Composition: Towards Universal Algebra of Diagrammatic Operations with Pre- and Post-Conditions

Zinovy Diskin\textsuperscript{1} and Patrick Stünkel\textsuperscript{2}

\textsuperscript{1} McMaster University, Centre for Software Certification, Hamilton, Canada
diskinz@mcmaster.ca
\textsuperscript{2} Høgskulen på Vestlandet, Institutt for Datateknologi, Elektroteknologi og Realfag, Bergen, Norway past@hvl.no

Abstract. A diagrammatic operation takes a diagram of a specified shape as its input and completes it with new elements forming a bigger diagram (e.g., arrow composition takes two consecutive arrows as its input and adds a third arrow resulting in a triangle). Pre- and post-conditions are an important part of the story (e.g., monic preservation by pullbacks) so that diagram chasing smoothly integrates logic and algebra. This technique transcends category theory and finds numerous applications in practice of software engineering, specifically in the multi-view approach to system design, and in model management (in the sense of model-driven engineering). These applications force us to consider a very general version of diagram chasing over generalized sketches (in the sense of Makkai) and give rise to a project (now in progress) reported in the present technical Report.

We introduce the notions of diagrammatic signatures, terms and algebras over an underlying sketch category, and prove several basic results about them; specifically, that the category of views is symmetric monoidal. Our views are a modification of the standard notion of a Kleisli mapping, in which each view $v : V \rightarrow Q_v(S)$ carries its own term $Q_v$ for extending the target $S$ of the mapping rather than using one unified but huge monad. Nevertheless, we prove that our localized Kleisli mappings generate a classical Kleisli triple and hence a monad.
1 Introduction

A diagrammatic operation takes a diagram of a specified shape as its input, and extends it with new elements forming a bigger diagram of a specified shape; we will say that the input diagram is completed by the operation. For example, arrow composition takes two consecutive arrows and completes them with a third arrow forming the corresponding triangle. The resulting diagram often satisfies some postconditions, e.g., a chosen pullback completes a cospan with a commutative square, and a chosen powerobject completes an object with its jointly-monic membership span. Preconditions are needed too, e.g., adhesive categories only provide pushouts for one-leg-monic spans (and then we know that the corresponding leg of the resulting cospan is also monic). Thus, categorical diagram operations come together with reasoning over them so that diagram algebra and logic are smoothly integrated in diagram chasing routines.
This paper stems from the observation that the diagram chasing idea transcends Category Theory (CT) and finds numerous applications in the practice of Software Engineering (SE), in which, however, it is currently used in an ad-hoc way: implicit, semi-formal or formal but buried in code so that the underlying diagram chasing procedures are blurred and distorted. The negative impact of these issues increases with the recent advent of Model Driven Engineering (MDE), in which diagrammatic models rather than code become the primary artifacts of software development.

Our goal in the project reported in the present technical report (further TR) is to provide an accurate formalization of diagrammatic reasoning routines adequate to their practical use in MDE. We will briefly discuss this in the next section (Sect. 2.1-2) along with several mathematically interesting structures and results emerged along the course (in Sect. 2.3); a brief survey of related work and our contributions are in Sect. 2.4. We also describe a simple example of domain modelling motivating generalized sketches as the main underlying machinery in Sect. 2.5, and overview our notation in Sect. 2.6. Section 3 gives three brief introductions to our three main subjects: diagrammatic logic, diagrammatic algebra, and diagrammatic reasoning. Sections 4-6 present technical development: in Sect.4 we consider sketches over predicate signatures (with or without dependencies between predicates), and in Sections 5-6 complex queries built from diagrammatic operations (the case of operation without dependencies is in Sect. 5 and the general case with dependencies in Sect. 5; the latter is in progress and currently only provides basic definitions). Section 7 concludes.

2 Background and Contributions

We discuss two main applications of the framework in SE in Sect.2.1-2, then discuss some mathematical aspects of the framework in Sect.2.3, and describe related work and contributions in Sect.2.4. Section 2.5 presents our main motivating example and Sect. 2.6 overviews our notation.

2.1 Views.

Developing a complex software system normally requires working with multiple views [ASB10] of the system: This is the only way to manage complexity and a cornerstone of system engineering, see [FKN\\'92] for an early discussion and [FKWVH19,BBCW19,CCP19] for more recent reports and surveys. A view of a source artifact S is, in general, not merely a part of S or a projection of S, but typically involves some computation over S, e.g. composing references, aggregating attributes by summing up their values etc, which extends S with derived elements produced by some predefined operations (we will say queries) applied to S. Then a view can be defined as a mapping v: V \rightarrow Q_v(S) in a suitable category of software artifacts where the source S and the view schema V live, and Q_v is a query that provides an extension Q_v(S) of source S with derived elements required by the view. Note an important distinction from the ordinary Kleisli mappings defined for a given monad Q or, equivalently, for a global mapping on objects, which only potentially exists and seems unnecessary abstract for software engineers not familiar with Haskell. In contrast, in our notion of view, each view mapping v carries its own query Q_v and maps to its own local extension Q_v(S) rather than to a unified but huge global extension.
Q(S). A simple example in Sect. 2.5 shows how it works, and also demonstrates the importance of constraints for the view mechanisms. It is placed aside to save time for the first reading, but we encourage the reader to take a look there to obtain a notion of the MDE world and our motivation for this paper.

2.2 Algebra for model management.

In model management, models are manipulated as holistic entities, and replacement the one-object-at-a-time approach (micromanagement) by one-model-at-a-time (macromanagement) turned out very productive [Ber03, FN05, PKR+13]. It requires considering high-level operations on models such as model slicing, transformation, merging, and synchronization, and composing them into complex workflows [DKM13]. Fig. 1 shows a simple example in which two views of the same source are first matched, and then merged. One may think of this as taking the pullback and then the pushout, but in practice these “pure” categorical operations are often interwoven with some normalization procedures (and may require conflict resolution) so that even in this simple case the real workflow will be much richer. Importantly, models in SE are typically considered up to their auto-isomorphisms generated by permutations of OIDs (in the sense of OO programming) which are irrelevant and normally not visible. It justifies making categories of models skeletal, which makes limits and colimits deterministic algebraic operations. Thus, for both the inter-model and the intra-model levels of reasoning, SE needs powerful and formally specified algebraic machinery for diagrammatic operations. On the other hand, it seems that even basic universal algebraic constructs (signature, term, parsing) have not been accurately specified for diagrammatic operations, especially when the latter are considered together with pre-and post-conditions and hence are interwoven with logic; in this sense, we are developing some basic universal algebra for diagrammatic operations.

2.3 Mathematical aspects

Why generalized graphs and generalized sketches As discussed above, categories hosting software models are categories of sketches. In general, a sketch is a graphical structure endowed with labelled diagrams (a.k.a. constraints); in the project, we consider a very general version of the notion: our ‘graphs’ are finitary presheaves, and our constraints of shape $G$ are any “devices” that classify mappings from $G$ as either valid or invalid. The generality of our graphs is caused by the great diversity and complexity of graphical structures used in SE [Ken02], and graphs used for behavioural and workflow models can even be more complex. Similarly, constraints used in domain modelling (banking, e-commerce, automotive etc.) can be complex and very elementwise-oriented so that their representation in a standard categorical way via universal properties can be rather bulky and unwieldy (a detailed discussion can be found in [DW07], see also Appendix). Thus, we need constraints and sketches of a very general nature that Michael Makkai called generalized sketches [Mak97].
2 Background and Contributions

**String diagrams do not work for diagram operations** For \( PROPs \), terms can be specified by string diagrams as shown in Fig. 2, which depicts a ternary term \((x + y) * (x + z)\) built from two binary operations. Note that reuse of the middle variable \( y \) is shown explicitly with \texttt{Copy}-operation (visualized via \( \leftarrow \)). This is possible as set \( \{x, y, z\} \) is the coproduct of its elements \( \{x\} \sqcup \{y\} \sqcup \{z\} \).

![Fig. 2: String diagram notation for ordinary algebraic terms](image)

However, diagram operations are graph-based, and a graph (while being a colimit of its elements) is not their coproduct!

Consider Fig. 3 presenting a simple diagram chasing piece. Part (A) uses a typical categorical notation: the diagram presents a workflow consisting of two pullbacks taken in parallel over spans with one monic leg, followed by yet another pullback. Monic preservation by pullbacks is also shown. Part (B) makes the above explicit by using labels for constraint declarations (ovals) and operation applications (chevrons), and showing their scope. Note that arrows going in and out of chevrons show the workflow and what elements are constructed/derived (blue and dashed) rather than given. Part (C) is similar to Fig. 2 and highlights the workflow of the query by making the order of operation applications explicit, e.g. the pullback \( c \) depends on elements produced by the other two pullbacks.

Note several important differences with Fig. 2. First, in discrete \( PROP\)-setting, the elements produced by an operation have no further relationship with the input elements, while in the graph-based setting, there is an overlap between the input and the output, e.g. element \( B \) in Fig. 3(B) is both in the input and the output of the pullback application \( a \), and similarly for application \( b \) so that \( B \) is also shared by \( a\)-output and \( b\)-output. Second, the workflow shown in Fig. 3(C) is a rather poor view of the actual workflow under the hood, which consists of several pushouts (POs): three POs are needed to produce graph \( X' \) and yet another one to produce \( X'' \). Workflow arrows with hollow ends are, in fact, pairs, \((X, a), (X, b)\) etc., while mappings that do the job, e.g., the one showing what part of graph \( X \) is used as an input for \( a\)[pb], are not shown. Thus, in contrast to Fig. 2, diagram Fig. 3(C) is an incomplete specification of the story.

2.4 Related Work and Contributions

By *logic* we refer to work on generalized sketches, and *algebra* refers to work on signatures of operation symbols, terms and algebras over them.

**Logic.** Several authors noticed that Ehresmann’s sketches can be applied to data modelling [LS91,PS95,JKR02]. However, these sketches are based on universal properties of the corresponding (limit, colimit, powerobject) diagrams, which (as we discussed above) is a too restrictive setting for applications. A
Fig. 3: Diagram Chasing
suitable generalization of sketches by exactly this reason was proposed by Diskin [Dis96,Dis97], and independently and even earlier, in a very different context of an abstract approach to logic, by Makkai [Mak97] (whose preprints had been circulating in early 90s). Makkai developed a coherent theory of generalized sketches and proved a fundamental result that any category of sketches is a presheaf topos. Hence, sketches in a signature $\mathcal{P}_1$ can be considered graphs over which a new signature $\mathcal{P}_2$ and hence a new category of $\mathcal{P}_2$-sketches can be defined, and so on. In this way, Makkai specified categories with several layers of additional structure as sketches. In our project, we use a different approach in which all discrete signatures $\mathcal{P}_i$ are integrated into one signature-category $\check{\mathcal{P}}_i$, whose arrows describe dependencies between predicate symbols extracted from arities (see Sect. 4 for details). We prove that Makkai’s theorem can be extended for categorical signatures so that the slogan “sketches are graph” persists (see Th. 1 for a precise formulation). Also, while traditionally sketches are defined in an *indexed* way as families of sets of diagrams indexed by predicate symbols (for the discrete signature case), or functors from the signature (for the categorical signature case [DW07]), we define them in an equivalent *fibrational* way via a discrete opfibration functor into the signature: the fibrational definition is more compact and often simplifies technicalities.

To give a detailed overview of what is done in the paper in comparison with other work, we need to distinguish between several notions of signatures and sketches. We call logic *basic* if signatures of predicate symbols are sets considered as discrete categories. We call it *advanced*, if signatures are categories whose arrows are dependencies between predicates; in addition, these arrows are required to go in one direction, which makes signature categories *hierarchical* (a version of Lawvere’s one-way categories). There is also an important distinction between *multi*-sketches that allow the same labelled diagram appear in a sketch for multiple times, and *mono*-sketches that disallow this. Finally, a partially de-categorificated approach (due to Makkai) is to consider signatures as layered sets so that each predicate in layer $(n + 1)$ implicitly depends on all predicates in layer $n$. However, when arities of $(n + 1)$-predicates are defined based on $n$-predicates, the actual dependencies appear on the stage. We call this approach *discrete hierarchical(dh)* and show that it is actually equivalent to the categorical one. In these terms, Table 1 shows the previous work and contributions of the project. Our main results here are an extension of a fundamental Makkai’s theorem (that multi-sketches are actually presheaves) for the case of categorical signatures with dependencies, Th. 1, and an equivalence of the two approaches to sketches: internally categorical (signatures are categories) and discrete hierarchical (a hierarchy of discrete signatures), Th. 2.

<table>
<thead>
<tr>
<th>Sketch Def.</th>
<th>Setting (Basic Logic)</th>
<th>Multi-Layer (Advanced Logic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indexed</td>
<td>mono-sketches</td>
<td>Makkai [Mak97]</td>
</tr>
<tr>
<td></td>
<td>multi-sketches</td>
<td>Diskin &amp; Wolter [DW07]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Sect. 4.2</td>
</tr>
<tr>
<td></td>
<td>Fibrational (both mono &amp; multi)</td>
<td>Sect. ?? (Motivational)</td>
</tr>
</tbody>
</table>

Table 1: Diagram Logic and Contributions
Table of Contents

Algebra. Diagram operation signatures and algebras were defined first in [Dis97] and with some additional details in [Dis96], but no mathematical results were obtained. For the special case of operations over ordinary graphs, Wolter et al. [WDK18] proved that any signature of such operations generates a monad (i.e., a global construct) over the category of graphs. The machinery of local Kleisli mappings explained informally was used in several papers of the first author, e.g., [DXC10, DGC17], and formalized as Kleisli mappings of a monad (i.e., globally) in [DMC12]; we are not aware of an accurate formalization of Kleisli mappings considered locally as is done in our project, nor of several other results reported in Sect. 5. Specifically, we prove that a) sketches and local Kleisli mappings (views) between them form a symmetric monoidal category, and b) the local Kleisli composition gives rise to a Kleisli triple and hence a monad (as expected). Table 2 provides an overview of related work in this area and where to find our results in this paper.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Setting</th>
<th>Discrete</th>
<th>Categorical (Dependencies)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signatures and Algebras</td>
<td>Sect. 3.2; [WDK18] (graph operations)</td>
<td></td>
<td>Sect. 6.1, 6.2</td>
</tr>
<tr>
<td>Multiqueries and Substitution</td>
<td>Sect. 5.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Query Semantics</td>
<td>Sect. 5.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Complex Queries &amp; Composition</td>
<td>Sect. 5.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monads</td>
<td>Sect. 5.5; [WDK18] (graph term algebra adjunction)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Diagram Algebra and Contributions

2.5 Domain Modelling with Class Diagrams and their Mappings

![Sample UML diagram](image)

Fig. 4: Sample UML diagram

2.5.1 Domain Modelling with Class Diagrams. Figure 4 presents a toy example of domain modelling with UML class diagrams [RJB04]. The source model S is specified by a UML class diagram describing a part of a university information structure in an almost self-explanatory way. Boxes refer to classes
2 Background and Contributions

of objects Person, Course etc. to be instantiated by object identifiers, and show their attributes (name, workload, etc.) instantiable by functions, e.g., for an object \( o \in [\text{Person}] \) with \([\text{name}](o) = \text{Bill_Clinton}\) (where \([x]\) denotes a semantic interpretation of an element \(x\) of the diagram). Arrows denote associations to be instantiated by mappings (relations or functions), and expressions \([m:n](m, n \in [0, 1, 2, \ldots, *])\) (shown in red) near associations ends are called multiplicities and constrain the type of relation/function, e.g. multiplicity \([3..5]\) at the source end of association student says that each student may only have from 3 to 5 subscriptions, i.e., for a valid instantiation \([\_\_\_\_\_]\), for any object \(s \in [\text{Student}]\) the following condition is to hold:

\[
\text{card}([b \in [\text{Subscription}]] \mid [\text{student}](b) = s) \in [3..5],
\]

and if this condition is violated, instantiation \([\_\_\_\_\_]\) is considered invalid (and some alarm is to be fired).

Constraints often appear as implementations of a set of requirements (typically conflicting), e.g., the Student Health Office may require the multiplicity above to be \([2..5]\) while the Graduate Office requires it to be \([3..*]\). Keeping the backward traceability to requirements is important and often crucial for a proper domain modelling. By default (and thus omitted in Fig. 4), every association arrow has the target multiplicity \([1..1]\) (i.e. exactly one) and the source multiplicity \([0..*]\) (i.e. arbitrary), which makes the relation a function.

Furthermore, there are different arrow-types, e.g. the arrows between Employee and UniversityRole, and between Student and UniversityRole, show the inheritance (subset) relationship between classes. It means that both Employee and Student inherit the person association from UniversityRole. Multiplicities constraints and inheritance are typical examples of built-in-constraints in UML. However, there may be more complex domain-specific rules that cannot be expressed with built-in constraints and require special attached constraints commonly formulated in the Object Constraint Language (OCL) [WK99]. For example, requirement Req 3 that student teaching assistants are not allowed to earn more than 15,000.00 a year due to tax regulations would be formalized as shown in Listing 1 and indicated in Fig. 4 with a note-symbol attached to class Employee. We can express this constraint in an element-free way by taking suitable pullbacks along inheritance monics and the monic \(\{15000\} \hookrightarrow \text{Real}\), but it would be a bulky and opaque construction for a practical engineer. Instead, in our generalized sketch framework, we consider such an OCL constraint as a formal constraint declaration \(C\) whose scope \(\text{diagr}(C)\) consists of the inheritance triangle (i.e., the corresponding cospan) and attribute salary (considered as an arrow from Employee to Real). Semantics of the constraints is the Boolean-valued procedure specified in the Listing, which classifies any configuration of sets and mappings over the scope as either valid or invalid.

Listing 1: OCL constraint

```plaintext
context Employee inv TA_Payment:
    Student.allInstances() -> exists(s | s.person = self.person) implies self.salary <= 15000.00
```

2.5.2 View definition example. An application working with the university database may only need a part of information stored in the database (due, say, to security concerns), e.g., it may only need data about students and courses
they take. Such data are specified by schema $V$ specified in the right part of the figure, which is called a view of $S$. To make the view definition formal, we need to map schema $V$ to schema $S$, but there is an issue: association takes is not present in $S$. The issue is resolved by noticing that the derived association $\text{student}^{-1}.\text{course}$ produced by, first, inverting $\text{student}$, and then composing it with association $\text{course}$, is actually what view $V$ means by $\text{take}$. Multiplicities of the new association are also derivable. In other words, by applying operations/-queries $q_1$ (inversion) and $q_2$ (composition), we derive new elements of model $S$ shown in blue (to suggest the mechanical nature of computation) and dashed (for the black-white printing); note that queries $q_1$, $q_2$ are diagram operations of obvious diagrammatic arities. They can be defined in OCL as shown in Listing 2 (where $\text{student}^{-1}$ is called $\text{student_inverse}$ and $\text{student}^{-1}.\text{course}$ is called $\text{student_to_course}$ due to a limited character palette).

Listing 2: Derived elements in OCL

```
context Student :: student_inverse : Subscription[2..*]
derive Subscription.allInstances()->select(s|s.student = self)
```
```
context Student :: student_to_course : Course[2..*]
derive Subscription.allInstances()->select(s|s.student = self)
  .collect(s|s.course)
```

Now we can build the view definition mapping by mapping arrow $\text{takes}$ in $V$ to the derived arrow $\text{student}^{-1}.\text{course}$ in $S$. Formally, we express this as a mapping $v: Q_v(S) \leftrightarrow V$ where $Q_v = \{ q_1, q_2 \}$ and $Q_v(S)$ is the full (black, red and blue) model/sketch shown in the figure, whereas the source model is the (black and red) subsketch $S$ not including the derived arrows and derived constraints.

Importantly, the view definition mapping $v$ is compatible with constraints:

\[
\langle \text{context :: student}^{-1}.\text{course} \rangle [3..5] \models [2..5]
\]

in the sense that any instantiation $\llbracket . \rrbracket _S$ satisfying $[3..5]$ will also satisfy $[2..5]$ and hence its view $\llbracket . \rrbracket _V$ also satisfies $[2..5]$ as required: compatibility of a view mapping with constraints ensures the correctness of views. Thus, an important requirement to a view definition mapping $v$ is to be a sketch morphism, so the category hosting our artifacts is to be a category of sketches rather than graphs.

2.6 Notation

The goal is to fix our notation and terminology. To manage size issues, we consider all categories in the paper are small and refer to the idea of Grothendieck universes for handling categories whose collections of objects and morphisms are “bigger” than sets. For variables referring to categories, we use blackboard font, e.g. $G$. Names of special categories such as $\text{Set}$ or $\text{Cat}$ are in bold italic.

**Notation for objects and arrows.** For a small category $C$, its set of objects and set of arrows (morphisms) will be denoted by $\text{Ob}(C)$ and $\text{Arr}(C)$ resp. We will also write $\text{Elem}(C)$ for the set $\text{Ob}(C) \sqcup \text{Arr}(C)$.

