Normal Forms and Reduction for Theories of Binary Relations

Dan Dougherty and Claudio Gutiérrez

Computer Science Group, Wesleyan University
Middletown CT 06459 USA
{ddougherty,cgutierrez}@wesleyan.edu

Abstract. We consider equational theories of binary relations, in a language expressing composition, converse, and lattice operations. We treat the equations valid in the standard model of sets and also define a hierarchy of equational axiomatisations stratifying the standard theory. By working directly with a presentation of relation-expressions as *graphs* we are able to define a notion of reduction which is confluent and strongly normalising, in sharp contrast to traditional treatments based on first-order terms. As consequences we obtain unique normal forms, decidability of the decision problem for equality for each theory. In particular we show a non-deterministic polynomial-time upper bound for the complexity of the decision problems.

1 Introduction

The theory of binary relations is a fundamental conceptual and methodological tool in computer science. The formal study of relations was central to early investigations of logic and the foundations of mathematics [11, 20, 24, 25, 26] and has more recently found application in program specification and derivation, [2, 6, 4, 18] denotational and axiomatic semantics of programs, [8, 10, 22, 19] and hardware design and verification [7, 16].

The collection of binary relations on a set has rich algebraic structure: it forms a monoid under composition, each relation has a converse, and it forms a Boolean algebra under the usual set-theoretic operations. In fact the equational theory in this language is undecidable, since it is possible to encode set theory [26]. Here we eliminate complementation as an operation, and investigate the set $E_{\mathcal{R}}$ of equations between relation-expressions valid when interpreted over sets, as well as certain equational axiomatic theories naturally derived from $E_{\mathcal{R}}$.

Now, the most popular framework for foundations and for implementations of theorem provers, proof-checkers, and programming languages remains the λ-calculus. It seems reasonable to say that this is due at least in part to the fact that the equational theory of λ-terms admits a computational treatment which is well-behaved: 'b-reduction is confluent, and terminating in typed calculi, so that the notion of *normal form* is central to the theory.

To our knowledge, no analogous notion of normal form for terms in $E_{\mathcal{R}}$ is known. In fact the calculus of relations has a reputation for being complex.

L. Bachmair (Ed.): RTA 2000, LNCS 1833, pp. 95–109, 2000.

Bertrand Russell (quoted in [21]) viewed the classical results of Peirce and Schröder on relational calculus as being "difficult and complicated to so great a degree as to doubt their utility." And in their recent monograph [4, page 81] Bird and de Moor observe that "the calculus of relations has gained a good deal of notoriety for the apparently enormous number of operators and laws one has to memorise in order to do proofs effectively."

But in this paper we suggest that a rather attractive syntactic/computational treatment of the theory of relations is indeed available, at least for the fragment of the theory not including complementation.

The essential novelty derives from the idea of taking certain graphs as the representation of relations. These graphs, called here "diagrams," arise very naturally and have been used since Peirce by researchers in the relation community (e.g. Tarski, Lyndon, Jónsson, Maddux, etc.); recent formalisations appear in [12, 1, 7]. What we do here is to take graphs seriously as a notation alternative to first-order terms, i.e., to treat diagrams as first-class *syntactic* entities, and specifically as candidates for *rewriting*.

One can see diagram rewriting as an instance of a standard technique in automated deduction. It is well-known that certain equations inhibit classical term-rewriting techniques — the typical examples are associativity and commutativity — and that a useful response can be to pass to computing *modulo* these equations. In Table 1 we exhibit a set $E_{\mathcal{D}}$ of equations such that diagrams are the natural data structure for representing terms modulo $E_{\mathcal{D}}$.

Summary of Results

It is not hard to see that in the absence of complementation equality between relation-expressions can be reduced to equality between expressions not involving union, essentially because union distributes over the other operations. So we ultimately restrict attention to the complement- and union-free fragment of the full signature (see Definition 1). It is known [1, 12] that the set of equations true in set-relation algebras in this signature is decidable.

We clarify the relationship between terms and diagrams by showing that the algebra of diagrams is precisely the free algebra for the set $E_{\mathcal{D}}$ of equations between terms. It is rather surprising that a finite set of equations accounts for precisely the identifications between terms induced by compiling them into diagrams.

Freyd and Scedrov [12] isolated the theory of *allegories*, a finitely axiomatisable subtheory of the theory of relations which corresponds to a certain geometrically-motivated restricted class of morphisms between diagrams. We refine this by constructing a proper hierarchy of equational theories, beginning with the theory of allegories, which stratifies the equational theory of set-relations.

Our main result is a computational treatment of diagrams via a notion of reduction. Actually each of the equational theories in the hierarchy induces its own reduction relation; but we prove *uniformly* that each reduction satisfies strong normalisation and Church-Rosser properties. Therefore each theory enjoys

unique (diagram-) normal forms and decidability. In fact the decision problem for each theory is in NP, non-deterministic polynomial time.