As a rule, we will use uppercase letters for objects and lowercase letters for morphisms; then we can use this agreement to ease formulas and write $X \in C$ meaning $X \in \text{Ob}(C)$ and $f \in C$ meaning $f \in \text{Arr}(C)$. The domain of an arrow $f$ will be denoted by $\text{dom}(f)$ or sometimes $\circ f$, and the codomain by $\text{cod}(f)$ or $f \circ$ (one may as well think of $\text{dom}$ and $\text{cod}$ as mappings from $\text{Arr}(C)$ to $\text{Ob}(C)$).
We write $f: X \to Y$ (or $X \xrightarrow{f} Y$ in diagrams) for an arrow $f \in \text{Arr}(\mathcal{C})$ with $\text{dom}(f) = X$ and $\text{cod}(f) = Y$.

The hom-set of all arrows between two objects $X,Y \in \mathcal{C}$ will be denoted $\mathcal{C}(X,Y)$. We write $\mathcal{C}(X,-)$ for $\bigcup_{Y \in \text{Ob}(\mathcal{C})} \mathcal{C}(X,Y)$ and similarly for $\mathcal{C}(-,Y)$.

> Normally we assume that $\mathcal{C}(X,Y)$ is the set of arrows from $X$ to $Y$, but in sect. we will have a special exception... ●

**Arrow composition and application.** Composition of two arrows $f: X \to Y$ and $g: Y \to Z$ will be denoted either in the diagrammatic order $f \circ g: X \to Z$ or the application order $g \circ f: X \to Z$.

For a mapping $f: X \to B^A$ and $x \in X$, we will often write $f^x$ or $f_x$ rather than $f(x)$ to make reading applications of $f(x)$ to elements of $A$ easier: $f^x(a)$ rather than $f(x)(a)$ (although the latter form will be used too). Similarly, if $f(x)$ is a domain of another arrow $g: f(x) \to y$, we will often write $x^f$ for $f(x)$ so that $g: x^f \to y$.

To highlight special properties of morphisms, we denote a monomorphism $m$ by $m: X \rightarrow Y$ and an epimorphism $e$ by $e: X \rightarrow Y$. Finally, we will interchangeably use expressions $\mathcal{C}^B$ or $[\mathcal{C}, \mathcal{D}]$ to denote the functor category between $\mathcal{C}$ and $\mathcal{D}$, $\mathcal{C}/X$ for the slice category under $X \in \text{Ob}(\mathcal{C})$, $\mathcal{C} \backslash C$ for the coslice category over $X \in \text{Ob}(\mathcal{C})$, and $\mathcal{C} \times \mathcal{D}$ for the product category. Class $\text{Ob}([\mathcal{C}, \mathcal{D}])$ will be often denoted by $[\mathcal{C}, \mathcal{D}]^*$ or just $[\mathcal{C}, \mathcal{D}]$ if the context makes discretization clear.

If $f: \mathcal{C} \to \mathcal{D}$ is a functor and $\mathcal{C}' \subseteq \mathcal{C}$ is a subcategory, then the restriction of $f$ on $\mathcal{C}'$ (i.e., composition $i_{\mathcal{C}'} \circ f$ where $i_{\mathcal{C}'}$ is the corresponding embedding) is denoted by $f|_{\mathcal{C}'}$.

**Carriers.** In some contexts, where the domain of a mapping gives the name of the story, we will use the following (categorically, rather odd) naming for components of an arrow $f: X \to Y$. We will consider this arrow as a pair $X = (|X|, f_X)$ where the domain of $f$ is now denoted by $|X|$ and called the **carrier** of $X$ while the arrow itself is denoted by $f_X$ and is referred to as the **mapping** of $X$. All these are fully terminological permutations as arrow $f$ and pair $X$ are the same formal construct. Moreover, by a common abuse of notation, we will overuse symbol $X$ to refer to the carrier $|X|$, and will often omit the superscript $X$ in $f_X$.

### 3 Diagram Logic and Algebra: Overview

We will give three brief introductions: to a sketch version of diagrammatic predicate logic in Sect. 3.1, to a diagrammatic algebra in Sect. 3.2, and to diagrammatic reasoning in Sect. 3.3

#### 3.1 Diagram Logic via Generalized Sketches: Basic Notions

A paradigmatic object of logical studies is a satisfiability relation between *theories*—syntactical constructs built from predicate symbols, variables and logical symbols, and *models*—semantic constructs built from predicate symbols, values and set-theoretical operations interpreting logical symbols. Categorical logic places syntax and semantics in a suitable category $\mathcal{L}_\mathcal{P}$ where $\mathcal{P}$ is a fixed signature and we will omit the subindex, and makes a model $M$ of a theory $T \in \text{Ob}(\mathcal{L})$
an \(\mathbb{L}\)-morphism \(\llbracket \cdot \rrbracket_M : T \to |M|\) where \(|M|\) is an \(\mathbb{L}\)-object called the carrier of model \(M\). In the sketch view of logic, the logical part is very simple: the only logical symbol is conjunction so that theories are merely sets of formulas, but the carrier part is interesting: carriers are graph-based structures that can be rather complex (2-graphs, hypergraphs, Petri nets, etc) and further are referred to as graphs. Thus, arities of predicate symbols are graphs, and a collection of variables \(X\) is a graph too, so that an atomic formula over \(X\) becomes a labelled diagram \((P, d)\) with \(P\) a predicate symbol (of arity, say, \(\alpha(P)\)), and \(d : \alpha(P) \to X\) a diagram (graph morphism) in \(X\) of shape \(\alpha(P)\). A sketch is nothing but a collection of labelled diagrams (i.e., of diagrammatic atomic formulas). A suitable choice of a predicate signature can make this simple formalism extremely expressive. In this section, we will give simple variants of several basic definitions to show how the machinery works, and postpone a more detailed study until Sect.4-5.

**Generalized graphs.** Ordinary (directed, multi-)graphs can be seen as functors (presheaves) \(G : \text{Graph} \to \text{Set}\) over schema specified in (1)

\[
\begin{array}{ccc}
\text{Node} & \xrightarrow{\text{src}} & \text{Arrow} \\
\downarrow{\text{trg}} & & \\
\end{array}
\]

Schema \text{Graph} for ordinary (directed multi-)graphs

Graph morphisms are natural transformations, and the category of all graphs is a presheaf topos \(\text{Set}^{\text{Graph}}\). Our applications require to consider more general notions of graphs that could be seen as presheaves over schemas \(G\) more complex than \(\text{Graph}\). In general, a graph is a presheaf \(G : \mathcal{G} \to \text{Set}\) over a small category \(\mathcal{G}\), and the category of graphs is a presheaf topos \(\mathcal{G} = \text{Set}^{\mathcal{G}}\). In Sect. 4.1 below we will consider this in more detail, but for the current Overview section, graphs can be understood as \(\text{Graph}\)-graphs. A graph \(G\) is finite if the set \(\text{Elem}(G)\) is finite. All finite graphs form a (sub)category (of \(\mathcal{G}\)) denoted by \(\mathcal{G}_\omega\).

**Predicate signatures and their models.**

**Definition 1 (Discrete Signatures and Mono Sketches).** Let \(\mathcal{G}\) be a schema for graphs and \(\mathcal{G} = \text{Set}^{\mathcal{G}}\) is the corresponding presheaf topos. A discrete predicate signature \(\mathcal{P}\) over \(\mathcal{G}\) is given by a set \(|\mathcal{P}|\) of predicate symbols and an arity function \(\alpha : |\mathcal{P}| \to \text{Ob}(\mathcal{G})\). For a \(P \in |\mathcal{P}|\), we say that graph \(P^\alpha = \alpha(P)\) is the arity of \(P\). We will often (inaccurately) write \(\mathcal{P}\) for \(|\mathcal{P}|\) too.

A (mono) sketch \(S\) over signature \(\mathcal{P}\) is given by a graph \(|S|\in\mathcal{G}\) and a mapping \(\llbracket \cdot \rrbracket_S\) such that diagram (2)(a) commutes.

\[
\begin{array}{ccc}
\mathcal{P} & \xrightarrow{\llbracket \cdot \rrbracket_S} & \mathcal{G}/|S| \\
\downarrow{\alpha} & & \downarrow{\text{dom}} \\
\text{Ob}(\mathcal{G}) & \xrightarrow{\text{dom}} & \text{Ob}(\mathcal{G}) \\
\end{array}
\quad (a)
\]

\[
\begin{array}{ccc}
\mathcal{P} & \xrightarrow{\llbracket \cdot \rrbracket_U} & \text{Pred}(U) \\
\downarrow{\alpha} & & \downarrow{\text{dom}} \\
\text{Ob}(\mathcal{G}) & \xrightarrow{\text{dom}} & \text{Ob}(\mathcal{G}) \\
\end{array}
\quad (b)
\]

Given two sketches \(S, S'\), a morphism \(f : S \to S'\) between them is a graph morphism \(|f| : |S| \to |S'|\) such that for any \(d \in \llbracket P \rrbracket_S\), we have \(d \circ f \in \llbracket P \rrbracket_{S'}\), i.e.,
postcomposition with \( f \) maps diagrams in \( S \) to diagrams in \( S' \) with the same label.

**Corollary 1 (Sketch category).** It is easy to see that composition \( |f| \circ |f'| : |S| \to |S'| \circ |S''| \) is a sketch morphism as soon as \( f \) and \( f' \) are such, and an identity graph morphism is a sketch morphism. This gives us a category \( G|\mathcal{P} \) of sketches over \( \mathcal{P} \), which is supplied with an obvious forgetful functor \(|.\) : \( G|\mathcal{P} \to G \). This functor is faithful and surjective on objects (any graph is a sketch with all diagram sets empty) but not full.

\( F \) or sketches of syntactical origin, graph \( X = |S| \) can be understood as a graph of variables typically countable or finite, and a diagram \( d \in \Pi \mathcal{P} \), i.e., a graph morphism \( d : P^\alpha \to X \), can be seen as substituting variables from \( X \) for placeholders in \( P^\alpha \). In MDE, pairs \( (P, d) \) are called constraint declarations (e.g., see Example in Fig. 4). For an ordinary string-based logic where graphs are (multi-sorted) sets, a diagram is a tuple of variables \((x_1, ..., x_i, ...)\), \( x_i = d(e_i) \), for some enumeration \( (e_1, ..., e_i, ...) \) of the arity \( P^\alpha \), and a labelled diagram \((P, d) \) is nothing but a formula \( P(x_1, ..., x_i, ...) \).

Two types of sketch semantics should be distinguished: elementwise (or value-oriented in terms of relational databases) and setwise (object-oriented as is typical for MDE). For the former, elements of the carrier \(|U|\) of a sketch \( U \) of semantic origin are values and links between them, and labelled diagrams are thus labelled tuples \( P(u_1, ..., u_i, ...) \) representing true facts about the model \( U \). Sets of such tuples are predicates, and diagram (2)(a) above is re-denoted as shown in diagram (b), where \( \text{Pred}(U) \) \( \overset{\text{def}}{=} \{P_X \subseteq \mathbb{G}(X, U) | X \in \text{Ob}(\mathbb{G})\} \) and \( \text{dom}(P_X) = X \). For the setwise semantics, the carriers are graphs rather than sets, there are just few sketches of semantics origin such as \( \text{Set} \) or \( \text{Rel} \) or \( \text{Span(Set)} \) considered as sketches, whereas individual user-defined models are sketch morphisms into these. The following example clarifies these remarks.

**Example 1 (Elementwise vs. Setwise Logics).** Let \( R \subseteq A \times B \times A \) be a two-sorted ternary predicate symbol, and let \( R_{\Pi} \subseteq \Pi[A] \times \Pi[B] \times \Pi[A] \) be a relation given elementwise, i.e., as a set of triples \((a_1, b, a_2) \in \Pi[A] \times \Pi[B] \times \Pi[A] \). This gives us a model of signature \( \mathcal{P} = \{R\} \) further referred to as \( \Pi \).

In our diagram logic, the graph schema is a two-element set (of sorts), \( G = \{A, B\} \), and a graph \( G \) is thus a pair of sorts \((G(A), G(B)) \). Arity \( R^\alpha \) is the following graph: \( R^\alpha(A) = \{a_1, a_2\} \) and \( R^\alpha(B) = \{b\} \), where \( a_1, a_2, b \) are some chosen “placeholders”. Although \( R^\alpha \) is a pair of sets, it is convenient to specify it as a labelled triple \((a_1 : A, b : B, a_2 : A)\).

A \( R \)-model \( \Pi \) described above is almost immediately a sketch \( S_{\Pi} \) whose carrier graph \( U := |S_{\Pi}| \) is a pair of sets, \( U(A) = |A|, U(B) = |B| \), and the diagram set is defined as follows:

\[
|R|_{S_{\Pi}} = \left\{d \in G(R^\alpha, U) | (d(a_1), d(b), d(a_2)) \in R_{\Pi}\right\}
\]

so that diagrams in \( S_{\Pi} \) are triples in \( R_{\Pi} \) (i.e., true facts about model \( \Pi \)).

Consider now a constraint for the model \( \Pi \), say, a universal first-order sentence \( \forall \varphi \) that states \( (\forall x \in A, \forall y \in B) R(x, y, x) \) (so that \( \varphi \) is formula \( R(x, y, x) \)). To specify this in our diagram logic, we introduce a formal syntactical sketch \( S_{\varphi} \) with the carrier graph \( X := |S_{\varphi}| \) defined by setting \( X(A) = \{x\}, X(B) = \{y\} \)
and the only (formal) diagram \( d \) in \([ R ]_G\) is defined by formula \( R(x, y, x) \) (i.e., \( d(a_1) = d(a_2) = x \) and \( d(b) = y \)). It is easy to see that the standard definition of the satisfiability relation, \([ \square ] = \forall \varphi\), actually means that any graph morphism \( f: X \to U \) (i.e., a mapping with \( f(x) \in \[ A \], f(y) \in \[ B \] \)) is a correct sketch morphism that maps the only formal true fact in \( S_{\varphi} \) to a true fact in \( \square \), that is, \((f(x), f(y), f(x)) \in R[\square]\). This is exactly the first-order logic semantics of universal formulas: any evaluation of variables results in a true fact about the model. Thus,

\[
\square \models \forall \varphi \quad \text{iff} \quad G^\text{ew}|P(S_{\varphi}^\text{ew}, S_{\varphi}^\text{ew}) \cong G^\text{ew}(|S_{\varphi}^\text{ew}|, |S_{\varphi}^\text{ew}|)
\]

(3)

where the superscript ‘ew’ stands for “elementwise”: we added it to distinguish the above from the setwise framework considered below. Recall that \( G^\text{ew} = \{A, B\} \).

In the setwise sketch logic, relation \( R \) is considered to be a sort rather than a predicate. It is a fundamental idea that changes everything. First, our carriers are now graphs with schema (1) and the carrier \( X \) of syntactical sketch \( S_{\varphi}^\text{sw} \) is a ternary span with apex \( R \) and three projection legs \( p_i: R \to A, i = 1, 2, q: R \to B \); in addition, we require these legs to be jointly monic and hence place predicate symbol \([\text{monic}_3]\) with arity being graph \( \text{Span}_3 \) into the signature, and include the corresponding diagram \( d: \text{Span}_3 \to X \) into \( S_{\varphi}^\text{sw} \).

The semantic story begins with a functor \([ \_ ]^\text{sw}: X \to \text{Set} \), which is required to be a \([\text{monic}_3]\)-sketch morphism (as \( \text{Set} \) is an obvious \([\text{monic}_3]\)-sketch too). To ensure that relation \([ R ] \) satisfies formula \( \forall \varphi \), we add to the signature a predicate symbol \( P_{\varphi} \) of arity \( \text{Span}_3 \), and add to sketch \( S_{\varphi}^\text{sw} \) one more constraint declaration (labelled diagram) described by formula \( P_{\varphi}(p_1, q, p_2) \). Finally, we convert \( \text{Set} \) into a \( P_{\varphi}\)-sketch by defining

\[
[P_{\varphi}]_{\text{Set}} = \{d \in \text{Set}^{\text{Span}_3} | d \models \forall \varphi\}
\]

. Now we can state

\[
\square \models \forall \varphi \quad \text{iff} \quad [\_ ]^\text{sw}: S_{\varphi}^\text{sw} \to \text{Set}_P \text{ is in } G^\text{sw}|P^\text{sw}
\]

(4)

where \( G^\text{sw} \) is \( \text{Graph} \) (see (1)), \( P^\text{sw} = \{[\text{monic}_3], P_{\varphi}\} \), and \( \text{Set}_P \) is \( \text{Set} \) considered as \( P \)-sketch. Note that the sketch treatment in (4) works for any constraint, not necessarily universal. Whatever logic is employed for writing a closed formula \( \psi \), we introduce a predicate \( P_{\psi} \) of the corresponding arity, and as satisfiability classifies models into valid and invalid, we can convert \( \text{Set} \) (or other semantic universe) into a \( P \)-sketch by including into class \( [P_{\psi}]_{\text{Set}} \) exactly those diagrams, which satisfy formula \( \psi \).

**Multi-diagram vs. unique-diagram sketches.** To be completed

If relation \([ P ] \) were stored in a database, duplication of its rows (triples as above) would normally be allowed and considered normal: typically, a row records a state of an object, and two different objects may have the same state at one database snapshot at some time moment and be different at another snapshot/time moment.

Duplication is useful in syntax too: See discussion in Sect. 2.5.

### 3.2 Diagram Operations and Algebras: Basic Intuitions

Below we assume given some fixed category of sketches \( S = G||P \) (which, according to Makkai’s extended theorem, can be considered graphs).
Definition 2 (Diagram Operation). An (diagrammatic) operation is a pair \((\phi, \alpha_\phi)\) with \(\phi\) the name (or the naming symbol) and \(\alpha_\phi: I_\phi \to E_\phi\) a sketch monic called \(\phi\)'s arity. Let \(A \in \mathbb{S}\) be a \(\mathcal{P}\)-sketch and \((\phi, \alpha_\phi)\) an operation. We say that \(A\) carries the operation \(\phi\), or \(A\) is an \(\phi\)-algebra, if for any sketch mapping \(b: I_\phi \to A\) (called binding) there is one and only one sketch mapping \(\|\phi\|^A(b): E_\phi \to A\) (the result) such that the triangle \((I_\phi, E_\phi, A)\) in diagram (5) below commutes: \(\alpha_\phi \circ \|\phi\|^A(b) = b\). This **commutativity condition** is fundamental: it ensures that an operation does not change its input data.

\[
\begin{array}{c}
I_\phi \\
\downarrow^b \\
A \\
\downarrow^f \\
A' \\
\end{array}
\quad \quad
\begin{array}{c}
\alpha_\phi \\
\|\phi\|^A(b) \\
\|\phi\|^A(b \circ f) \\
\end{array}
\quad \quad
\begin{array}{c}
E_\phi \\
\rightarrow \\

\end{array}
\quad (5)
\]

Given another algebra \(A'\) for \(\phi\), a sketch morphism \(f: A \to A'\) is called \(\phi\)-compatible iff \(\|\phi\|^A(b) \circ f = \|\phi\|^A(b' \circ f)\) holds for every \(b: I_\phi \to A\).

Definition 3 (Discrete Operation Signatures and Algebras). A discrete operation signature \(\mathcal{O}\) over \(\mathbb{S}\) is given by a set \(|\mathcal{O}|\) of operation symbols and an arity function \(\alpha: |\mathcal{O}| \to [\cdot \to, \mathbb{S}]^*\) (whose target is considered discrete). For a symbol \(\phi \in |\mathcal{O}|\), we say that monic \(\alpha(\phi): I_\phi \to E_\phi\) is the arity of \(\phi\). We will also denote this monic by \(\phi^\alpha\) or \(I E_\phi\) and simply write \(\mathcal{O}\) for \(|\mathcal{O}|\).

A signature \(\mathcal{O}\) is finitary if \(\mathcal{P}\) is finitary and sketches \(I_\phi, E_\phi\) are finite for all \(\phi \in \mathcal{O}\). This is what we have in applications, but many of our results do not depend on the signature’s (in)finitarity.

An algebra \(A\) for \(\mathcal{O}\) is given by a carrier sketch \(|A|\) and an operation

\[
\|\phi\|^A: \mathbb{S}(I_\phi, |A|) \to \mathbb{S}(E_\phi, |A|)
\]

for each \(\phi \in \mathcal{O}\) (which means commutativity of the corresponding triangle for each \(\phi\)). An algebra homomorphism is a sketch morphism compatible with all operations in \(\mathcal{O}\).

The following is obvious.

**Proposition 1 (Category of \(\mathcal{O}\)-algebras).** There is a category \(\mathbb{S}^\mathcal{O}\) whose objects are algebras with \(\mathbb{S}\)-carriers and arrows are \(\mathcal{O}\)-homomorphisms.

**Example 2 (Precategories).** A precategory is a graph with an arrow composition and identity loops. We formalize this in our framework with the following signature: \(\mathcal{G} = \text{Graph}, \mathcal{P} = \emptyset,\) and \(\mathcal{O}_{\text{preCat}} = \{[\text{id}], [\cdot]\}\) with arities specified in Fig. 5. Then any \(\mathcal{O}\)-algebra is a precategory, and \(\mathcal{O}\)-homomorphisms are functors between precategories. Of course, precategories are not very interesting objects as they are not required to satisfy associativity and unitality conditions. We will return to the example later when we discuss reasoning capabilities of diagrammatic operations.

**Example 3 (Universal algebra).** Consider \(\mathcal{G} = \mathbb{1}\) so that \(\mathcal{G} \cong \text{Set}\), and \(\mathcal{P} = \emptyset\) so that \(\mathcal{G}||\mathcal{P} \cong \mathcal{G} \cong \text{Set}\). Every “classical” operation signature \(\Omega\) with an
arity function \( \text{ar} : |\Omega| \to \text{Nat} \) can be interpreted as a diagrammatic signature \( \mathcal{O}_\Omega = (|\Omega|, \alpha) \) such that for each symbol \( \omega \), its arity is given by the input set \( I_\omega = \text{ar}(\omega) \) considered as a set, the scope set \( E_\omega = I_\omega + \{\ast\} \) and the obvious monic \( m_\omega : I_\omega \twoheadrightarrow E_\omega \). Any \( \mathcal{O}_\Omega \)-algebra \( A \) is an \( \Omega \)-algebra \( A \): let \( n = \text{ar}(\omega) \) and \( \alpha \in A^n \), then setting \( [\omega]^A(a) \overset{\text{def}}{=} [\omega]^A(\ast) \) defines an \( \Omega \)-algebra structure on \( |A| \). The converse construction from a classical \( \Omega \) algebra \( A \) to a diagrammatic \( \mathcal{O}_\Omega \) algebra \( \hat{A} \) is equally obvious: for any \( a \in A^{|\Omega|} \) and any \( x \in I_\omega \subset E_\omega \), \( [\omega]^\hat{A}(x) \overset{\text{def}}{=} a(x) \), and \( [\omega]^\hat{A}(\ast) \overset{\text{def}}{=} [\omega]^A(\ast) \). Moreover, \( \hat{A} \cong A \) and \( \hat{A} \cong A \). It is also easy to check that the categories \( \mathbb{S}^{\mathcal{O}_\Omega} \) and \( \Omega^{-\text{Alg}} \) are isomorphic. Thus, the ordinary universal algebra is a specialization of the discrete diagrammatic algebra.

### 3.3 Reasoning via diagrammatic operations

A principal feature of diagram chasing is simultaneous derivation of both new elements and their properties. A typical example is “monics are stable under pullbacks”. Moreover, operations only deriving properties also make sense and can be seen as implications. For example, the passage from precategories to categories can be described by requiring precategories to carry three more diagrammatic operations ensuring associativity of the composition and left and right unitality of the identity. To see this, we fist note that any diagram operation \( \phi \) with arity \( \phi^\alpha : I_\phi \to E_\phi \) gives rise to a diagram predicate \([\phi]\) with arity \( [\phi]^\alpha = E_\phi \). Thus, the operational signature \( \mathcal{O}_{\text{preCat}} \) gives us a predicate signature \( \mathcal{O}_{\text{preCat}}^E \).

Now we can specify the associativity of composition by two operations. The first has arity shown in Fig. 6. In sketch \( I_{\text{assoc}} \), the labelled diagram 1:[\_] says that \( p = a; b \) and label 2:[\_] says that \( r = p; c \). Sketch \( E_{\text{assoc}} \) extends \( I_{\text{assoc}} \) with three new elements: arrow \( q \) and two labelled diagrams 3:[\_] and 4:[\_], which say that \( r = a; (b; c) \). Hence, if we require a precategory to be also an \( \text{assoc} \)-algebra, it would satisfy the implication \( r = (a; b); c \Rightarrow r = a; (b; c) \). In a similar way we can specify operation \( \text{assoc}' \) imposing implication \( r = a; (b; c) \Rightarrow r = (a; b); c \). Hence, a precategory which is also an \( \{\text{assoc}, \text{assoc}'\} \) algebra has its composition operation associative.

In a similar way, we can specify the left and right unitality laws for the identity operation. First, we need to extend our signature of predicate symbols with symbols \([=\text{Node}]\) and \([=\text{Arr}]\) for equality of nodes and arrows: the arity of symbol \([=\text{Node}]\) is a pair of nodes and the arity of \([=\text{Arr}]\) is a pair of arrows with
different sources and different targets; in addition, we have a dependency: equality of arrows implies equality of nodes. Obviously, any graph is automatically a sketch in this equality signature (this is a diagrammatic version of logic with equality). However, in our sketches below we will not use \([=_{\text{Node}}]\) as equality of the corresponding nodes is automatic in the diagram.

The diagram in Fig. 7 specifies the required law. In sketch \(I_{\text{leftUnit}}\), label \([\text{id}]\) declares arrow \(a\) to be the identity of node \(x\), and label \([\cdot]\) says that \(r = a; b\). Sketch \(E_{\text{leftUnit}}\) extends \(I_{\text{leftUnit}}\) with one more labelled diagram that declares arrows \(b\) and \(r\) to be equal. In a similar way we can specify operation \(\text{rightUnit}\). Hence, a precategory which is also \(\{\text{assoc, assoc}', \text{leftUnit, rightUnit}\}\)-algebra is a category.

4 Advanced Diagram Logic: Predicate Symbols with Dependencies

4.1 Generalized Graphs

As our discussion in Sect.2 suggests, we need a very general notion of graphs as presheaves \(G: \mathcal{G} \to \text{Set}\) for a suitable schema category \(\mathcal{G}\). In all our applications, graphs have hierarchical structure that allows their traversal from the top elements to the bottom. For example, ordinary directed multigraphs are given by
Hierarchical categories

Definition 4 (Hierarchical categories). We consider a natural number $N$ as a category

$$[0, N) = \{0 \to 1 \to \ldots \to (N-1)\}$$

consisting of $N$ objects and $N-1$ non-identity arrows.

A hierarchical category (h-category in brief) of height $N$ is given by a small category $H$ along with a surjective on objects functor $\text{hei}: H \to [0, N)$ so that all layers $L_n(H) = \text{hei}^{-1}(n)$ for $n < N$ are not empty. We refer to $N$ by $\text{Hei}(H)$. Note that if $N = 0$, then category $[0, N)$ is empty and we have an empty hierarchical category of height 0. Thus, $H = \varnothing$ iff $\text{Hei}(H) = 0$.

Moreover, the following acyclicity condition holds for any two objects $x, y$ in $H$: if $H(x, y) \neq \emptyset$, then either $\text{hei}(x) > \text{hei}(y)$ or $x = y$, but in the latter case we require $H(x, x) = \{\text{id}_x\}$. Thus, all $H$-arrows go down from higher to lower layers, and the only arrows within a layer are identities. For an h-category $H$, we call elements of layer $L_n(H)$ n-cell types or n-(hyper)edge types and typically denote them by $H$. We will also call 0-cell types node types.

Corollary 2. If $H$ is an h-category, then for any $n < N$, the full subcategory $H_n \subset H$ generated by objects of height less than $n$ is a h-category as well. We have a cumulative chain of h-categories

$$H_0 \subset H_1 \subset \ldots \subset H_{N-1} = H$$

and for any object $H \in H$, $\text{hei}(H) = \min \{n \mid H \in H_n\}$.

Definition 5 (Discrete h-categories). A hierarchical category is called discrete if its only arrows are identities. However, a discrete h-category is more than a set: it is a set equipped with a height function, i.e., a layered set $H = \bigcup_{n < N} L_n(H)$. We will also write $L_{\leq n}(H)$ for the set $\bigcup_{0 \leq i \leq n} L_i(H)$.

Definition 6 (Finitary hierarchical categories). A h-category is called finitary if for any object $x$ the set $H(x, _)$ is finite.

Corollary 3 (Finite h-categories). A finitary h-category $H$ is finite as soon as its object set $\text{Ob}(H)$ is finite.

Remark 1 (Finitarity conditions). Def. 6 is the first in the family of similar definitions, which require finiteness of some construct assigned to the notion at hand. If an h-category $H$ is finite, then it is finitary, but the converse is obviously not true, and we will often deal with infinite but finitary objects (signatures, sketches, graphs).

3 The reason we start counting at 0 rather than 1 is to make the generalization for the first infinite ordinal $\omega$ straightforward: category $\varnothing$ has all natural numbers $n < \omega$ as objects.
Definition 7 (Graphs). Let \( \mathcal{G} \) be an h-category. An \( \mathcal{G} \)-graph is a functor \( G : \mathcal{G} \to \text{Set} \), i.e., a presheaf over schema \( \mathcal{G} \). We will normally refer to such functors as merely graphs, but if we want to stress their distinction from ordinary directed multigraphs (which are functors \( G : \text{Graph} \to \text{Set} \) over schema specified in (1)), we will refer to them as \( \mathcal{G} \)-graphs or, sometimes, as ‘graphs’ in single quotes. For a graph \( G : \mathcal{G} \to \text{Set} \), elements of set \( G(H) \) with \( H \in \mathcal{G}_n \) are called \( n \)-cells (of type \( H \)), and if \( \mathcal{G}_n \) is not a singleton, we say that \( n \)-cells are multi-sorted.

We write \( \text{Elem}_p G_q \) for the set \( \text{Ob} p \mathcal{G} q \). The presheaf topos of all \( \mathcal{G} \)-graphs will be denoted by \( \mathcal{G} \), and we assume that throughout the paper the schema \( \mathcal{G} \) is arbitrary but fixed. A graph \( \mathcal{G} \) is finite if \( \text{Elem}_p G_q \) is finite. All finite graphs form a (sub)category (of \( \mathcal{G} \)) denoted by \( \mathcal{G}_\omega \).

We will try to depict ‘graphs’ as graphical objects. For example, if object \( H \in \mathcal{G} \) has exactly two non-looping outgoing arrows, say, \( x, y \), we will call them source and target, and depict each element \( a \in G(H) \) as an arrow from element \( a.G_p x \) to element \( a.G_p y \).

Example 4 (Fancy Graph). Assume \( \mathcal{G} \) is the category in Fig. 8(a) (identities omitted); and Fig. 8(b) depicts a graph given by the following functor \( G : \mathcal{G} \to \text{Set} \):

\[
\begin{align*}
G(\text{Node}) &= \{P, Q, A, B\} \\
G(\text{Edge}) &= \{f, g, i\} \\
G(2-\text{Edge}) &= \{\alpha\} \\
G(\text{Edge}_{(0,4)}) &= \{n\} \\
G(\text{src}) &= \{f \to A, g \to A, i \to Q\} \\
G(\text{ trg}) &= \{f \to B, g \to B, i \to Q\}
\end{align*}
\]

Elements of \( G(0) \) can be called nodes, elements of \( G(1) \) and \( G(2) \) can be visualized as arrows (since the respective object in \( \mathcal{G} \) has exactly two outgoing non-identity morphisms. There is an edge between edges, \( \alpha \), and a hyperege \( n \) connecting four elements ("sources" of the hyperegde) as shown (as the respective object in \( \mathcal{G} \) has more than two outgoing non-identity morphisms). Lines connecting a hyperegde to its sources will be called tentacles.

4.2 Signatures and Sketches

Classic Ehresmann’s sketches consider signatures that only contain classes of limit and colimit predicates whose semantics is given by the respective universal properties. Generalized sketches consider arbitrary predicates, whose semantics is defined externally. This allows for more compact specifications because one does not need to add auxiliary structures as it is often required with classical sketches, see, e.g., [JR02] and discussion in [DW07].

Definition 8 (Predicate Signature). Let \( \mathcal{G} \) be a schema category for graphs and \( \mathcal{G} = \text{Set}^{\mathcal{G}} \) is the corresponding presheaf topos. A predicate signature \( \mathcal{P} \) over \( \mathcal{G} \) is given by a h-category \( |\mathcal{P}| \), whose objects are called predicate symbols and arrows are dependencies, and an arity functor \( \alpha : |\mathcal{P}| \to \mathcal{G}^{op} \). Thus, if a symbol
Fig. 8: Example of a higher-order graph

$P$ depends on symbol $P'$ via dependency $p: P \to P'$, then there is a graph morphism $p^\alpha: P^\alpha \to P'^\alpha$, where we write $X^\alpha$ for $\alpha(X)$ for an $X \in \text{Ob}(\mathcal{P})$ but we will use the bracketed notation too. For a predicate symbol $P \in \mathcal{P}$, we say that graph $P^\alpha$ is the arity of $P$. We will often (inaccurately) use symbol $P$ for both a signature and its carrier category $\mathcal{P}$.

A symbol $P$ is called finitary if its arity graph $P^\alpha$ is a finite graph, and a signature is finitary if all its symbols are such so that $\mathcal{G}^{op}$ can be considered as the codomain of functor $\alpha$. Below we assume finitarity as a default assumption.

Fig. 9: A sample signature $\mathcal{P}_{\text{pbm}}$
Example 5 (Pullbacks over monics). Consider ‘graphs’ being ordinary graphs, and a predicate signature with three symbols, $\text{Ob}(\mathcal{P}) = \{\text{monic}, \text{pb}_m, \text{=sqr}\}$, whose arities are as expected and shown in Fig. 9(b,c,d) resp. We assume these arities to be concrete fixed graphs whose elements are given fixed names formed by numeric strings (but we could also used colours or whatever names we like). We assume that predicate $\text{pb}_m$ can only be declared for squares in which arrows 10 and 32 are declared to be $\text{monic}$. Later we will describe $\text{pb}_m$ as an operation for which the requirement for arrow 10 to be monic will be a precondition while a similar requirement for arrow 32 will be a postcondition, but when $\text{pb}_m$ is a predicate, both requirements are “postconditions”. Another postcondition is the requirement for the entire square to be $\text{sqr}$ (read “commutative”). Then we partition $\mathcal{P}$ into two layers, $|\mathcal{P}|_0 = \{\text{monic}, \text{=sqr}\}$ and $|\mathcal{P}|_1 = \{\text{pb}_m\}$ connected by dependency arrows $p_1, p_2, p$ as shown in diagram (a). This gives us an h-category $\mathcal{P}$ and a functor $\alpha: \mathcal{P} \to \text{G}^{\text{op}}$ with $p_1\alpha(10) = 10$, $p_2\alpha(10) = 32$, and $p\alpha = \text{id}$ (or isomorphism if these squares are different).

We need predicate symbols to declare atomic constraints, e.g., $P(x, y)$ in ordinary first-order logic, and simplest logical theories are built from conjunctions of atomic formulas, e.g., $P(x, y) \land P(y, z)$. A diagrammatic counterpart of conjunctive theories is the notion of a sketch: a set of atomic formulas over a graph of variables, which is closed w.r.t. all dependency arrows in the signature.

Definition 9 (Sketch over a signature $\mathcal{P}$ with dependencies). A (generalized) sketch $S$ over $\mathcal{P}$ is given by a graph $|S| \in \mathcal{G}$ called the carrier, and a span of two functors $\text{label}$ and $\text{diagr}$ making a commutative outer square in diagram (7). The apex of the span is a small category $\text{Index}(S)$ of (diagram) indexes and dependencies; moreover, functor $\text{label}: \text{Index}(S) \to |\mathcal{P}|$ is required to be a discrete opfibration so that all arrows in $\text{Index}(S)$ are lifts of $\mathcal{P}$-arrows—this is the closure condition mentioned above (it is discussed in detail in [DW07]).

\[
\begin{array}{ccc}
\text{Index}(S) & \xleftarrow{\text{diagr}} & \mathcal{G}^{\text{op}}/|S| \\
\xrightarrow{\text{label}} & & \downarrow \alpha^* \\
\text{LDiagr}(|S|) & \xrightarrow{\alpha'} & \mathcal{G}^{\text{op}} \\
\downarrow & & \downarrow \\
|\mathcal{P}| & \xrightarrow{\alpha} & \mathcal{G}^{\text{op}}_{|S|}
\end{array}
\]

We call the definition above fibrational to distinguish it from a typical indexed sketch definition, in which the discrete opfibration $\text{label}$ is replaced by presheaf $\mathbb{J} \cdot |\mathcal{P}| \to \text{Set}$ and $\text{diagr}$ is a functor $\mathbb{J} \cdot |\mathcal{P}| \to \mathcal{G}^{\text{op}}$ (this is how sketches were defined in [DW07] assuming also that $\text{diagr}$ is injective).

Corollary 4 (Sketches are fibration morphisms). As $\text{dom}^{\text{op}}_{|S|}$ is a discrete fibration (see Lemma 1), commutativity of the outer square in diagram (7) makes the pair $(\alpha, \text{diagr})$ an opfibration morphism. This is an exact formulation of the major requirements for sketches with dependencies: the set of diagrams is to be closed under “inference rules” arising from dependencies.

Definition 10 ((Multi)sketches vs. mono-sketches). Let the internal square in diagram (7) be a pullback in $\text{Cat}$, then its apex is the category whose objects
are all \( \mathcal{P} \)-labelled diagrams over graph \(|S|\) and arrows are lifts of dependency arrows in \( \mathcal{P} \) (recall that pulling back a split fibration results in a split fibration). Objects of \( \text{Index}(S) \) can be identified with indexed pairs \((P, d)\) with \( P \) a predicate symbol and \( d: P^o \to |S| \) a diagram. In the MDE parlance, pairs \((P, d)\) would be called constraints, and then objects in \( \text{Index}(S) \) are constraint indexes or indexed constraints. The diagonal mapping \# is provided by the universality of pullbacks.

A sketch is called **mono-sketch** if the pair of functors \((\text{label}, \text{diagr})\) is jointly monic and hence mapping \# is monic.

**Definition 11 (Finite Sketches).** A sketch \( S \) is **carrier-wise finite** if the graph \(|S|\) is finite, and **constraint-wise finite** if the set \( \text{Index}(S) \) is finite. A sketch is **finite** if both finiteness conditions hold.

**Definition 12 (Sketch morphisms).** Let \( S = (|S|, \text{label}, \text{diagr}) \) and \( S' = (|S'|, \text{label}', \text{diagr}') \) be two sketches. A **sketch morphism** \( f: S \to S' \) is a pair consisting of an \( \mathcal{G} \)-arrow \(|f|: |S| \to |S'|\) and a functor \( f_{\text{const}}: \text{Index}(S) \to \text{Index}(S') \) such that \( f_{\text{const}} \circ \text{label}' = \text{label} \) and \( f_{\text{ind}} \circ \text{diagr}' = \text{diagr} \circ (\mathcal{G}/|f|) \) with \( \mathcal{G}/|f|: \mathcal{G}/|M| \to \mathcal{G}/|M'| \) defined by postcomposition with \(|f|\). This gives us the category of \( \mathcal{P} \)-sketches denoted by \( \mathcal{G}/\mathcal{P} \) (this is Makkai’s notation that turned out convenient).

![Diagram](image)

Note that if the target sketch is a mono sketch, then the second component of a sketch morphism becomes redundant as it can be derived from \(|f|: |S| \to |S'|\): given \( f \in \text{Index}(S) \) with \( P = \text{label}(I) \) and \( d = \text{diagr}(I): P^o \to |S| \), we define \( f_{\text{ind}}(I) \) to be the index over \(|S'|\) determined by the pair \((P, d \circ |f|)\).

### 4.3 Makkai’s Construction: Sketches as Graphs

A fundamental and not obvious property of the sketch construct (first noticed by Makkai for the case of a discrete signature without dependencies [Mak97]) is that multisffes are, in fact, presheaves — Makkai justified the notion of a multisketch by exactly this reason. However, as we have seen above, the notion of a multisketch has quite practical reasons to be considered: mathematical models often turn out smarter than their creators could foresee.

In this section we will extend Makkai’s theorem for the case of predicate signatures with dependencies, whose proof, unexpectedly, turned out to be much more complex. We begin with a simple example to illustrate the idea and then give a general proof.

**Example** Assume signature \( \mathcal{P}_{\text{pb}} = \{[\text{monic}], [\text{pb}], [\text{eq}],[\text{eq}]) \) described in Example 5, which is defined over h-graphs with schema Graph described in Fig. ??(a)
with binding maps, e.g., we require that dependant diagrams are obtained by postcomposition of inter-arity maps (Fig. 9), and we require commutativity of the corresponding triangles to ensure (see Fig. 9(c)).

Fig. 10: Collage of $\mathcal{P}_{pb_m}$ and Graph: elements of $\mathcal{G}$ and $\mathcal{P}$ are, resp., blue and orange, binding (wavy) arrows are black

on p. ???. We want to build a metamodel (metasketch, schema) for $\mathcal{P}_{pb_m}$-sketches, that is, a category $\mathcal{S}(\mathcal{P}_{pb_m})$ such that any $\mathcal{P}_{pb_m}$-sketch could be seen as a functor $S: \mathcal{S}(\mathcal{P}_{pb_m}) \to \textbf{Set}$.