We feel that the existence of computable unique normal forms is our most striking result. The virtue of treating diagrams as syntax is highlighted by the observation that $E_{\mathcal{R}}$ is not finitely axiomatisable [15], so no finite *term* rewriting system can even claim to correctly present the theory, much less be a convergent presentation.

In light of the characterisation of the set of diagrams as the free algebra for the set $E_{\mathcal{D}}$ of equations, these results can be seen — if one insists — as results about rewriting of terms modulo $E_{\mathcal{D}}$. But for us the diagram presentation is the primary one and is ultimately the closest to our intuition.

Related Work

The case for using a calculus of relations as a framework for concepts and methods in mathematics and computer science is compellingly made by Freyd and Scedrov in [12]. They define allegories as certain *categories*; the structures modeled in this paper are in that sense one-object allegories. One may view this as the distinction between typed and untyped calculi.

Bird and de Moor's book [4] is an extended presentation of the application of relational calculus to program specification and derivation, building explicitly on the theory of allegories. There, terms in relation calculus are not programs *per se*, but the authors do raise the question of how one might *execute* relation-expressions [4, page 110]. As noted there, a promising proposal is made by Lipton and Chapman in [18], where a notion of rewriting terms using the allegory axioms is presented. It should be very interesting to explore the relationship between the Lipton-Chapman model and the one presented here.

Brown and Hutton [7] apply relational methods to the problems of designing and verifying hardware circuits. They observe that people naturally reason informally about pictures of circuits and seek to provide formal basis, again based on allegories, for such reasoning; their vehicle is the relational language RUBY used to design hardware circuits. To our knowledge they do not claim decidability or normal forms for the theory they implement. An implementation of their method is distributed at [16].

Two other investigations of graphical relation-calculi are the work of Kahl [17] and that of Curtis and Lowe [9].

The general topic of diagrammatic reasoning has been attracting interest in several areas lately (see for example [3]). The present research might be viewed as a case-study in reasoning with diagrams in the general sense.

Further indication of the range of current investigations into relations and relation-calculi may be found in, for example, the books [23] or [5] or the proceedings of the roughly annual RelMiCS conferences.

2 Preliminaries

Definition 1. *The signature Σ is composed of the binary operations* intersection $\cap$ *and* composition $;$ *(usually written as concatenation), two unary operations* converse $(\,)^\circ$ *and* domain dom*, and a constant* 1*.*

When we exhibit terms, the composition is to be interpreted in "diagrammatic order" so that xy means "x first then y. The operation dom has a natural interpretation as the domain of a relation, and under this interpretation it is definable from the other operations as $\mathrm{dom}\, x = 1 \cap xx^\circ$. The inclusion of dom in the signature is non-traditional, but there is a very good technical reason for its inclusion, made clear in the remark following Theorem 2.

The standard models are sets and binary relations; the following definition is from [1].

Definition 2. *A* (subpositive) set relation algebra *is an algebra of the form $\langle A, \cap, ;, (\,)^\circ, \mathrm{dom}, 1\rangle$ where A is a set of binary relations on some base set closed under the operations, which have the standard relational meaning (1 being the identity relation).*

By $\mathcal{R}$ we will denote the class of algebras isomorphic to set relation algebras and by $E_{\mathcal{R}}$ the set of equations valid in $\mathcal{R}$.

Definition 3. *A* undirected graph *g is a pair (V_g, E_g) of sets (vertices and edges) together with a map $E_g \longrightarrow [V_g]^2$, where the elements of $[V_g]^2$ are the 2-element multisets from V. Such a graph is* connected *if there is a path between any two vertices. An undirected graph h is a* minor *of g if h can be obtained from a subgraph g' of g by a sequence of contractions of vertices of g'.*

The notion of directed graph *is obtained by replacing $[V]^2$ by $V \times V$ in the definition above; note that any directed graph g obviously has an undirected graph underlying it. A directed graph g is* labelled by a set X *if there is a function $l(g) : E_g \longrightarrow X$. We will be interested in this paper in directed labelled graphs g with a distinguished* start vertex s_g *and a distinguished* finish vertex f_g*. We allow these to be the same vertex.*

For the sake of brevity the term *graph* will always mean: directed, labelled graph with distinguished start and finish vertices, whose underlying undirected graph is connected.

Let $\mathcal{G}$ denote the set of such graphs. Strictly speaking the set $\mathcal{G}$ depends on the particular set of labels chosen, but this set will never change in the course of our work, so we suppress mention of the label-set in the notation. We do assume that the set of labels is infinite.

A *morphism* φ between graphs g and h is a pair of functions $\varphi_V : V_g \longrightarrow V_h$ and $\varphi_E : E_g \longrightarrow E_h$ which

- preserves edges and direction, *i.e.*, for all $v, w \in V_g$, if e is an edge in g between v and w, then $\varphi_E(e)$ is an edge in h between $\varphi_V(v)$ and $\varphi_V(w)$,
- preserves labels, *i.e.*, for all $e \in E_g$, $l(e) = l(\varphi_E(e))$, and
- preserves start and finish vertices, *i.e.*, $\varphi_V(s_g) = s_h$ and $\varphi_V(f_g) = f_h$.