The carrier of sketches are specified by schema Graph (the top two nodes and two vertical arrows between them in Fig. 10) so that category Graph is to be included into $\mathcal{S}(\mathcal{P}_{pb_m})$. We enrich this schema with three new nodes [monic], [pb], and [=sqr] as shown in Fig. 10, which in a $\mathcal{P}_{pb_m}$-sketch $S$ will be instantiated by the corresponding constraints. To ensure that elements of sets $S(x), x \in \{\text{monic}, \text{pb}, [sqr]\}$ are indeed diagrams of the corresponding shapes, we add to constraint nodes suitable binding arrows, e.g., arrows $\overline{01}, \overline{1}: \text{monic} \to \text{Node}$ ensure that any constraint labelled [monic] will refer to the two nodes in the carrier graph $|S|$ corresponding to the two nodes in the arity graph $\text{monic}^{10}$ (see Fig. 9(b)), and arrow $\overline{01}: \text{monic} \to \text{Edge}$ ensures that any [monic]-constraint in $S$ will refer to an arrow in the carrier graph corresponding to arrow $01$ in the arity graph. Moreover, we require $\overline{01} \circ \text{src} = \overline{0}$ and $\overline{01} \circ \text{trg} = \overline{1}$ so that any functor $S$ necessarily respecting these constraint would make any element in set $S(\{\text{monic}\})$ a correct diagram $b_S: [\text{monic}]^{10} \to |S|$.

We repeat this trick for two other constraint nodes [pb] and [=sqr] as shown in the figure. To simplify drawing, we use the following visualization: if an arrow is labelled by a set, it actually abbreviates a set of parallel arrows — one arrow per label, e.g., the vertical arrow from [pb] to Edge abbreviates four arrows whose names are over-lined names of the four arrows in the arity graph [pb]$^{10}$ (see Fig. 9(c)).

Arrows $p_1, p_2, p$ are the three dependency arrows in the signature $\mathcal{P}_{pb_m}$ (see Fig. 9), and we require commutativity of the corresponding triangles to ensure that dependant diagrams are obtained by postcomposition of inter-arity maps with with binding maps, e.g., we require $p_1 \circ \overline{01} = \overline{10}$ and $p_2 \circ \overline{01} = \overline{32}$. This gives us a category $\mathcal{S}(\mathcal{P}_{pb_m})$ consisting of three parts: a) category Graph is embedded into $\mathcal{S}(\mathcal{P}_{pb_m})$, b) category $\mathcal{P}_{pb_m}$ is embedded, and c) a set of intermediate bind-
ing arrows from \( \mathcal{P}_{\text{pb}} \)-objects to \( \mathcal{G} \)-objects such that some commutativity conditions hold.

(\text{ignore node [\text{constr}]} \text{ and four adjoint arrows for a moment}).

It is easy to see that any functor \( \hat{S}: \mathcal{S}(\mathcal{P}_{\text{pb}}) \to \mathcal{S} \) is actually a \( \mathcal{P}_{\text{pb}} \)-sketch \( \hat{\mathcal{S}} \). Indeed, the union \( \hat{\mathcal{S}}([\text{constr}]) = \bigcup_{x \in \text{Ob}(\mathcal{P}_{\text{pb}})} \hat{S}(x) \) gives us the sketch’s set of constraints \( \text{Index}(\hat{\mathcal{S}}) \) with an obvious mapping \( \text{label}: \text{Index}(\hat{\mathcal{S}}) \to \text{Ob}(\mathcal{P}_{\text{pb}}) \).

Moreover, functions \( \hat{S}(x), x \in \{p_1, p_2, p\} \) give us discrete opcartesian lifts of the three dependencies in \( \mathcal{P}_{\text{pb}} \), so that \( \text{label} \) becomes a discrete opfibration. Finally, binding arrows attached to a node \( X \in \{ \text{monic}, [\text{pb}], [\text{=sqr}] \} \subset \mathcal{S}(\mathcal{P}_{\text{pb}}) \) along with commutativity conditions ensures a correct binding mapping for every constraint \( x \in \hat{\mathcal{S}}(X) \), and as \( \hat{\mathcal{S}}([\text{constr}]) \) is a coproduct, we have a correct binding functor \( \text{diag}: \text{Index}(\hat{\mathcal{S}}) \to \mathcal{G}/[\hat{\mathcal{S}}] \).

It is also easy to see that the converse is true as well: any \( \mathcal{P}_{\text{pb}} \)-sketch \( \mathcal{S} \) gives rise to a functor \( \hat{S}: \mathcal{S}(\mathcal{P}_{\text{pb}}) \to \mathcal{S} \) such that the two transformations are mutually inverse (below we will prove it for the general case).

**Signatures as profunctors and their collages** Given a schema \( \mathcal{G} \) for graphs, a predicate signature over \( \mathcal{G} \) is a functor \( \alpha: \mathcal{P} \to (\mathcal{S}^\text{op})^\text{op} \), whose opposite is \( \alpha^\text{op}: (\mathcal{P}^\text{op}) \to \mathcal{S}^\text{op} \). Currying gives us a profunctor

\[
\alpha: \mathcal{G} \times \mathcal{P}^\text{op} \to \mathcal{S}.
\]

which maps a pair of objects \( H \in \mathcal{G}, P \in \mathcal{P} \) to the set

\[
\alpha(H, P) = \{ a \in \mathcal{P}^\alpha(H) \}
\]

where writing \( \alpha \) is a notational trick to show the change of the role of element \( a \) from being an element of an arity shape to being a binding arrow \( \overline{a}: H \leftarrow P^4 \); note that arrows go form the second component of the pair to the first one). In the opposite direction, we write \( a \in \mathcal{P}^\alpha(H) \) for a given binding arrow \( a \in \alpha(H, P) \), thus,

\[
\mathcal{P}^\alpha(H) = \{ a \in \alpha(H, P) \}
\]

and \( a = a \) for any \( a \in \alpha(H, P) \). It is known that any profunctor is equivalent to a cospan of categories

\[
\mathcal{G} \overset{\mathcal{G}^\bullet a}{\twoheadleftarrow} \mathcal{G}^\bullet \mathcal{P} \overset{\mathcal{P}}{\twoheadleftarrow} \mathcal{P}
\]

whose apex is the collage category of the profunctor and legs are full embeddings [?][Benabou; sc Add citation: 1) Distributors at Work by Jean Bénabou June 2000 2) Joyal’s CatLab Distributors and barrels. The collage category is defined as follows. \( \text{Ob}(\mathcal{G}^\bullet \mathcal{P}) = \text{Ob}(\mathcal{G}) \sqcup \text{Ob}(\mathcal{P}) \), and for a pair of objects \( X, Y \in \text{Ob}(\mathcal{G}^\bullet \mathcal{P}) \), we set

\[
\mathcal{G}^\bullet \mathcal{P}(Y, X) = \begin{cases}
\mathcal{P}(X, Y) & \text{if } X, Y \in \mathcal{P} \\
\mathcal{G}(X, Y) & \text{if } X, Y \in \mathcal{G} \\
\alpha(Y, X) & \text{if } Y \in \mathcal{G} \text{ and } X \in \mathcal{P} \\
\emptyset & \text{if } Y \in \mathcal{P} \text{ and } X \in \mathcal{G}
\end{cases}
\]

Note that for category \( \mathcal{G}^\bullet \mathcal{P} \), the usual notation for hom-sets is reversed and formula \( \mathcal{G}^\bullet \mathcal{P}(\_ , \_ ) \) denotes the set of arrows from \( P \).

---

\[\text{with a more formal attitude, we can consider } \alpha \text{ as a pair } (a, \ast) \text{ with } \ast \text{ being some predefined token}\]
Arrow composition is given by functions \( \alpha(h, \text{id}) \) and \( \alpha(id, p) \) as shown in the inset figure, and its associativity is ensured by the functoriality of \( \alpha \). Arrows in \( \alpha(X, Y) \) are sometimes called formal or wavy; in our context, they are exactly binding arrows we considered in the example above.

It is easy to see that the collage category above is an instance of the following construct.

**Definition 13.** A split h-category is a pair \((\mathbb{H}, m)\) with \(\mathbb{H}\) a hierarchical category of height \(N \geq 2\) and \(m \in (0, N)\) a natural number. Objects of height less than \(m\) (there is at least one) are called cell types or graphical; they and arrows between them form a full category \(\mathbb{H}_{[0, m]} \subset \mathbb{H}\). Objects of height more or equal to \(m\) (there is at least one) are predicate symbols (logical elements); they and arrows between them form a full subcategory \(\mathbb{H}_{[m, N]} \subset \mathbb{H}\). Arrows from predicate symbols to cell types are called wavy.

**Proposition 2.** The constructs of a predicate signature and a split h-category are naturally equivalent.

**Proof.** Let \((\mathbb{H}, \text{hei}, m)\) is a split h-category of height \(N\). The intermediate \(m\) generates a functor \(m: [0, N - 1] \rightarrow [0, 1]\) whose composition with hei gives us a barrel \(b = \text{hei} \cdot m: \mathbb{H}^{\text{op}} \rightarrow [0, 1]\) and hence a profunctor \(b: \mathbb{H}_{[0, m]} \times \mathbb{H}_{(m, N)} \rightarrow \text{Set}\) (see Joyal [5]); uncurrying the latter gives us a signature \(\alpha_m: \mathbb{H}^{\text{op}}_{(m, N)} \rightarrow \text{Set}^{\mathbb{H}_{[0, m]}}\). The opposite transformation was considered above. Natural isomorphism of exponentiation and currying, and of profunctors and barrels, give us required natural equivalence. Thus, we can consider any split h-category as the collage \(\mathcal{G} = \mathbb{H}_{[0, m]}\) and a signature \(\mathcal{P} = \mathbb{H}_{[m, N]}\) with arity \(\alpha_m\) defined as above.

This results is the first step in our general program of showing that distinction between graphs and theories over them is conventional. The second (and final) step is given by the following result.

**Theorem 1 (Extended Makkai’s Theorem).** The category of \(\mathcal{P}\)-sketches \(\mathcal{G}|\mathcal{P}\) is equivalent to the presheaf topos \(\text{Set}^{\mathcal{G} \ast \mathcal{P}}\). In other words, \(\mathcal{P}\)-sketches over \(\mathcal{G}\)-graphs can be interpreted as \((\mathcal{G} \ast \mathcal{P})\)-graphs (and thus diagrammatic constraints are just higher-order cells).

**Proof (Sketch)** We will show equivalence by building two mutually inverse functors.

From presheaves to sketches. Having a presheaf (graph) \(G \in \text{Set}^{\mathcal{G} \ast \mathcal{P}}\), we build a sketch \(\mathcal{S}(G)\) with

(a) \(|\mathcal{S}(G)| = G|_{\mathcal{G}} = \iota_{\mathcal{G}} \circ \mathcal{G}\), and obviously we have a naturally bijective correspondence between these parts of presheaf on \(\mathcal{G} \ast \mathcal{P}\) and \(\mathcal{G}|\mathcal{P}\)-sketch structures.

(b) We have a presheaf (indexed set/family) \(\iota_{\mathcal{G}} \circ \mathcal{S}(G) = G|_\mathcal{G} = \iota_{\mathcal{G}} \circ \mathcal{G}\), and define label\(_{\mathcal{S}(G)}\) to be the respective discrete opfibration (see, e.g., Emily Riehl’s survey).

(c) it remains to build the diagram functor \(\text{diag}_{\mathcal{S}(G)}: \iota_{\mathcal{G}} \circ \mathcal{S}(G) \rightarrow \mathcal{G}|\mathcal{S}(G)|\) that makes a commutative square as needed, and show that the construction provides a natural isomorphism between presheaves and sketches. This is the most difficult part, but it’s managed by the Yoneda lemma. The latter states
that for any predicate symbol \( P \), we have an isomorphism natural wrt. both \( P \) and \( G \):

\[
G(P) \cong_{P,G} \text{Set}^{\mathcal{G} \ast \alpha P}(\mathcal{G} \ast \alpha \mathcal{P}(\_, P), G)
\]  

(14)

(recall that for category \( \mathcal{G} \ast \alpha \mathcal{P} \), the usual notation for hom-sets is reversed and formula \( \mathcal{G} \ast \alpha \mathcal{P}(X, P) \) denotes the set of arrows from \( P \) to \( X \)). We need to unpack this statement in our context for our notation.

We begin with the following observation. For any predicate \( P \), the hom-functor \( \mathcal{G} \ast \alpha \mathcal{P}(\_, P) \) actually works as follows:

\[
\mathcal{G} \ast \alpha \mathcal{P}(x, P) = \begin{cases} 
\mathcal{P}(P, \_)(x) & \text{if } x \in \text{Elem}(\mathcal{G}) \\
\mathcal{P}(\_)(x) & \text{if } x \in \text{Elem}(\mathcal{P}) \\
\forall : \mathcal{P}(H) \leftarrow \{ \ast \} & \text{if } x : H \leftarrow P \text{ is a wavy arrow}
\end{cases}
\]  

(15)

(recall that \( \mathcal{P}(P, P) = \{ \text{id}_P \} \) for each \( P \) and thus any wavy arrow from \( P \) is an element in \( P^n \)).

Applying (15) to Yoneda (14) for \( p : P \rightarrow P' \) and restricting each of the natural transformations in the set to objects of subcategory \( \mathcal{G} \), we get the following family of commutative diagrams indexed by \( p \):

\[
\begin{aligned}
G(P) & \xrightarrow{\text{Yoneda}} \text{Set}^{\mathcal{G} \ast \alpha P}(\mathcal{G} \ast \alpha \mathcal{P}(\_, P), G) \xrightarrow{\text{diagr}} \text{Set}^G(P^n, G|_G) \\
G(P') & \xrightarrow{\text{Yoneda}} \text{Set}^{\mathcal{G} \ast \alpha P}(\mathcal{G} \ast \alpha \mathcal{P}(\_, P'), G) \xrightarrow{\text{diagr}} \text{Set}^G(P'^n, G|_G)
\end{aligned}
\]

(16)

whose right square gives us the sketch-structure view of diagrams extracted from the presheaf \( G \).

Applying (15) to Yoneda (14) for \( g : G \rightarrow G' \) and again restricting each of the natural transformations in the set to objects of subcategory \( \mathcal{G} \), we get another family of commutative diagrams indexed by \( g \):

\[
\begin{aligned}
G(P) & \xrightarrow{\text{Yoneda}} \text{Set}^{\mathcal{G} \ast \alpha P}(\mathcal{G} \ast \alpha \mathcal{P}(\_, P), G) \xrightarrow{g \circ \text{diagr}} \text{Set}^{\mathcal{G}'}(P^n, G'|_G) \\
G'(P) & \xrightarrow{\text{Yoneda}} \text{Set}^{\mathcal{G} \ast \alpha P}(\mathcal{G} \ast \alpha \mathcal{P}(\_, P'), G') \xrightarrow{g' \circ \text{diagr}} \text{Set}^{\mathcal{G}'}(P'^n, G'|_G)
\end{aligned}
\]

(17)

which shows that translation of presheaves to sketches is functorial.

From sketches to presheaves. Having a sketch \( S \in \mathcal{G} \parallel \mathcal{P} \), we build a presheaf (graph) \( G^S \) with

(a) \( G^S|_P = |S| \), and obviously \( |G^S| = |S| \).

(b) \( G^S|_P = \llbracket S, \text{label} \rrbracket \) is the presheaf corresponding to the discrete opfibration component of \( S \)—functor \( S, \text{label} \). As is well known, the correspondence between presheaves and discrete opfibrations is naturally bijective (it’s just a special discrete case of the Grothendieck construction).

(c) it remains to define functions \( G^S(p) \) for each wavy arrow \( p : H \leftarrow P \), and we set for any constraint index \( c \in \llbracket P \rrbracket \),

\[
G^S(c) = (S, \text{diagr})(c)_H(a).
\]
Recall that $\text{diagr}(c)$ (here and below we will omit attribution to $S$) is a diagram of shape $P^\alpha$, i.e., a natural transformation $\text{diagr}(c): P^\alpha \Rightarrow |S|$ of functors into $\text{Set}$, i.e., a family of functions $\text{diagr}(c)_H$ for each $H \in \mathcal{G}$, and $a$ is the element in $P^\alpha(H)$ corresponding to the wavy arrow $a \in \mathfrak{a}(P, H)$ (see (10)).

For further references, we summarize the above as shown below (we write $\text{label}$ for $\text{S.label}$ and $\text{diagr}$ for $\text{S.diagr}$):

$$G^S(x) = \begin{cases} |S|(x) & \text{if } x \in \text{Elem}(\mathcal{G}) \\ \llbracket x \rrbracket_{\text{label}} & \text{if } x \in \text{Elem}(\mathcal{P}) \\ \{ c \mapsto \text{diagr}(c)_H(x) : c \in \llbracket P \rrbracket \} & \text{if } x \in \mathfrak{a}(P, H) \end{cases} \quad (18)$$

We need to check the functoriality of so defined functions $G^S(a)$ wrt. compositions of wavy arrows with predicate and cell arrows, i.e., for any triple of arrows $p: P' \rightarrow P$, $a \in \mathfrak{a}(P, H)$ and $h: H \rightarrow H'$, compositions (i) $a \circ h$ and (ii) $p \circ a$ are preserved.

Case (i) is depicted in diagram (19)(a). As $\text{diagr}(c)$ is a graph morphism for any predicate $P$ and index $c \in \llbracket P \rrbracket$, diagram (19)(b) also commutes for any $H \in \mathcal{G}$:

$$\begin{array}{ccc}
P & \xrightarrow{\alpha} & H \\
\downarrow & & \downarrow \\
P' & \xrightarrow{a' = \mathfrak{a}(P, h)(a)} & H'
\end{array} \quad \begin{array}{ccc}
P^\alpha(H) & \xrightarrow{\text{diagr}(c)_H} & |S|(H) \\
\downarrow & & \downarrow \\
P^\alpha(H') & \xrightarrow{\text{diagr}(c)_H} & |S|(H')
\end{array} \quad (19)$$

Now we compute (we will write $\mathfrak{a}(P, h)$ for $\mathfrak{a}(\text{id}_P, h)$):

$$G^S(a \circ h)(c) = G^S(a')(c) \quad \text{by Def. of composition in } \mathcal{G}_* \mathcal{P}$$

$$= \text{diagr}(c)_{H'}(a') \quad \text{by Def. of } G^S (18), \text{3rd case}$$

$$= \text{diagr}(c)_{H'}(P^\alpha(h)(a)) \quad \text{by Def. of } \alpha$$

$$= |S|(h) \left( \text{diagr}(c)_H(a) \right) \quad \text{by commutativity (19)}$$

$$= G^S(h) \left( \text{diagr}(c)_H(a) \right) \quad \text{by Def. of } G^S (18), \text{1st case}$$

$$= G^S(h) \left( G^S(a)(c) \right) \quad \text{by Def. of } G^S (18), \text{3rd case}$$

$$= \left( G^S(a) \circ G^S(h) \right)(c)$$

Since the above holds for all $c \in \llbracket P \rrbracket_{\text{label}}$, we conclude that $G^S(a \circ h) = G^S(a) \circ G^S(h)$.

Case (ii) is in diagram (20)(a). For any $c' \in \llbracket P \rrbracket$, let $p: c' \rightarrow c$ be the lifting of arrow $p: P' \rightarrow P$ in $\mathcal{P}$ at $c' \in \text{Index}(S)$, i.e., $\llbracket p \rrbracket(c') = c$. As $\text{diagr}(p)$ is a diagram morphism, i.e., a commutative triangle, then diagram (20)(b) also commutes for any $H \in \mathcal{G}$:

$$\begin{array}{ccc}
P & \xrightarrow{\alpha} & H \\
\downarrow p & & \downarrow \\
P' & \xrightarrow{a' = \mathfrak{a}(p, H)(a)} & H \\
\downarrow p^\alpha & & \downarrow p^\alpha \\
P^\alpha(H) & \xrightarrow{\text{diagr}(c')_H} & |S|(H) \\
\downarrow & & \downarrow \\
P^\alpha(H) & \xrightarrow{\text{diagr}(c')_H} & |S|(H)
\end{array} \quad (20)$$
Now we compute:

\[
G^S(p \xrightarrow{a} c) = G^S(a')(c') \quad \text{by Def. of composition in } \mathcal{G}^\bullet \mathcal{P}
\]

\[
= \text{diagr}(c')_{H} (a') \quad \text{by Def. of } G^S \text{ (18), 3rd case}
\]

\[
= \text{diagr}(c')_{H} (p'_{H}(q)) \quad \text{by Def. of } \alpha
\]

\[
= \text{diagr}(c')_{H} (q) \quad \text{by commutativity (20)}
\]

\[
= G^S(a) (G^S(p)(c')) \quad \text{by Def. of } G^S \text{ (18), 2nd case}
\]

\[
= (G^S(p \xrightarrow{a} G^S(a)) (c')
\]

Since the above holds for all \(c' \in \mathbb{P}^\mathrm{lab}_{\mathrm{label}}\), we conclude that \(G^S(p \xrightarrow{a} G^S(p \xrightarrow{a})
\]

Thus, we have built a presheaf \(G^S \in \mathcal{S}^{\mathbb{S}_{\mathbb{P}}^\ast - \mathbb{P}}\) from sketch \(S \in \mathcal{G}||\mathbb{P}\) and checking functoriality of sketch-to-presheaf translation is straightforward (commutativity of diagram mapping with a sketch morphism is exactly naturality of presheaf morphism wrt. wavy arrows). It is easy to see that the two translations are exactly mutually inverse: \(S(G^S) \cong S\) and \(G^S(G) \cong G\). Yoneda lemma now implies that categories \(\mathcal{S}^{\mathbb{S}_{\mathbb{P}}^\ast - \mathbb{P}}\) and \(\mathcal{G}||\mathbb{P}\) are naturally isomorphic.

### 4.4 Discrete Hierarchical Signatures and Sketches

We will redefine the notions of a predicate signature with dependencies and sketches over it in a discrete (but hierarchical) way: it is somewhat unexpected that this partial de-categorification does not actually lose information and the semi-discrete (below) and the fully categorical (above) notions are equivalent. To distinguish the two families, we will refer to the definition of a signature with dependencies as functorial (or f-signature) and a corresponding notion of a sketch as fibrational (f-sketch). Their counterparts defined below are dh-signature and dh-sketch with prefix ‘dh’ referring to discrete hierarchical.

Also, in this section we will denote discrete signatures and categories of sketches over them (defined in Def. 8) by symbols \(\mathcal{P}_{\bullet}^\ast\) and \(\mathbb{S}||\mathcal{P}_{\bullet}^\ast\), where superscript \(\bullet\) is intended to recall that a signature carrier is a set rather than a category, and \(\mathbb{S}\) refers to the category of carriers which can be the category of graphs \(\mathbb{G}\), or some category of sketches defined before (recall that “sketches are graphs” by Makkai’s theorem). Finally, in this section we will begin counting layers of hierarchical constructs at 1 rather than 0 simply because it’s convenient to assign an empty signature of predicate symbols to index 0 rather than -1.

**Definition 14 (Dh-signatures and (dh-)sketches over them).** Let \(\mathcal{G}\) be a presheaf topos of \(\mathcal{G}\)-graphs as before. A predicate dh-signature of height \(N\) is given by the inductively defined data specified by the commutative diagram (21).
The upper row is provided by a discrete hierarchical category $\mathcal{P}^\bullet$ of predicate symbols of height $N$, i.e., a finite chain of non-empty sets shown in the top row of the diagram. Each of these sets $\mathcal{P}^\bullet_i$ is equipped with an arity function $\alpha^\bullet_i$ into the corresponding category of sketches $\mathcal{S}_{i-1} \overset{\text{def}}{=} \mathcal{S}_{i-2} || \mathcal{P}^\bullet_{i-1}$, $i < N$ as shown in the bottom row of the diagram. Note that this sequence starts from $\mathcal{G} = \mathcal{S}$, i.e., sketches with the empty signature, which allows us to defined mapping $\alpha^\bullet_0$, which in turn defines sketch category $\mathcal{S}_0$ and so on. Bottom arc arrows show the carriers. Diagonal dotted arrows labelled $\colon\text{sig}$ are pairs rather than mappings (and can be seen as instances of some big mapping $\text{sig}$ from the big category of all sketches to the big category of all signatures, which we don’t consider in the paper).

The above defines a dh-signature $\mathcal{P}^\bullet$. Sketches over $\mathcal{P}^\bullet$ and their morphisms are defined as objects and arrows of the category $\mathcal{S}_{N-1} \overset{\text{def}}{=} \mathcal{S}_{N-2} || \mathcal{P}^\bullet_{N-1}$. We will denote this category by $\mathcal{G}||\mathcal{P}^\bullet$: strictly speaking, we should write $\mathcal{G}||^\bullet\mathcal{P}^\bullet$ as it’s another construct of building sketches, but we will usually omit the superscript near $||$ if the right term unambiguously shows the dh-context of $||$.

**Theorem 2.** Any dh-signature $\mathcal{P}^\bullet$ gives rise to a functorial signature $[\mathcal{P}^\bullet]$, and any functorial signature $\mathcal{P}$ gives rise to a dh-signature $\mathcal{D}^\bullet\mathcal{P}$. Moreover, $\{\mathcal{D}^\bullet\mathcal{P}\} \cong \mathcal{P}$ and $\mathcal{D}^\bullet\{\mathcal{P}^\bullet\} \cong \mathcal{P}^\bullet$.

**Proof.** Let $\mathcal{P}^\bullet$ be a dh-signature as defined above. Makkai’s theorem allows us to rewrite $\alpha_2: \mathcal{S}_1 \hookrightarrow \mathcal{P}^\bullet_2$ as $\alpha_2: \text{Set}^{\mathcal{G}^\bullet_1 \times \mathcal{P}^\bullet_2} \hookrightarrow \mathcal{P}^\bullet_2$, then currying gives us a profunctor

$$\hat{\alpha}_2: \text{Set} \leftarrow (\mathcal{G} \circ \alpha_1, \mathcal{P}^\bullet_1) \times \mathcal{P}^\bullet_2$$

(we don’t write $\mathcal{P}^{\bullet opp}_2$ as $\mathcal{P}^\bullet_2$ is discrete), whose collage is the category

$$(\mathcal{G} \circ \alpha_1, \mathcal{P}^\bullet_1) \circ \alpha_2 \mathcal{P}^\bullet_2.$$

The latter, for any $P \in \mathcal{P}^\bullet_2$ and $P' \in \mathcal{P}^\bullet_1$, has a wavy arrow $\overline{\alpha}: P \rightarrow P'$ iff there is a diagram index $c \in P^{\alpha_2}(P')$ (i.e., an element in $[P']$ for sketch $P^{\alpha_2}$), and similarly, any wavy arrow $\overline{\alpha}: P \rightarrow H$ is generated by a cell $a \in P^{\alpha_2}(H)$ in the carrier of sketch $P^{\alpha_2}$.— we use overlining and underlining as earlier in definitions (10) and (11) (Example in Fig. 10 shows how it works). We continue for $i = 3...N$ and obtain the commutative diagram formed by the upper lane of squares in (22):

\[
\begin{array}{cccccc}
\emptyset & \rightarrow & \mathcal{P}^\bullet_1 & \rightarrow & \mathcal{P}^\bullet_2 & \rightarrow & \ldots & \rightarrow & \mathcal{P}^\bullet_N \\
\mathcal{G} & \rightarrow & \mathcal{G}\mathcal{P}^{\alpha_1}_1 & \rightarrow & \mathcal{G}\mathcal{P}^{\alpha_2}_2 & \rightarrow & \ldots & \rightarrow & \mathcal{G}\mathcal{P}^{\alpha_N}_N \\
\emptyset & \rightarrow & \mathcal{P}^\bullet_1 & \rightarrow & \mathcal{P}^\bullet_2 & \rightarrow & \ldots & \rightarrow & \mathcal{P}^\bullet_N
\end{array}
\]

In this lane, the lower line (starting at $\mathcal{G}$) is a chain of collage categories obtained by inductively adding binding/wavy arrows (taken from the arity sketches elements) layer by layer. (Example in Fig. 10 would illustrate the idea if we make all orange arrows black wavy arrows.) By removal of the subcategory $\mathcal{G}$ (and resp. all arrows from predicates to $\mathcal{G}$-objects) from each of the collages, we obtain the
bottom chain of categories $\mathcal{P}^*_i$. (With the colour legend of the example in Fig. 10, we remove all black arrows from predicate symbols to $\mathcal{G}$ but keep arrows between predicates and colour them in orange. Thus, the difference between set $\mathcal{P}^*_i$ and category $\mathcal{P}^*_i$ is the presence of (formerly wavy and black but now) orange arrows $\tau$ between predicate symbols specified above.) The arrows that left in categories $\mathcal{P}^*_i$ actually specify dependencies, and we can consider category $\mathcal{P}^*_N$ as the carrier of a f-signature with dependencies. To complete the story and make $\mathcal{P}^*_N$ a signature, we need to define the corresponding (integrated) arity functor $\mathfrak{f}_\alpha: \mathcal{P}^*_N \to \mathcal{G}^{\text{op}}$.

On objects, for any $i \leq N$ and $P \in \mathcal{P}^*_i$, we set $\mathfrak{f}_\alpha(P) \overset{\text{def}}{=} |\alpha_i(P)|$.

For arrows, given a binding/wavy arrow $\tau: P \to P'$ for $P' \in \mathcal{P}^*_{i-1}$, we define

$$\mathfrak{f}_\alpha(\tau) \overset{\text{def}}{=} \text{diagr}(\alpha_i(P))|_{\alpha_{i-1}(P')}: |\alpha_{i-1}(P')| \rightarrow |\alpha_{i}(P)|$$

Checking the functoriality of so defined $\mathfrak{f}_\alpha$ is straightforward, which makes the pair $(\mathcal{P}^*_N, \mathfrak{f}_\alpha)$ a correct f-signature. We will denote it by $\mathcal{P}^*$ (so that now the carrier $\mathcal{P}^*_N$ is also denoted by $\mathcal{P}^*$).

Now we consider the transition in the opposite direction. Let $\mathcal{P}$ be an f-signature of height $N$ with arity functor $\alpha: \mathcal{P} \to \mathcal{G}^{\text{op}}$. Recall that f-signatures are layered, layers are sets, and we define $\mathcal{P}^*_i = \mathcal{P}_i$ for all $1 \leq i \leq N$. For each $P \in \mathcal{P}^*_i$, we define sketch $\alpha_i(P)$ as follows. The carrier $|\alpha_i(P)| = \alpha(P)$.

Diagrams are given by dependency arrows: for any $P' \in \mathcal{P}^*_{i-1}$, we set $P^{\alpha_i}(P') = \{ [P] p \in \mathcal{P}(P, P') \}$ (recall that formally $p$ is just a couple $(p, \tau)$ of $p$ with any aside token $\tau$), and for any index $p \in [P]$, the diagram $\text{diagr}(p)$ is defined to be mapping $\alpha(p)$ (the construction is exactly the inverse to (24)). Thus, we have decomposed f-signature into a dh-signature $\mathcal{D}^*\mathcal{P}$, in which dependency arrows are hidden in diagrams of the arity sketches. Isomorphisms immediately follow from the constructions above.

**Corollary 5.** a) Given an f-signature (i.e., predicate signature with dependencies) $\mathcal{P}$, two categories of sketches, $\mathcal{G}|\mathcal{P}$ and $\mathcal{G}|\mathcal{D}^*\mathcal{P}$ are isomorphic. b) Given a dh-signature $\mathcal{P}^*$, two categories of sketches, $\mathcal{G}|\mathcal{P}^*$ and $\mathcal{G}|\mathcal{P}^*$, are isomorphic as well.\(^5\)

### 4.5 Some operations on sketch subobjects

**Definition 15 (Operations between subobject semi-lattices).** Let $\mathcal{P}$ be a predicate signature and $S \subseteq \mathcal{G}|\mathcal{P}$ a sketch. There is a subobject functor

$$\text{subske}_S : \text{Sub}(|S|) \rightarrow \text{Sub}(S)$$

which adds to a subgraph $G \subseteq |S|$ all $S$-constraint that can be declared on $G$. It is defined via the right adjoint $\forall (id, !)$ of the pullback functor $(id, !)^{-1}: \text{Sub}(S) \rightarrow \text{Sub}(|S|)$, see Remark ??. Note that $|S|$ can be interpreted as a sketch without diagrams, thus there is a canonical sketch morphism $(id, !): |S| \rightarrow S$ in $\mathcal{G}|\mathcal{P}$.

\(^5\) The bulky formulation of this theorem should be shorter and properly stated if we introduce categories of f-signatures and dh-signatures, the corresponding categories of sketches over them, and two functorial isomorphisms. We leave it for future work.
Furthermore, there is another functor
\[
\text{subsig}_P : \mathbb{G}||\mathcal{P} \to \text{Sub}(\mathcal{P})
\]
which maps a \(\mathcal{P}\)-sketch \(S\) to its effective predicate signature, i.e. those predicates from \(\mathcal{P}\) that are actually used in the sketch, \(\text{subsig}_P(S) \subseteq \mathcal{P}\). It is defined as the image of \(\text{label}\), i.e. \(\text{subsig}_P(S) = \text{img}(\text{label})\). Arrows in \(\text{Sub}(\mathcal{P})\) are inclusions and \(\text{subsig}_P\) is obviously a functor.

Finally, if \(\mathcal{P}'\) is a subsignature of \(\mathcal{P}\), i.e., \(|\mathcal{P}'| \subset |\mathcal{P}|\) in \(\text{Cat}\), and \(\alpha'|_{\mathcal{P}} = \alpha\), then we have a forgetful functor \(\_|_{\mathcal{P}'} : \mathbb{G}||\mathcal{P} \to \mathbb{G}||\mathcal{P}'\).

5 Basic Diagram Algebra: Operations without Dependencies

In this section we consider operation signatures and algebras for them when there are no dependencies between operations. Syntax and semantics of the parallel composition of queries (we say multiqueries) are described in, resp., Sect. 5.1 and 5.2, and sequential composition (complex queries) is in Sect. 5.3. In Sect. 5.4 we define views and their compositions, and show that they can be organized into an sm-category. Section 5.5 discusses how our query building mechanism is related to monads.

Our technical machinery in this section is basically a “pushout calculus”: we chase PO squares and wide POs over categories of sketches (which are presheaf toposes by the extended Makkai’s theorem). All these sketch categories are assumed to be skeletal by default, which makes POs over them deterministic operations. The skeletality assumption is motivated by our applications as discussed in Sect. 2.2. However, while applications need finitary operations, until Sect. 5.3 on complex queries, we consider the general case without the finitariness assumption as it does not bring any additional technical difficulties (and notionally it is even simpler to work with indexing in general, say, \(x_i, i \in I\), rather than with finite enumerations \(i = 1, 2, ..., n\)). Moreover, although complex queries are finite sequences of multiqueries, the latter are not assumed to be finitary until Sect. 5.5, where size restriction become important for building the query monad, and we prove that complex queries built from finitary multiqueries give rise to a finitary monad.

5.1 Multiqueries and Operations over them

Basic Notions. We now consider how we can build complex operations (queries) from elementary ones given in the signature. We first need the notion of an atomic query that makes a derived operation from a single operation, e.g., \(x + x\) makes a unary operation from a binary one (and later we will consider multi-queries).

**Definition 16 (Atomic Queries (= atomic terms)).**
Let \(\mathcal{O}\) be a discrete operation signature over a category of sketches \(\mathbb{S} = \mathbb{G}||\mathcal{P}\). An atomic query is a triple \(q = (I_q, \phi, b_q)\) of an input sketch \(I_q \in \mathbb{S}\), an operation symbol \(\phi \in \mathcal{O}\) (we omit the subscript \(q\)) and a binding sketch morphism \(b_q : I_\phi \to I_q\). These data determine pushout squares (25) in \(\mathbb{S}\), in which the resulting objects \(E_q\) is determined up to iso.
An atomic query is finitary if $\phi_q$ is finitary, i.e., sketch $E_{\phi}$ (and hence $I_{\phi}$) is finite.

**Remark 2 (Concrete Syntax).** By fixing a concrete syntax for specifying terms, we fix a particular object $E_q$, while the general pushout formulation gives us a concrete-syntax-free description. Consider, for example, a binary operation $\cdot$ in the ordinary universal algebra setting (Example 3) and an atomic query $q = (X, +, b_q)$ over a set of variables $X = \{x, y, z\}$ in the role of $I_q$ and a binding mapping $b_q : 2 \to X$, where $2 = \{0, 1\}$ and $b_q = \{0 \mapsto x, 1 \mapsto y\}$. The diagrammatic arity of $\cdot$ is an inclusion $IE_{\cdot} : \{0, 1\} \to \{0, 1, *\}$ with $*$ being an arbitrarily chosen token, and taking pushout of $(IE_{\cdot}, b_q)$ gives us a four-element set, e.g., $E_q = \{x, y, \emptyset, z\}$ with $\emptyset$ being a new element representing the application of $\cdot$ to $(x, y)$ (indeed, $b^E_q(*) = \emptyset$ irrespective of what tokens are chosen for $*$ and $\emptyset$). If elements of set $X$ are chosen to be strings (a typical case), then for the new element we may take a string $x + y$ (infix) or $+(x, y)$ (prefix) or $xy+$ (reversed polish notation) etc. depending on the respective choice of the concrete syntax.

Thus, choosing a concrete syntax for an atomic term $q$ amounts to choosing a concrete pushout square for the set of binding squares determined by $q$. However, as we will not use any special properties of a chosen pushout square besides its universality, our considerations are actually concrete-syntax-free, i.e., do not depend on the chosen concrete syntax. Thus, working with chosen POs would mean working with some chosen but unspecified concrete syntax (analogously to defining a function with an unspecified parameter). However, we work with an even stronger skeletality assumption, which makes our PO constructions deterministic and analogous to classical ASTs (abstract syntax trees) in the ordinary algebra.

An algebraic analog of the notion of sketch (a conjunctive theory) is the notion of multiquery—a "conjunction" of atomic queries executable in parallel.

**Definition 17 (Multiqueries).** A multiquery $Q$ over a discrete one-layer signature $O$ (over a category of sketches $S = G[P]$) is given by a sketch $I_Q \in S$ called the input, and a family of atomic queries with the same input sketch $I_Q$. This family is given by a span of two functions $\text{label}$ and $\text{diagr}$ making a commutative diagram (26b) in $\text{Set}$. The apex of the span is a set $\text{Index}(Q)$ of multiquery indexes, and for an index $q \in \text{Index}(Q)$, diagram $\text{diagr}(q)$ or $\text{diagr}^q$ is a pushout square in $S$ of shape shown in (26a) in which the value $\text{diagr}^q(I_a)$ is predefined to be $I_Q$. Moreover, $\text{diagr}^q(IE) = (\text{label}^q)^\alpha$ due to commutativity, and hence
5 Basic Diagram Algebra: Operations without Dependencies

the triple \((I_Q, label^q, diagr^q(II))\) is an atomic query.

\[
\begin{aligned}
\text{Index}(Q) & \quad \text{diagr} \quad \# \\
I & \xrightarrow{\text{IE}} E \\
I_I & \xrightarrow{\text{II}} EE \\
E & \xrightarrow{\text{EE}} E^*_I
\end{aligned}
\]

\[
\begin{aligned}
LDiagr(I_Q) & \xrightarrow{\alpha} \downarrow \begin{VertMatrix}
\vdots \quad \downarrow \quad \vdots \\
[po] \quad \downarrow \quad S
\end{VertMatrix} \\
\downarrow \text{(PB)} & \xrightarrow{\alpha} \downarrow \begin{VertMatrix}
\vdots \quad \downarrow \quad \vdots \\
[O] \quad \downarrow \quad S
\end{VertMatrix}
\end{aligned}
\]

a) Shape of diagr-diagrams

b) Multiquery Definition

(26)

The image of function \(label\), the set \(\{label(q)\mid q \in \text{Index}(Q)\}\) of all operation symbols involved in \(Q\) (together with their arities) will be denoted by \(O(Q) \subseteq \mathcal{O}\).

The internal square in (26b) is a pullback in \(\text{Set}\), and hence there is a unique mapping \(#\). A multiquery is monic, if the pair \((label, diagr)\) is jointly monic and hence \(#\) is monic. A multiquery \(Q\) is finitary if all operations \(label(q), q \in \text{Index}(Q)\) are finitary, and \(Q\) is finite if the indexing set \(\text{Index}(Q)\) is finite.

Example 6 (Running Example). In Fig. 3 on p. 6, the two applications \(a, b\) of the pullback operation (see two chevrons in diagram B) form a multi-query \(Q\) with \(\text{Index}(Q) = \{a, b\}\), \(\text{label}(a) = \text{label}(b) = [pb]\) and the following binding diagrams (see Fig. 9(C) on p.20 for naming the elements in the arity of \([pb]\)):

\[
\begin{aligned}
diagr(a)(II) & = \{10 \to x, 20 \to y\}; \\
diagr(a)(EE) & = \{10 \to x, 20 \to y, 32 \to x_y, 31 \to y_x\}, \\
diagr(b)(II) & = \{10 \to z, 20 \to y\}; \\
diagr(b)(EE) & = \{10 \to z, 20 \to y, 32 \to z_y, 31 \to y_z\},
\end{aligned}
\]

and the mappings for horizontal monics are easily restored. The result of applying this multiquery is a black-and-blue \([\text{monic}]\)-sketch in Fig. 3(B).

We can use multi-queries to extend our operation signature in the same way as we can define new methods in a programming language by combining built-in methods. For example, the multiquery above is an operation that takes diagram \(I = (x, y, z)\) as an input and completes it with arrows \(x_y, y_x, y_z, z_y\) as shown in the figure. On the other hand, multiqueries can be seen as sketches in an extended signature. For example, the multiquery above has two chevron labels of the square arity \(E_{[pb]}\) (the scope of operation \([pb]\)) and, implicitly, two labels of the cospan arity \(I_{[pb]}\) that describe the inputs to \([pb]\) applications. Moreover, there is a dependency between the corresponding operation symbols \(P^E_{[pb]}\) and \(P^I_{[pb]}\); as soon as a diagram labelled \(P^E_{[pb]}\) is found in a sketch, the corresponding diagram \(P^I_{[pb]}\) must be there too. The following definition makes it precise.

Example 7 (Empty multiqueries). For any sketch \(X\), there is a unique multiquery \(0_X\) with \(\text{Index}(0_X) = \emptyset\) and empty leg mappings too.
**Operations over Multiqueries.** We begin with an important construction of extended signature.

**Definition 18 (Extended Signature).** For a given predicate signature \((\mathcal{P}, \alpha_\mathcal{P})\) and an operation signature \((\mathcal{O}, \alpha_\mathcal{O})\) over \(\mathcal{P}\), their **collage signature** is a predicate signature \(\mathcal{P} \parallel \mathcal{O}_\mathcal{E} \parallel \mathcal{P}\) with \(\text{Ob}(\mathcal{P}) \parallel \mathcal{O}_\mathcal{E} = \text{Ob}(\mathcal{P}) \sqcup \mathcal{O} \) and \(\text{Arr}(\mathcal{P}) \parallel \mathcal{O}_\mathcal{E} = \text{Arr}(\mathcal{P})\) so that \(\mathcal{P} \subset \mathcal{P} \parallel \mathcal{O}_\mathcal{E}\); further, \(\alpha_{\mathcal{P}} \parallel \mathcal{O}_\mathcal{E} \parallel \mathcal{P}\) and the object part of \(\alpha_{\mathcal{P}}\) is provided by the following diagram:

\[
\begin{align*}
\mathcal{P} & \rightarrow \text{Ob}(\mathcal{P}) \rightarrow \text{Ob}(\mathcal{P}) \sqcup \mathcal{O} \leftarrow \mathcal{O} \\
\alpha_\mathcal{P} & \downarrow \downarrow \downarrow \text{Ob}(\alpha_\mathcal{P}) \downarrow \downarrow \downarrow \text{Ob}(\alpha_{\mathcal{P}}) \\
\mathcal{G} & \rightarrow \text{Ob}(\mathcal{G}) \rightarrow \text{Ob}(\mathcal{G}) \parallel \mathcal{P}_\mathcal{E} \leftarrow \text{Ob}(\mathcal{G} \parallel \mathcal{P}) \leftarrow \text{Ob}(\mathcal{G} \parallel \mathcal{P}) \parallel \mathcal{O}_\mathcal{E} \\
\end{align*}
\]

The sketch category \(\mathcal{G} \parallel (\alpha_{\mathcal{P}})\) will be denoted by \(\mathcal{G} \parallel \mathcal{P} \parallel \mathcal{O}_\mathcal{E}\) or \(\mathcal{S} \parallel \mathcal{O}_\mathcal{E}\). Any sketch \(\mathcal{S} \in \mathcal{G} \parallel \mathcal{P} \parallel \mathcal{O}_\mathcal{E}\) can be seen as a \(\mathcal{P}\)-sketch \(\mathcal{S} \parallel \mathcal{P}\) by forgetting its \(\mathcal{O}\)-diagrams.

**Lemma 1 (Multi-queries as operation symbols)** Any multi-query \(Q\) with input sketch \(X \in \mathcal{S} := \mathcal{G} \parallel \mathcal{P}\), gives rise to a \(\mathcal{P} \parallel \mathcal{O}_\mathcal{E}\)-sketch \(Q_{\mathcal{P}}\) and a monic \(\eta_Q : X \rightarrow Q_{\mathcal{P}}\) with \(\text{Arr}(\mathcal{G} \parallel \mathcal{P} \parallel \mathcal{O}_\mathcal{E})\). Restriction \(Q_{\mathcal{P}}\) will be denoted by \(E_Q\) and we thus also have a \(\mathcal{P}\)-sketch monic

\[
\{ IE_Q : X \rightarrow E_Q \} \in \text{Arr}(\mathcal{S})
\]

called the arity monic of \(Q\) (\(\eta_Q\) is also re-denoted as \(IE_Q\); the change in notation will become clear after Lemma 3, in which we show that each \(\mathcal{O}\)-algebra \(A\) is also a \(IE_Q\)-algebra).

**Proof.** We will first build monic \(IE_Q : X \rightarrow E_Q\) in \(\mathcal{S}\), and then will show \(E_Q\)’s enrichment with \(\mathcal{O}\)-diagrams. We take projection

\[
\left| IE_\mathcal{S} : [X \rightarrow , \mathcal{S}] \rightarrow \left[ \begin{smallmatrix} [\mathcal{P} \parallel \mathcal{E}] \parallel \mathcal{S} \\ [X \rightarrow , \mathcal{S}] \end{smallmatrix} \right] \right|
\]

and obtain

\[
\text{diagr}' \overset{\text{def}}{=} \text{diagr} \upharpoonright IE_\mathcal{S} : \text{Index}(Q) \rightarrow [X \rightarrow , \mathcal{S}]
\]

Currying gives us a wide span with apex \(X\)

\[
\text{ diagr}' : \text{Index}(Q) \times \{ X \rightarrow \} \rightarrow \mathcal{S}
\]

whose colimit (wide pushout) in \(\mathcal{S}\) gives us a monic \(IE_Q : X \rightarrow E_Q\) together with a family of monics \(IE_\mathcal{Q} : E^q \rightarrow E_Q \mid q \in \text{Index}(Q)\) as shown in the diagram below for the case of a countable \(\text{Index}(Q)\) (think of a countable pile of binding squares glued together at \(I_Q\), and the colimit cocone is shown with dashed arrows):
5 Basic Diagram Algebra: Operations without Dependencies

In addition, we have a family of diagrams indexed by $q \in \text{Index}(Q)$:

$$b^q_E \vdash EE_q^q : E_{\text{label}(q)} \to E_Q,$$

which makes $E_Q$ a $G\|P\|O_E$-sketch $Q(X)$ or just $Q(X)$.

**Corollary 6 (Empty Queries as Identities).** For any sketch $X$, the arity of the empty query over $X$ is identity: $\eta_X = \text{id}_X: X \to X$.

**Multiquery Merge**

**Definition 19 (Multiquery Morphisms).** Let $X$ be a sketch in $S=G\|P; and $Q, Q'$ are two multiqueries over $X$ (see Def. 17). A morphism $f: Q \to Q'$ is a function $f: \text{Index}(Q) \to \text{Index}(Q')$ such that $f \vdash \text{label}' = \text{label}$ and $f \vdash \text{diagr}' \vdash_{\text{diagr}} \vdash_{\text{diagr}}$ $\vdash_{\text{diagr}}$. Clearly, due to commutativity, we also have $f \vdash \text{diagr}' \vdash_{\text{diagr}} \vdash_{\text{diagr}}$ and hence, as pushouts are unique for sketch categories, $f \vdash \text{diagr}' = \text{diagr}$. Hence, multiquery morphism is nothing but a span morphism.

**Corollary 7.** A multiquery morphism $f: Q \to Q'$ is also a $G\|P\|O_E$-sketch morphism making the following diagram commutative:

$$
\begin{array}{c}
\begin{array}{c}
X \\
\eta_Q \\
\eta_{Q'}
\end{array}
\end{array}
\begin{array}{c}
\begin{array}{c}
\downarrow f \\
Q(X) \\
\downarrow \eta_Q \\
\downarrow \eta_{Q'}
\end{array}
\end{array}
\begin{array}{c}
\begin{array}{c}
\downarrow \eta_{Q'} \\
Q'(X)
\end{array}
\end{array}
$$

**Proposition 3.** For any $P$-sketch $X$, there is a (big) category of multiqueries over $X$, $MQ(X)$, equipped with an injective on objects and faithful functor $mq2sk$ as shown below:

$$MQ(X) \xrightarrow{mq2sk} G\|P\|O_E \xrightarrow{\text{colim}_P} G\|P.$$ 

We will also write $mQQ(X)$ for the class $\text{Ob}(MQ(X))$.
Proof. Obviously, multiquery morphism composition is (strictly) associative, and the empty multiquery \( 0_X \) is the unit. This gives us category \( \mathbb{MQ}(X) \). We will consider its (co)skeletal version, in which isomorphic multiqueries are identified. Injectivity and faithfulness of \( \text{mq2sk} \) follow from Def. 17 and 19. The functor is not full, e.g., there are sketch morphisms whose carrier graph part is an endomorphism but multiquery morphisms assume this part to be identity on \( X \).

**Proposition 4 (Multiquery Merge).** For any \( X \), category \( \mathbb{MQ}(X) \) has coproducts (denoted by \( +_{\mathbb{MQ}(X)} \) or \( +_X \) or just \( + \) if the context is clear). Moreover, pushout of span \( (\eta_Q, \eta_{Q'}) \) in \( \mathbb{G}||\mathbb{P}||\mathbb{O}_E \) (see diagram (29a)) produces \( (Q + Q') \) with legs being coprojections: \( \eta_{Q'} = \iota: Q \rightarrow Q + Q' \) and \( \eta_Q^* = \iota': Q' \rightarrow Q + Q' \) (note switching the primes in our notation).

**Proof.** Let \( Q \) and \( Q' \) be two multi-queries over \( X \). We construct a multi-query \( Q + Q' \) called the merge in the following way. We define the apex set

\[
\text{Index}(Q +_{\mathbb{MQ}(X)} Q') := \text{Index}(Q) \cup_{\text{Set}} \text{Index}(Q'),
\]

and then the legs label and diag are determined by the universal property of the coproduct in \( \text{Set} \); the required commutativity of the rectangle is straightforward. Checking the universality of \(+_{\text{MQ}(X)}\) is straightforward.

Now we prove that diagram (29a) in \( \mathbb{G}||\mathbb{P}||\mathbb{O}_E \) is a pushout. We use construction of Lemma 1 and note that the wide span in diagram (27) for \( +_{\mathbb{MQ}(X)} Q' \) is partitioned into two wide spans, and hence its colimit is given by the pushout diagram of subspans colimits, but the latter are specified by the two \( \eta \)-arrows. Diagram (b) shows the \( -|_\mathbb{P} \) projection of (a) (note that \( X \) is a \( \mathbb{P} \)-sketch).

**Multiquery Substitution** In universal algebra, we can freely exchange the variable names in a term algebra \( T(X) \). Thus every variable substitution \( \sigma: X \rightarrow Y \) freely induces a unique term-algebra substitution \( T(\sigma): T(X) \rightarrow T(Y) \). The same is true for our notion of multiqueries.

**Lemma 2 (Multiquery Substitution)** Any \( \mathbb{P} \)-sketch morphism \( f: X \rightarrow X' \) It gives rise to a functor \( f^{\mathbb{MQ}}: \mathbb{MQ}(X) \rightarrow \mathbb{MQ}(X') \). Moreover, for any m-query \( Q \in \mathbb{MQ}(X) \), we have a pushout square in \( \mathbb{G}||\mathbb{P}||\mathbb{O}_E \) shown in diagram (30)(b) below, where \( Q' \) denotes \( f^{\mathbb{MQ}}(Q) \) and \( f_\sigma^* \) is the lifting of \( f \) at \( Q \) (we will show later that it is indeed an opcartesian lifting). Diagram (c) is merely another notation for diagram (b), which may be useful in some cases, e.g., for stating compositionality of the construction accurately specified in diagram (d), but in terms of (c) taking the form \( Q(f \circ g) = Q(f) \circ Q(g) \). Diagram (a) is the \( -|_\mathbb{P} \) projection of (b), and (c) will be used in the proof.
Proof. We define multiquery query $Q^f = (I_{Q^f}, \text{label}^f, \text{diag}^f)$ as follows. The input sketch is $I_{Q^f} := Y$ and the indexing span is

$$\begin{array}{c}
\xymatrix{
X \ar[rr]^-{\eta_Q} \ar[dd]_-{f} & & Q(X) \ar[dd]^-{f} \\
Y \ar[rr]^-{\eta_{Q^f}} \ar[rrd]_-{g} & & Q^f(Y) \ar[dd]^-{f} \\
Z & & (Q^f)^2(Z) \ar[lld]_-{g_{(Q^f)}} \\
& & \eta_{Q^f(q)}(q)
}
\end{array}$$

(30)

\begin{align*}
&I_{label(q)} \xrightarrow{IE_{label(q)}} E_{label(q)} \\
&\xymatrix{
I_Q \ar[rr]^-{IE_Q} \ar[dd]_-{f} & & E_Q \ar[dd]^-{f} \\
Y \ar[rr]^-{IE_{Q^f}} \ar[rrd]_-{g} & & E_{Q^f} \ar[dd]^-{f} \\
& & \eta_{Q^f(q)}(q)
}
\end{align*}

(1) and (2)

where $\text{Index}(Q^f) = \text{Index}(Q)$ and $\text{label}^f = \text{label}$ are taken from multiquery $Q$. Functor $\text{diag}^f$ is defined as follows. For each $q \in \text{Index}(Q)$, we define the diagram $\text{diag}^f(q)$ by setting:

(a) $\text{diag}^f(q)(IE) = \text{diag}^f(q)(II) = IE_{label(q)}$;

(b) $\text{diag}^f(q)(II) = \text{diag}^f(q)(II) \uplus f = b^f_1 \uplus f$;

(c) arrows $\text{diag}^f(q)(IE)$ and $\text{diag}^f(q)(EE)$ are given by the pushout of the two arrows above, which can be decomposed into two pushouts (1) and (2) as shown in diagram (30) (note also that monomorphisms are preserved by pushouts since $G|P$ is adhesive). We can thus say that diagram (e) gives us substitution for each atomic query $q \in \text{Index}(Q^f)$, and query $Q^f$ is defined. Commutativity of the square $\text{label}^f \uplus \alpha = \text{diag}^f \uplus |IE$ is obvious by construction.

By Lemma 1 both multiqueries $Q$ and $Q^f$ have their arity monics $IE_Q$ and $IE_{Q^f}$ (see diagram (32) hosted by category $G|P$) computed by the respective colimits: in the diagram, the upper face is the colimit cocone for $Q$ and the bottom face is the colimit cocone for $Q^f$. Squares $I_Q E_d, E_{q^f} Y$ are pushout squares.
Universality of the upper colimit gives us arrow $f^*_Q$, but we need to prove that the face square $I_Q E Q E Q Y$ is a PO. We prove it but a direct demonstration that the square is universal. Let $(x, y)$ in be a cospan making a commutative square with $I Q$ and $f$. Then for each index $q$ (shown in the diagram as $q_i$), as the square $I_Q E Q E Q Y$ is a pushout, we have a unique arrow $u_q$ (shown in the diagram as $u_i$). The family of these $u_q$ makes a cocone over $I E Q$, and as $E Q$ is the colimit cocone, there is a unique arrow $u!$ making all diagrams commutative. Hence, the front face is a pushout, and we built diagram (30)(a)

It is easy to see that this diagram is also a pushout in $G||P|O_E$. Recall from Lemma 1 that both multiqueries provide also monics $\eta_Q$, $\eta_{Q!}$ in $G||P|O_E$, and as label $l = label$ by construction, mapping $f^*_Q$ is a correct $G||P|O_E$-morphism. Now, if arrow $x$ in the diagram above is a $G||P|O_E$-morphism, then $G||P$-arrow $u!$ is automatically a correct $G||P|O_E$-arrow, which gives us diagram (30b) The functoriality of the construction immediately follows from the pushout composition lemma applied to diagram (30d).

Now suppose that $r: Q \rightarrow Q'$ is a m-query morphism. We extend diagram (30b) to a triangle prism as follows. The upper face is given by the triangle diagram in Corollary 7 (on p. 35), and then we take PO of $(f, \eta_{Q'})$ by applying PO decomposition to $\eta_{Q'} = \eta_{Q'} \cdot r$ which completes the prism. The right face is a PO square $(r, f_Q, f^*_{Q'}, r^*)$ and we define $f^{MQ}(r) = r^*$. If $r': Q' \rightarrow Q''$ is
another m-query morphism, then PO decomposition applied again gives us required functoriality $f^{\mathcal{MQ}}(r \uparrow r') = f^{\mathcal{MQ}}(r) \uparrow f^{\mathcal{MQ}}(r')$ — this is a horizontal version of diagram (30d).

The Grothendieck construction gives us the following

**Corollary 8.** The multiquery construction gives rise to an indexed category functor $\mathcal{MQ} : \mathcal{G} || \mathcal{P} \to \text{Cat}$ or, equivalently, split opfibration $\mathcal{G} || \mathcal{P} \leftarrow \mathcal{MQ}$, where $\mathcal{MQ}$ is the total category of all multiqueries. Moreover, reindexing is compatible with multiquery merge.

Merge preservation easily follows from the construction considered in the proof of Lemma 2.

### 5.2 Multiquery Semantics

The next step is to show that any $\mathcal{O}$-algebra is also a carrier of the $\mathcal{Q}$-operation of arity $IE_Q$. We will begin with an accurate re-definition of the notion of an $\mathcal{O}$-algebra.

**Definition 20 (Algebras (redefined)).** An algebra $A$ over $\mathcal{O}$ is given by a carrier sketch $|A|$ and a mapping $\llbracket |A| \rrbracket$, shown in diagram (33b), where the span $(\alpha_A', |A|')_{IE}$ is derived as a pullback in $\text{Set}$. Elements in its apex $LDiagr(p_A,q)\alpha_1$ can be identified with pairs $q = (\phi, \delta)$ in which $\delta$ is a pushout diagram in $\mathcal{S}$ with the left-bottom corner being $|A|$ and the upper monic being $\phi^a$ due to commutativity.

Function $\llbracket |A| \rrbracket$ maps any such pair to a triangle diagram in $\mathcal{S}$, in which

- the left-side (binding) arrow is the same as in diagram $\delta$ due to the commutativity of the trapezoid (a),
- the upper monic is $\phi^a$ due to commutativity $\alpha_A' \uparrow_{IE} = \llbracket |A| \rrbracket \uparrow_{IE}$;
- hence, the right-side arrow can be considered as the result of the operation: in fact, the right-side “formal” arrow in the pushout square $\delta$ freely added to graph $|A|$ is replaced by an actual arrow in $|A|$, namely, $\llbracket q \rrbracket |A|'_{IE}$.

Thus, diagrams at the target of arrow $\llbracket |A| \rrbracket$ are actually commutative triangles from Def. 2.

---

\[ \begin{array}{c}
I \ar[r]^-{IE} & E \\
H \ar[r] & EE \\
I_h \ar[r]^-{IE_h} & E_h
\end{array} \]

a) the shape of $\text{diagr}$-diagrams

\[ \begin{array}{c}
\text{Index}(Q) \ar[r]^-{\text{diagr}} & \text{label} \\
\llbracket |A| \rrbracket \ar[r]^-{L\text{Diagr}(A)} & \text{label} \\
I_q \ar[r]^-{\llbracket \alpha \rrbracket} & \text{label}
\end{array} \]

b) Multi-Queries and Algebras

(33)
Lemma 3 (Multiqueries as operations) Let $Q$ be a multiquery in a discrete signature $\mathcal{O}$. Then any $\mathcal{O}$-algebra is also an $\mathcal{O} + Q$-algebra, i.e. any binding $b : I_Q \to |A|$ gives rise to a unique $[Q]^A(b) : E_Q \to |A|$ such that $IE_Q \Downarrow [Q]^A(b) = b$.

Instead of saying that $A$ is an $\mathcal{O} + Q$-algebra we may simply say $A$ is a $Q$-algebra when $\mathcal{O}$ is clear from the context.

Proof. In diagram (33), we add to the internal rectangle defining an algebra $A$, the external rectangle defining a multiquery $Q$. Let $b_Q : I_Q \to |A|$ be a sketch morphism. Then post-composition gives us arrow $\_ : b_Q$ as shown in the diagram, and universality of pullbacks gives us arrow $!_{b_Q}$. Composing the latter with $\Downarrow [Q]^A$ gives us evaluation of all atomic queries in $Q$ and a wide cospan $(EA_q : E_q \to A \mid q \in \text{Index}(Q))$, where $EA_q = (!_{b_Q} \Downarrow [Q]^A)_{|E_A}$. Now diagram (28) can be completed as shown with dashed lines in diagram (34) below for the case of finite $\text{Index}(Q)$, and the universality of the colimit gives us a required arrow $!$ such that $b_Q = IE_Q \Downarrow !$. That is, we have defined an operation $[Q]^A$ that maps any input binding $b_Q$ to a scope binding $b_Q^E = !$.

![Diagram](image-url)

5.3 Complex Queries

Multiqueries can be sequentially composed into complex queries.

Definition 21 (Complex queries). Let $X$ be a sketch in $\mathbb{G}||\mathcal{P}$. A complex query $Q$ over $X$ is a finite sequences of multiqueries $\{Q_1, ..., Q_N(Q)\}$ such that $I_{Q_1} = X$ and $I_{Q_{i+1}} = E_{Q_i}$ for all $i \leq N(Q)$. The number $N(Q)$ is the length of the query, and we will simply write $N$ if the query $Q$ is clear. Thus, we have a sequence of monics in $\mathbb{G}||\mathcal{P}||\mathcal{O}_E$:

$$I_Q = X_0 \xrightarrow{\eta_1} X_1 \xrightarrow{\eta_2} \cdots \xrightarrow{\eta_N} X_N = Q(X)$$

where $X_i := Q_i(X_{i-1})$, $i \leq N$; its composition will be denoted by $\eta_Q : X \xrightarrow{} Q(X)$.

We also have the projection of the above to $\mathbb{G}||\mathcal{P}$:

$$I_Q = I_{Q_1} \xrightarrow{IE_{Q_1}} E_{Q_1} = I_{Q_2} \xrightarrow{IE_{Q_2}} \cdots \xrightarrow{IE_{Q_{N-1}}} E_{Q_{N-1}} \xrightarrow{IE_{Q_N}} E_{Q_N} = E_Q$$

where $|\mathcal{O} + Q| = |\mathcal{O}| \sqcup |Q|$ and $\alpha^{\mathcal{O} + Q} = \alpha^{\mathcal{O}}$, $\alpha^{\mathcal{O} + Q}(Q) = IE_Q : I_Q \to E_Q$
5 Basic Diagram Algebra: Operations without Dependencies

whose composition will be denoted by \( IE_Q \).

Note that an multi-query \( Q \) over sketch \( X \) can be interpreted as a complex query \( \eta_Q: X \to Q(X) \) of length 1. And empty multiquery \( 0_X \) is a complex query of length 0. Below we will often write \( m \)-query and \( c \)-query for, resp., multiqueries and complex queries.

Two \( c \)-queries over \( X \) of equal length, \( Q = (Q_1, \ldots, Q_N) \) and \( Q' = (Q'_1, \ldots, Q'_N) \), are called equivalent, \( Q \cong Q' \), if there is a sequence of multiquery isomorphisms \( \iota_i: Q_i \to Q'_i \) in \( MQ(X_i) \), \( i = 1, \ldots, N \) such that \( \eta_i \circ \iota_i = \iota_{i-1} \circ \eta'_i \) and \( \iota_0 = \text{id}_X \).

The class of all \( c \)-queries over \( X \) modulo \( \cong \) will be denoted by \( cQQ(X) \) (and similarly, \( \text{Ob}(MQ) \) can be referred to as \( mQQ(X) \)).

**Definition 22 (Sequential Composition).** Let \( Q = (Q_1, \ldots, Q_N) \) and \( Q' = (Q'_1, \ldots, Q'_N) \) be two complex queries with units \( \eta_Q: X \to Q(X) \) and \( \eta_{Q'}: E_Q \to Q'(X) \). Then we define a complex query \( Q \circ Q' \) of length \( N + N' \) by the concatenation of sequences \( Q \) and \( Q' \):

\[
Q \circ Q' = (Q_1, \ldots, Q_N, Q'_1, \ldots, Q'_{N'})
\]

with \( \eta_{Q \circ Q'} = \eta_Q \circ \eta_{Q'} \).

**Lemma 4 (Parallel Composition)** Let \( Q = [Q_1, \ldots, Q_N] \) and \( Q' = [Q'_1, \ldots, Q'_N] \) be two complex queries over sketch \( X \in S \) with \( \eta: X \to Q(X) \) and \( \eta': X \to Q'(X) \) being their respective units. Then there exists a complex query \( Q + Q' \) (alternatively written \( Q_+ \)) with unit \( \eta_+: X \to Q_+(X) \) given by the following pushout diagram

\[
\begin{array}{ccc}
X & \xrightarrow{\eta_Q} & Q(X) \\
\downarrow{1_{E_Q}} & & \downarrow{\eta} \\
E_Q & \xrightarrow{\eta_Q} & Q_+(X)
\end{array} 
\quad \begin{array}{ccc}
X & \xrightarrow{\eta_{Q'}} & Q'(X) \\
\downarrow{1_{E_{Q'}}} & & \downarrow{\eta'} \\
E_{Q'} & \xrightarrow{\eta_{Q'}} & Q_+(X)
\end{array}
\]

\( (a) \quad (b) \)

**Proof.** Suppose first that \( N = N' \) and consider the matrix of pushout squares in (36) below built inductively in the following way. We begin with merging queries \( Q_1 \) and \( Q'_1 \) over \( X \) (by Prop. 4) and obtain diagonal pushout square (11) with monic \( \eta_1^* \) denoting \( \eta_{Q_1 \oplus Q'_1} \). Then we apply \( m \)-query substitution Lemma 2 to query \( Q_2 \) over \( Q_1(X_0) \) along mapping \( \eta_1^{*\ast} \) and obtain query \( Q_2^{*\ast} \) (not shown in the diagram) with pushout square (12) and specifically monic \( \eta_2^{*\ast} \). Similarly, we obtain query \( Q_2^{*\ast} \) (not shown) with square (21) and monic \( \eta_2^{*\ast} \) by applying substitution along \( \eta_1^{*\ast} \) to query \( Q_2 \). Now we have two \( m \)-queries, \( Q_2^{*\ast} \) and \( Q_2^{*\ast} \) over the same sketch \( X_1^+ \), which can be merged and produce monic \( \eta_2^*: \eta_2^{*\ast} + Q_2^{*\ast} \), and so on until we get the last \( m \)-query \( Q_N^{*\ast} + Q_N^{*\ast} \) with unit \( \eta_N^*: \eta_N^{*\ast} + \eta_N^{*\ast} \): \( X \to X_N \). If \( N \neq N' \) we extend the shorter query (say, \( Q' \)) with empty multiqueries so that \( Q'' = (Q'_1, \ldots, Q'_{N'}, 0_{N'+1}, \ldots, 0_N) \) and define \( Q + Q' := Q + Q'' \).
Table of Contents

Corollary 9 (Interchange law for complex queries). Let $X$ be a $P$-sketch, and $Q_1, Q_2, Q_3$ and $Q_4$ be $c$-queries with $X = I_{Q_1} = I_{Q_2}$, $I_{Q_3} = Q_1(X)$, and $I_{Q_4} = Q_2(X)$. Then the interchange law holds:

$$Q_{1+2} \cdot Q_{3+4}(X) = (Q_{1;3} + Q_{2;4})(X)$$

Proof. We take the left part, apply the matrix from Lemma 4 to $Q_{1;3}$ and to $Q_{2;4}$ separately, which will give us two matrices of size $N_{12} \times N_{12}$ and $N_{34} \times N_{34}$ (where $N_{12} = \max\{N(Q_1), N(Q_2)\}$ and $N_{34} = \max\{N(Q_3), N(Q_4)\}$) adjoint at the vertex $N_{12}$. Then we apply pushout decomposition lemma to fill-in the empty spaces and obtain a complete pushout matrix of size $(N_{12} + N_{34}) \times (N_{12} + N_{34})$, which gives us the right side of (37).

Lemma 5 (Complex Query Substitution) Let $Q = [Q_1, \ldots, Q_n]$ be a complex query with input sketch $I_Q = X$, and $f: I_Q \rightarrow Y$ a sketch morphism. These data uniquely determine a complex query $Q'$ with input sketch $I_{Q'} = Y$ and unit $\eta_{Q'}$ given by pushout diagrams (b) in (38) (similar to that in Lemma 2 on p.36) and $(a, c)$ similar to those $(a, c)$ Likewise, $Q$ is functorial: $Q(f \cdot g) = Q(f) \cdot Q(g)$.

Proof. The proof consists of iterated applications of Lemma 2 to the sequence of $m$-queries $Q_1, \ldots, Q_n$, which results in the sequence of substituted $m$-queries $Q'_1, \ldots, Q'_n$ together with respective pushout squares depicted in (39a). The pushout pasting lemma allows us to combine all these pushouts into a big pushout. We can apply the same principle not only horizontally but also verti-
cally, which provides us with functoriality of $Q$.

\[
\begin{align*}
X &= I_{Q_1} \xrightarrow{1_{E_Q}} E_{Q_1} \xrightarrow{1_{E_Q}} \ldots \xrightarrow{1_{E_Q}} E_{Q_n} = E_Q \\
Y &= I_{Q'_1} \xrightarrow{1_{E_{Q'_1}}} E_{Q'_1} \xrightarrow{1_{E_{Q'_2}}} \ldots \xrightarrow{1_{E_{Q'_n}}} E_{Q'_n} = E_{Q'}
\end{align*}
\]

\((39)\)

**Corollary 10 (Category of complex queries).** The collection of all $\mathcal{P}$-sketch as objects (modulo their isomorphism) and $\mathcal{C}$-queries (modulo $\cong$) as arrows form a category denoted by $\mathcal{CQ}$:

\[
\mathcal{CQ}(X, X') = \{ Q \in \mathcal{CQ}(X) | \eta_Q: X \rightarrow X' \}
\]

and identity $\text{id}_{\mathcal{CQ}}(X) = 0_X: X \rightarrow X$.

### 5.4 Views and their Composition

**Definition 23 (Views).** Let $\mathcal{S}$ be a $\mathcal{P}$-sketch called the source, and $\mathcal{O}$ an operation signature over $\mathcal{P}$. An $\mathcal{O}$-view of $\mathcal{S}$ is a pair $v = (Q_v, \hat{v})$ with $Q_v$ a complex $\mathcal{O}$-query over $\mathcal{S}$ and $\hat{v}: Q(S) \rightarrow V$ a sketch morphism called view (definition) mapping; the domain of the mapping is called the view schema.

Thus, a view is a cospan $\mathcal{S} \xrightarrow{\eta_v} Q_v(S) \xleftarrow{\hat{v}} V$, which we will write as $\mathcal{S} \xleftarrow{\eta_v} Q_v(S) \xrightarrow{\hat{v}} V$ to ease notation. Note that the left leg is monic while the right one is arbitrary. We will also write views as $v: V \rightarrow S$ or $v_Q: V \rightarrow S$ to make the query explicit.

**Theorem 3 (Category of Views (local Kleisli construction)).** Views are associatively composable and thus make a category $\mathcal{S}_\mathcal{O}$ whose objects are sketches in $\mathcal{S} = \mathcal{G}\|\mathcal{P}$, and arrows are equivalence classes $\mathcal{O}$-views between them.

This equivalence relation $\sim$ is defined as follows: Two views $\mathcal{S} \xleftarrow{\eta_{v_1}} Q_{v_1}(S) \xrightarrow{v_1} V$ and $\mathcal{S} \xleftarrow{\eta_{v_2}} Q_{v_2}(S) \xrightarrow{v_2} V$ are equivalent, $v_1 \sim v_2$, iff there exists an $\mathcal{G}\|\mathcal{P}\|\mathcal{O}_E$-isomorphism $i: Q_{v_1}(S) \rightarrow Q_{v_2}(S)$ such that $\eta_{v_1} \circ i = \eta_{v_2}$ and $v_1 \circ i = v_2$.

**Proof.** To define a category we have to define two operations (identities and composition) and show three properties to hold (left unitor, right unitor and associativity).

Composition:
Let $v_1: V \rightarrow S$ and $v_2: S \rightarrow T$ be two views with $V, S$ and $T$ being sketches in $\mathcal{G}\|\mathcal{P}$. Their composition $v_1 \circ v_2: V \rightarrow T$ is defined as follows: Consider the diagram in (40): first, we employ the substitution Lemma 5 on p.42 to substitute $Q_{v_2}$ along $v_2$ into $Q_{v_1}$ resulting in the pushout square (a) and a new substituted complex query $Q_{v_1}^{v_2}$ with input $I_{Q_{v_1}^{v_2}} := Q_{v_2}(T)$. Evidently we can merge two adjacent complex queries into one complex query by concatenating their sequences.
of multiqueries. Thus, we get a complex query $Q_{v_2} \triangleright Q_{v_1}$ (or $Q_{v_1} \circ Q_{v_2}$) with arity monic

$$IE_{Q_{v_1} \circ Q_{v_2}} = \eta_{v_2} \triangleright \eta_{v_1} ; T \mapsto Q_{v_1}(Q_{v_2}(T)).$$

Together with the composite $v_1 \triangleright v_2^*$ for the view mapping, we get the view $v_1 \triangleright v_2$.

$$\begin{array}{c}
T \\
\downarrow \eta_{v_2} \\
Q_{v_2}(T) \quad (a) \quad Q_{v_1}(S) \\
\downarrow \eta_{v_1}^* \\
Q_{v_1}(Q_{v_2}(T)) \\
\downarrow \eta_{v_1}^{**} \\
S \\
\downarrow \eta_{v_2}^* \\
V \\
\end{array}$$

**Identities:**

For any sketch $S \in S$ we can consider a special complex query $Q^1_S$, whose sequence of multiqueries is empty and thus $I_{Q^1_S} = E_{Q^1_S}$ and $IE_{Q^1_S} = \eta_{Q^1_S} = id_S$.

Hence, we define the identity view $id_S : S \rightarrow S$ on $S$ as the span $\eta_{Q^1_S} : Q^1_S(S) = S \xrightarrow{id} S$.

**Unitors:**

Consider the diagrams in (41) depicting the compositions $v_1 \triangleright id_S$ and $id_V \triangleright v_1$ respectively, where $v_1 : V \rightarrow S$ is a arbitrary view and $id_S$ and $id_V$ are identity views over $S$ and $V$. Substituting $Q_{v_1}$ along $id_S$ (l) results in the same query and concatenating $Q^1_V$ (empty) with $Q_{v_1}$ results in $Q^1_S$ and since $v_1 \triangleright id = v_1$ it follows $v_1 \triangleright id_S = v_1$. On the other hand, substituting (r) an empty query $Q^1_V$ along a morphism changes the input $I_{Q^1_S}$ (which is also the output) and the sequence of multiqueries stays empty. Concatenating it with $Q_{v_1}$ results in $Q_{v_1}$ and again $id \triangleright v_1 = v_1$ and thus it follows $id_V \triangleright v_1 = v_1$.

**Associativity:**

McSCert report TR-2020-33 44 May 19 2020
Let \( v_1: V \to S \), \( v_2: S \to T \) and \( v_3: T \to U \) be three views, which are depicted in the upper half of (42).

We claim that \( (v_1 \circ v_2) \circ v_3 = v_1 \circ (v_2 \circ v_3) \), i.e. they both result in the same view: It does not matter if we first substitute \( Q_{v_1} \) along \( v_2 \) (a), concatenate with \( Q_{v_2} \) along \( v_3 \) (b) + (c) and finally concatenate with \( Q_{v_3} \), or if we first substitute \( Q_{v_2} \) along \( v_3 \) (a), concatenate with \( Q_{v_3} \) along \( v_2 \circ v_3 \) (a) + (c) to finally concatenate \( Q_{v_3} \circ Q_{v_2} \) with \( Q_{v_1} \). In both cases the result is a configuration of three uniquely determined PO squares (recall that our sketch categories are skeletal) and a complex query \( Q_{v_3} \circ Q_{v_2} \circ Q_{v_1} \) with unit \( v_1 \circ v_2 \circ v_3 \).

**Theorem 4 (Views form a SMC).** The category of \( \mathcal{O} \)-views \( \mathcal{S}_\mathcal{O} \) can be endowed with a symmetric monoidal structure \((\otimes, I)\) (via coproducts and initial objects in \( \mathcal{S} \)).

**Proof.** To show has \( \mathcal{S}_\mathcal{O} \) has a symmetric monoidal structure we have to extend the proof in Thm. 3 with definitions of the monoidal product \( \otimes \), the monoidal unit \( I \) and have to show that the monoidality laws hold.

**Monoidal Product:**

The monoidal product of two sketches \( S_1 \) and \( S_2 \) is simply their coproduct (note that \( \mathcal{S} \) is a topos). The monoidal product of two morphisms (= views)

\[
S_i, \eta_i \rightarrow Q_i(S_i) \leftarrow V_i, i = 1, 2
\]
the monoidal (parallel) composition of views:

any four views has to hold:

universal property of coproducts. Now the cospan is a result of parallel composition (Lem. 4) and mapping \( u \) is specified by diagram (43) in S:

\[
\begin{align*}
Q_1(S) & \xrightarrow{\iota_1^*} Q_1(S_1 \sqcup S_2) & (l) \\
& \xrightarrow{\eta_1^*} \eta_1^* + \eta_2^* & \\
& \xrightarrow{Q_2^*(S_1 \sqcup S_2)} \xleftarrow{(Q_1^* + Q_2^*)(S_1 \sqcup S_2)} & (r) \\
& \xleftarrow{\iota_2^*} S_2 & \\
& \xleftarrow{\eta_2^*} & \\
V_1 \sqcup V_2 & \xrightarrow{\eta_1^*} & V_2 & \xrightarrow{\eta_2^*} & V_2
\end{align*}
\]

Here, \( S_1 \sqcup S_2 \) and \( V_1 \sqcup V_2 \) are coproducts with their respective injections, the diamonds \((l)\) and \((r)\) are substitution squares (Lem. 5), mapping \( \eta_1^* + \eta_2^* \) is a result of parallel composition (Lem. 4) and mapping \( u \) is given by the universal property of coproducts. Now the cospan \((\eta_1^* + \eta_2^*), u\) is defined to be the monoidal (parallel) composition of views:

\[
(v_1 \oplus v_2)_{Q_1 \oplus Q_2} : S_1 \oplus S_2 \leftarrow V_1 \oplus V_2
\]

We have to show that \( \oplus \) is a bifunctor a.k.a as the interchange law, i.e for any four views \( v_1 : T \leftarrow W \), \( v_2 : S \leftarrow V \), \( v_3 : P \leftarrow T \) and \( v_4 : U \leftarrow S \) the following has to hold:

\[
(v_1 \oplus v_2) \\
\downarrow \downarrow \downarrow \downarrow
(v_3 \oplus v_4) = (v_1 \downarrow v_3) \oplus (v_2 \downarrow v_4)
\]

The left hand side of this diagram is explicated in (45) below (the final result is highlighted in blue):

\[
\begin{align*}
(43)
(44)
\end{align*}
\]

McSCert report TR-2020-33 46 May 19 2020
and the right hand side in the diagram in (46):

To show that (44) holds, we need to show that there exists an isomorphism between $Q_{3||4} \dashv Q_{3||2}(U \sqcup P)$ and $Q_{3\|1\|2}(U \sqcup P)$ that makes the cospan triangles commute. This isomorphism is provided to us by (37) in Lemma 9 which also makes the left hand side triangle commute and the universal coproduct property of $V \sqcup W$ makes the right hand side triangle commute.

**Monoidal Unit:**

The monoidal unit is given by the empty sketch $0 \in \text{Ob}(S)$ (initial object in $S$).

**Monoidality:**

Note that $S$ is a topos and thus it has a symmetric monoidal structure $(+, 0)$ given by coproduct and initial objects. Let now $\alpha, \lambda, \rho$ and $\sigma$ denote the associator, left-, right-unitor and braiding isomorphism families. Furthermore, there exists an identity-on-objects functor $(\cdot): S \to S_\mathcal{O}$ that sends a sketch morphism $f: A \to B$ to the trivial view $B \sqcup_{0_B} (B \sqcup f) \downarrow A$. Functors preserve isomorphisms and thus the images of $\alpha, \lambda, \rho$ and $\sigma$ under this functor establish the coherence on $S_\mathcal{O}$.

### 5.5 Query Monad

**Background.** Complex queries are the diagrammatic equivalent of complex terms in the ordinary algebra, in which the term construction can be specified by the notion of a monad or, equivalently, a Kleisli triple, over some carrier category $\mathcal{C}$.

**Definition 24 (Global Kleisli triple).** A Kleisli triple over $\mathcal{C}$ is $(Q, \eta, \alpha^*)$ where $Q: \text{Ob}(\mathcal{C}) \to \text{Ob}(\mathcal{C})$ is an endofunction over $\text{Ob}(\mathcal{C})$, $\eta$ a family of $\mathcal{C}$-arrows
Table of Contents

\[ \eta_S: X \to Q(X), \] \( X \in \text{Ob}(\mathbb{C}) \), and \_*\text{ is an operation that send any } v: Q(S) \leftarrow V \text{ in } \mathbb{C} \text{ to } v^*: Q(S) \leftarrow Q(V) \text{ illustrated by the diagram below:} \]

\[
\begin{array}{ccc}
S & \xrightarrow{\eta_S} & V \\
\downarrow v & & \downarrow w \\
Q(S) & \xleftarrow{v^*} & Q(V) \\
\end{array}
\]

such that the following three equation hold: (a) \( \eta_X^* = \text{id}_X \), (b) \( \eta_V \circ v^* = v \) and (c) \( (w \circ v^*)^* = w^* \circ v^* \). Such a triple is called \textit{mono} if all \( \eta_X \) are monic.

A triple is \textit{finitary} if \( Q \) commutes with filtered colimits: \( Q(\text{colim } \{ X_i \mid i \in \mathbb{I} \}) = \text{colim } \{ Q(X_i) \mid i \in \mathbb{I} \} \) for some functor \( X: \mathbb{I} \to \mathbb{C} \) from a filtered category \( \mathbb{I} \).

Note that Kleisli arrows above are defined \textit{globally} in the sense that object (sketch) \( Q(S) \) is generated by applying \textit{all} complex terms/queries and different views \( v, w \) use the same construction on the target. In contrast, our category of views \( \mathcal{S}_O \) in Thm. 3 is defined \textit{locally}; each view carries its own query necessary to define the view mapping as illustrated below.

\[
\begin{array}{ccc}
S & \xrightarrow{\eta_S} & V \\
\downarrow w & & \downarrow w \\
Q_w(S) & \xleftarrow{w^*} & Q_w(V) \\
\end{array}
\]

where square (1) is provided by the query substitution (based on Lemma 5) as explained above.

Given a sketch \( X \in \mathcal{S} \), let \( c_\mathcal{Q}X_n(X) \) be the set of all complex queries \( Q \) over \( X \) with \( I_Q = X \) and the length equal or less than \( N \). We thus have a chain

\[ c_\mathcal{Q}X_1(X) \subset c_\mathcal{Q}X_2(X) \subset \ldots \]

of wide spans. Taking wide PO of each of them gives us a commutative diagram of monics in \( \mathcal{G}/|\mathcal{P}|/\mathcal{O}_E \), \( \eta_n \circ \iota_n = \eta_{n+1}, \ i = 1, 2, \ldots \)

\[
\begin{array}{ccc}
X & \xrightarrow{\eta_1} & Q_1(X) \\
\downarrow \eta_2 & & \downarrow \eta_2 \\
Q_2(X) & \xrightarrow{\eta_3} & Q_3(X) \\
\ldots & & \ldots \\
Q_X(X) & \xrightarrow{\eta_3} & Q_X(X) \\
\end{array}
\]

whose colimit is a \( \mathcal{G}/|\mathcal{P}|/\mathcal{O}_E \)-monic denoted by \( \eta^X_X: X \to Q_X(X) \).

Now Kleisli composition is specified by the following diagram, where the middle diamond is a query substitution PO provided by Lemma 5.

The Kleisli star operation requires the existence of arrow \( v^*_X \), and such would be provided by the universal property of \( Q_X(V) \) if for each query \( Q \) over \( V \), we would have the diagonal arrow \( d_{Q,v} \) as shown in the diagram, which in turn would be provided by the “monadic multiplication” arrow \( \mu_{Q,v} \), so that we could define \( d_{Q,v} := v^*_Q \circ \mu_{Q,v} \). The existence of such arrows \( \mu \) is a typical feature of the term formation procedure (“a term over a term is a term”), but if the image
of \( v \) spans the entire \( Q_x(S) \), this construction does not work. Suppose, e.g., that \( V = \{ x_1, \ldots, x_n, \ldots \} \) is countable and \( v(x_n) \in cQx_n(S) \) for all \( n \). Then, if a query \( Q \) over \( V \) uses all “variables” \( x_i \), then its substitution \( Q^f \) along \( v \) will span the entire \( Q_x(S) \) and disable the construction. To make it working, we need to refine the query substitution mechanism and introduce some restrictions on the number of variables (i.e., the size of the sketch) involved into queries.

**Effective Substitutions.** Recall that presheaf toposes have image factorization, and hence we can factorize the binding square for atomic queries into two POs as shown below, where \( \eta_{Q}^{\text{eff}} \) denotes the image of \( b_{q} \) in sketch \( I_{\phi} \).

\[
\begin{array}{c}
I_{q} \xrightarrow{b_{q}} I_{\phi}^{\text{eff}} \xleftarrow{b_{q}^{\text{eff}}} I_{\phi} \\
E_{q} \xrightarrow{I_{E_{q}}} I_{\phi}^{\text{eff}} \xleftarrow{I_{E_{q}^{\text{eff}}}} E_{\phi}
\end{array}
\]

This factorization can be taken into account for m-query and c-query substitution, and results in the following.

**Lemma 6 (Effective Substitutions)** For any \( P \)-sketch \( X \) and query \( Q \) over \( X \), we have a composition of three POs as shown in diagram (a) below, where we write \( \eta_{Q}^{\text{eff}} \) instead of an accurate but monstrous \( \eta_{Q}^{\text{eff}}_{Q} \). Diagram (b) is the \( G||P \)-projection of (a). \( \square \)
Now we introduce necessary size restrictions. Rather than doing this in a general setting with cardinal arithmetics, we will do it for the simple finitary case (but the pattern is generalizable for any infinite cardinal).

**Definition 25 (Finitary Queries).** An operation signature \( \mathcal{O} \) is called finitary if for all \( \phi \in \mathcal{O} \), sketch \( E_\phi \) (and hence \( I_\phi \)) is finite. A multiquery \( Q \) is finitary if \( \mathcal{O} \) is finitary and set \( \text{Index}(Q) \) is finite. A complex query \( Q \) is finitary if all \( m \)-queries in the sequence \( (Q_1, \ldots, Q_N) \) are finitary.

**Corollary 11 (Finitarity and Substitutions).** If query \( Q \) over \( X \) is finitary, then sketch \( Q(X_\text{eff}) \) is finite. Moreover, for any mapping \( f: x \rightarrow Y \), the sketch \( Q|_{Q_f}^{\text{eff}}(Y_\text{eff}_{Q,f}) \) (and hence \( E_\text{eff}_{Q,f}^f \)) is finite too.

Now we can implement the proof idea outlined above. For any global view \( v: Q_{\mathcal{X}}(S) \leftarrow V \) and a finitary query \( Q \) over \( V \), we apply Lemma 6 on effective substitution and build the diagram in Fig. 12 (ignore the dotted arrows for a moment), where we use a new lighter notation for the two intermediate sketches. As sketch \( Q_{\mathcal{X}}^* \) is finite, there is a number \( N \) (depending on \( Q, f \)) such that \( V_{\mathcal{X}}^* \subseteq Q_{N+1}(S) \), and hence sketch \( Q_{\mathcal{X}}^* \) can be mapped to \( Q_{N+1}(S) \subseteq Q_{\mathcal{X}}(S) \) (note arrow \( \mu_{Q,\mathcal{X}} \)). Composition of \( \mu_{Q,\mathcal{X}} \) with \( (v_\text{eff}_{Q,f})^* \) gives us \( d_\mathcal{X,v} \) for each query \( Q \) over \( V \), and thus a cocone with vertex at \( Q_{\mathcal{X}}(S) \). However, \( Q_{\mathcal{X}}(V) \) is a universal (minimal) cocone over the same base, and hence we have a unique arrow \( d_{\mathcal{X,v}} \). This arrow can be taken as the result the global Kleisli triple operation \( \_^* \) (denoted by \( \_^* \) in our notation). Checking monoidal properties (a,b) of Def. 24 is straightforward; property (c) is our local Kleisli composition defined above. We thus come to the following result.

**Theorem 5 (Query Monad).** A finitary diagrammatic query-building mechanism (Def. 25) gives rise to a finitary \( \eta \)-mono (global) Kleisli triple \( (Q_{\mathcal{X}}, \eta_{Q,\mathcal{X}}, \_^*) \) whose components were defined above. As is well-known, such a triple is equivalent to a finitary monad satisfying the \( \eta \)-mono requirement.
6 Advanced Diagram Algebra: Dependencies between Operations

In the previous section we investigated diagram algebra for the discrete case, i.e. $\mathcal{O}$ being a set. Now we consider signatures of operations with dependencies, which will provide us with reasoning capabilities. We will begin with a motivating example.

**Example 8 (Universal Pullback Property).** To specify commutativity of the pullback square, we need to a) extend the signature $\mathcal{P}_{\text{monic}}$ with a new predicate symbol $[; ;]$ whose input arity $I[; ;]$ are two consecutive arrows and its scope arity is the triangle $E[; ;]$, and b) add to the sketch $E_{[pb]}$ a new arrow (the diagonal of the square) plus two $[; ;]$-constraints declaring the two triangles dividing the square to be commutative. Thus, $E_{[pb]}$ becomes a $\{[\text{monic}], [; ;]\}$-sketch. Now, if we want to specify the universal property of pullbacks, we add to the signature a new operation symbol $\text{upb}$, whose input arity is a sketch in the signature $\mathcal{P}_{\text{upb}} = \{[\text{monic}], [\text{pb}], [; ;]\}$ (where $[\text{pb}]$ is a predicate symbol with arity $E_{[pb]}$). Fig. 13b depicts the arity of this $\text{upb}$. Black-colored elements are part of $I_{[\text{upb}]}$ and blue-colored elements are added by $E_{[\text{upb}]}$. The input sketch comprises two squares together with respective diagonals and commutativity conditions expressed via $[; ;]$-triangles, additionally one of the squares has a $[\text{pb}]$-predicate. The scope sketch adds a new arrow $u$ together with two commutative triangle predicates $[; ;]$. 

Figure 13 presents a formalized description of the above. We have a predicate signature (A) consisting of a predicate symbol $\text{monic}$ and above it an operation signature $\mathcal{O}$ consisting of three operation symbols $\{[; ;], [\text{pb}], [\text{upb}]\}$. There are two types of dependencies between symbols. Preconditions (shown in orange) specify conditions to make an operation applicable, and any such dependency
arrow is assigned with a mapping of the scope sketch of the dependency’s target into the input sketch of the dependency’s source. Postconditions (shown in blue) specify conditions that the result of operation must always satisfy, and any such dependency arrow is assigned with a mapping between the scopes of the related symbols.

![Diagram](image)

(a) Categorical Notation

![Diagram](image)

(b) Arity of [upb]

![Diagram](image)

(c) Workflow

Fig. 13: Example: Universal Pullback Property

6.1 Functorial Operation Signatures (with Dependencies)

We will need the notion of diagonal arrow category described in Def. 31, 32 on p. 58.
Definition 26 (Operation Signature with Dependencies). An operation signature $\mathcal{O}$ is given by two contravariant functors $\alpha_{IE}: \mathcal{O}_{\text{op pre}} \to [\cdot \rightarrow \cdot, \mathcal{S}]_{\text{diag}}$ and $\alpha_{EE}: \mathcal{O}_{\text{op post}} \to \mathcal{S}$ where $\mathcal{O}_{\text{pre}}$ is a subcategory of $\mathcal{O}_{\text{post}}$ sharing the same set of objects and the diagram (52) in $\textbf{Cat}$ commutes.

\[ \begin{array}{ccc}
\mathcal{O}_{\text{op pre}} & \xrightarrow{\alpha_{IE}} & [\cdot \rightarrow \cdot, \mathcal{S}]_{\text{diag}} \\
\mathcal{O}_{\text{op post}} & \xrightarrow{\alpha_{EE}} & \mathcal{S}
\end{array} \]  

(52)

The diagram category $[\cdot \rightarrow \cdot, \mathcal{S}]_{\text{diag}}$ is a subcategory of $[\cdot \rightarrow \cdot, \mathcal{S}]$ with chosen *diagonals*, thus every morphism $(\Delta_I, \Delta_E): \delta \Rightarrow \delta'$ in $[\cdot \rightarrow \cdot, \mathcal{S}]$ is identified by a morphism $d: \delta(E) \to \delta'(I_E)$ in $\mathcal{S}$. Furthermore, we call arrows in $\text{Arr}(\mathcal{O}_{\text{pre}})$ *preconditions* and those in $\text{Arr}(\mathcal{O}_{\text{post}})$ *postconditions*. However, amongst those, only arrows in $\text{Arr}(\mathcal{O}_{\text{post}}) \setminus \text{Arr}(\mathcal{O}_{\text{pre}})$ are really interesting postconditions, and later by postconditions we will only mean the latter.
Definition 27 (O-algebra). Given an operation signature \( O \), an \( O \)-algebra \( A \) is given by a sketch \( |A| \) and two mappings \( \emptyset \rightarrow \emptyset \) and \( \emptyset \rightarrow \emptyset \), see (53), where \( LDiagr(A)_pre \) is given by the pullback of \( |E| \) and \( \alpha_E \) and \( LDiagr(A)_post \) by the pullback of \( \text{dom} \) and \( \alpha_E \).

Definition 28 (O-multiquery). Given an operation signature \( O \), an \( O \)-multiquery is given by two spans \( (\text{diag}_pre, \text{label}_pre) \) and \( (\text{diag}_post, \text{label}_post) \) where the apex of the former \( \text{Index}(Q)_pre \) is a subcategory of \( \text{Index}(Q)_post \).

6.2 Hierarchical Operation Signatures and Queries

In this section, the previous categorical/functorial definitions are restated in the discrete but hierarchical way.

Definition 29 (Hierarchical operational signatures). Let \( G \) be a presheaf topos of \( G \)-graphs as before. A predicate dh-signature of height \( N \) is given by the inductively defined data specified by the commutative diagram (54).

\[
P \rightarrow O^*_0 \rightarrow O^*_1 \rightarrow \cdots \rightarrow O^*_{N-1}
\]

The upper row is provided by a discrete hierarchical category \( O^* \) of predicate symbols of height \( N \), i.e., a finite chain of non-empty sets shown in the top row of the diagram. Each of these sets \( O^*_i \) is equipped with an arity function \( \alpha^*_i \) into the corresponding class of sketch monics

\[
S_{i-1} \overset{\text{def}}{=} S_{i-2}||O^*_{i-1}, i < N
\]
as shown in the bottom row of the diagram (see Def. 18 on p. 34 that explains the notation). The sequence starts from \( i = 0 \) for which \( S_0 = G||P \), i.e., sketches
without operation symbols, which allows us to define mapping $\alpha_0^*$, which in turn defines sketch category $S_0$ and so on. Recall that diagonal dotted arrows labelled $\sigma$ are pairs rather than mappings.

The above defines a dh-signature $O^*$. Sketches over $O^*$ and their morphisms are defined as objects and arrows of the category $S_{N-1} \equiv S_{N-2}/\parallel O_{N-1}^*$. We will denote this category by $G/\parallel O^*$: strictly speaking, we should write $G/\parallel O^*$ as it’s another construct of building sketches, but we will usually omit the superscript near $\parallel$ if the right term unambiguously shows the dh-context of $\parallel$.

To be Completed

7 Conclusion and Future work

To be completed.

\[ \begin{array}{c}
I_{\text{label}(q)} \xrightarrow{IE_{\text{label}(q)}} E_{\text{label}(q)} \\
\downarrow b_f^* \quad \quad \quad \downarrow b_k^* \\
I_Q \xrightarrow{I_E q^*} E_q \\
\downarrow f \quad \quad \quad \downarrow f_q^* \\
X_q^* \xrightarrow{I_E q^*} E_{q^*} \\
\end{array} \]
References


McSCert report TR-2020-33 56 May 19 2020
A Appendix. Diagram Categories

We will recall a few basic definitions and results and provide several new auxiliary constructs needed in the paper.

A diagram in a category $C$ is a functor $\delta: D \to C$ whose domain is a category freely generated by a graph $D$ called the shape of the diagram (and $D$ is also called the shape). Sometimes we will make $D$ a sketch by adding to it corresponding declarations, e.g., a valid diagram of shape $\begin{array}{c} \text{po} \\ \text{an} \end{array} \to \begin{array}{c} \text{an} \\ \text{po} \end{array}$ must map two horizontal arrows into monics, and the entire square into a pushout square in $C$. A precise formalization of this construct in general is exactly the subject of the paper, but in the metatheory we will only use simple sketches like above that can well be explained like above.

A diagram morphism is a natural transformation between functors, $\Delta: \delta \Rightarrow \delta'$, i.e., a family of arrows $\Delta_x: x^\delta \to x'^{\delta'}$ indexed by $x \in \text{Ob}(D)$ such that $a^\delta \circ \Delta_y = \Delta_x \circ a'^{\delta'}$ for any arrow $a: x \to y$ in $D$. This gives us the category $[D, C]$ of all diagrams in $C$ of shape $D$.

In subsection Sect. A we will consider categories of diagrams whose shape is a single arrow.

Arrow categories.

**Definition 30 (Arrow Categories).** Any category $C$ gives rise to an arrow category $[\to, C]$ whose objects are $C$-arrows and morphisms are commutative squares: given two $C$-arrows $a, b$, their morphism is a pair of $C$-arrows, $f: \circ a \to \circ b$. 

---


For any $r$, such a pair an $\circ b$ and $f'$: $a \circ b \to b \circ f$, such that the square commutes: $a : f' = f \circ b$. We call such a pair an sq-arrow and write $(f, f')$: $a \Rightarrow b$.

Any object $X \in C$ gives rise to a dom-pointed arrow category $[\rightarrow \cdot, C] \subset [\rightarrow \cdot, C]$ whose objects are arrows from $X$ and the corresponding component of all sq-arrows is $\text{id}_X$. Similarly, we have cod-pointed arrow category $[\rightarrow X, C] \subset [\rightarrow X, C]$. These categories usually denoted by $C \backslash X$ and $C/X$ resp.

The following facts can be proven easily:

**Fact. 1** There are two obvious functors $\text{dom}: [\rightarrow \cdot, C] \to C$ and $\text{cod}: [\rightarrow \cdot, C] \to C$, for which the following holds.

1. For any $C$, functor $\text{dom}$ is a split fibration (by precomposition).
2. The restriction $\text{dom}_X: C/X \to C$ (for a given object $X$) is a discrete fibration.
3. For any $C$, functor $\text{cod}$ is a split opfibration (by postcomposition).
4. The restriction $\text{cod}_X: C \backslash X \to C$ is a discrete opfibration,

When furthermore $C$ has all pushouts (pullbacks) we can build more.

**Fact. 2** 1) If $C$ has all pushouts, the functor $\text{dom}$ is an opfibration as well where lifts are constructed via pushouts.

2) If $C$ has all pullbacks, the functor $\text{cod}$ is a fibration as well where lifts are constructed via pullbacks.

In Sect. 6.1 we will heavily use the following construct.

**Definition 31 (Diagonal arrow categories).** In an arrow category $[\rightarrow \cdot, C]$, a sq-arrow $(f, f')$: $a \Rightarrow b$ is called diagonal if there is a “diagonal” arrow $d$: $a \circ b \Rightarrow a \circ b$ such that $f = a \circ d$, $f' = d \circ b$; then we write $(f, f') = d^z$.

The associativity of composition ensures the following result.

**Lemma 7** Composition of two diagonal arrows is again diagonal: if $(f_1, f'_1) = d_1^z: a \Rightarrow b$ and $(f_2, f'_2) = d_2^z: b \Rightarrow c$, then $(f_1, f'_1) \circ (f_2, f'_2) = (d_1 \circ b \circ d_2)^z$.

**Definition 32 (Diagonal subcategories).** Given a category $C$, category $[\rightarrow \cdot, C]_{\text{diag}}$ has the same objects as $[\rightarrow \cdot, C]$, diagonal sq-arrows as arrows between different objects (i.e., different $C$-arrows), and the only looping sq-arrows $(f, f')$: $a \Rightarrow a$ being (non-diagonal) identity sq-arrows $f = \text{id}(\text{dom} a)$, $f' = \text{id}(\text{cod} a)$.

Category $[\rightarrow \cdot, C]_{\text{diag}}$ is a full subcategory of $[\rightarrow \cdot, C]_{\text{diag}}$, whose objects are monic arrows in $C$, and we have the following obvious commutative diagrams:

\[
\begin{array}{ccc}
[\rightarrow \cdot, C]_{\text{diag}} & \xrightarrow{[\text{wide}]} & [\rightarrow \cdot, C] \\
\downarrow \text{dom}' & & \downarrow \text{dom} \\
[\rightarrow \cdot, C]_{\text{diag}} & \xrightarrow{[\text{full}]} & [\rightarrow \cdot, C]
\end{array}
\quad
\begin{array}{ccc}
[\rightarrow \cdot, C]_{\text{diag}} & \xrightarrow{[\text{wide}]} & [\rightarrow \cdot, C] \\
\downarrow \text{cod}' & & \downarrow \text{cod} \\
[\rightarrow \cdot, C]_{\text{diag}} & \xrightarrow{[\text{full}]} & [\rightarrow \cdot, C]
\end{array}
\]

with labels $[\text{wide}]$ and $[\text{full}]$ referring to wide and full embedding resp. Note that functor $\text{cod}'$ is faithful: if two parallel diagonals are different, then their compositions with the target monic arrow are different too. To ease notation, below we will denote all four $\text{dom}$-functors and all four $\text{cod}$-functors by name $\text{dom}$ and $\text{cod}$, resp., without primes.