If it is clear from context, we will simply write φ instead of φ_V or φ_E.

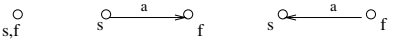

Fig. 1. The distinguished graphs 1, 2_a, and 2_a^{-1}.

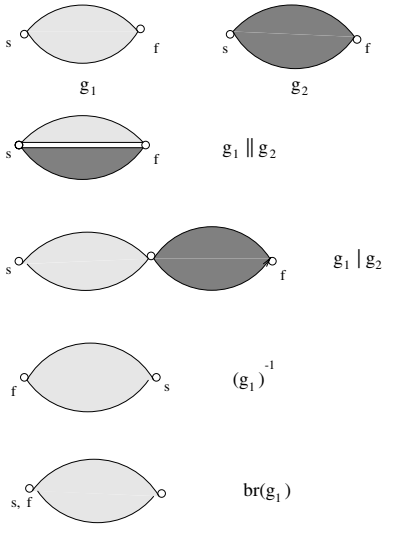

Fig. 2. Operations on graphs

2.1 Diagrams for Binary Relations

Here we introduce the central notion of *diagram* for a relational term. The set of diagrams supports an algebraic structure reflecting that of relations themselves, and a categorical structure in which morphisms between diagrams correspond to equations between relational terms valid in all set relation algebras. This material is standard and provides a foundation for our work in the rest of the paper.

There are some distinguished graphs in $\mathcal{G}$. The graph with only one vertex which is at the same time the start and finish, and no edges, will be denoted by 1. The graph with edge labelled a from the start vertex to the (distinct) finish vertex is denoted 2_a; the graph obtained by reversing the sense of the edge is $2_{a^{-1}}$. (See Figure 1.)

Definition 4. *Let g, g_1, g_2 be graphs in $\mathcal{G}$. We define the following operations in $\mathcal{G}$ (see Figure 2 for a graphical presentation.)*

1. *The* parallel composition*, $g_1 \| g_2$, is the graph obtained by (1) identifying the start vertices of the graphs g_1, g_2 (this is the new start), and (2) identifying the finish vertices of the graphs g_1, g_2 (the new finish).*
2. *The* sequential composition*, $g_1 | g_2$, is the graph obtained by identifying the finish of g_1 with the start of g_2, and defining the new start to be s_{g_1} and the new finish to be f_{g_2}.*

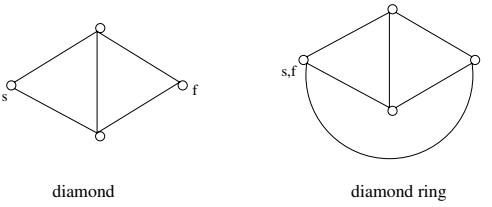

Fig. 3. Some graphs not in $\mathcal{D}$. Note the vertices s, f in each case.

3. *The* converse *of g, denoted by g^{-1}, is obtained from g by interchanging its start and finish. It is important to note that neither labels nor direction of edges changes.*
4. *The* branching *of g, denoted by* $\mathrm{br}(g)$*, is the graph obtained from g by redefining its finish to be the same as the start.*

Not every graph in $\mathcal{G}$ can be built using these operations. Figure 3 gives two examples (the edges in these pictures can be directed at will). Further significance of these graphs is given in Theorem 5.

Definition 5. *Let $\mathcal{D}$ denote the the set of graphs generated by 1, the 2_a, and the operations of branching and sequential and parallel composition.*

The set $\mathcal{D}$ is a Σ-algebra in a natural way; Theorem 2 below says more. $\mathcal{D}$ will play a key role in the normalisation process and has interesting properties in its own right.

Let $T_\Sigma(X)$ be the set of first-order terms over Σ with the labels X as variables. Then there is a surjective homomorphism

$$T_\Sigma(X) \longrightarrow \mathcal{D}, \quad t \mapsto g_t$$

defined recursively by $g_1 = 1$, $g_a = 2_a$ for $a \in X$, $g_{t_1;t_2} = g_{t_1} \mid g_{t_2}$, $g_{t_1 \cap t_2} = g_{t_1} \| g_{t_2}$, $g_{t^\circ} = (g_t)^{-1}$, and $g_{\mathrm{dom}\, t} = \mathrm{br}(g_t)$.

We can see the power of diagrams in the following important representation theorem. Recall that $\mathcal{R}$ denotes the class of algebras isomorphic to subpositive set relation algebras.

Theorem 1 ((Freyd-Scedrov, Andreka-Bredikhin)). *Let r, t be terms in the signature Σ. Then the equation $r = t$ is valid in $\mathcal{R}$ if and only if there are morphisms $g_r \longrightarrow g_t$ and $g_t \longrightarrow g_r$.*

Proof. The relationship between graphs in $\mathcal{D}$ and set relation algebras goes as follows. For an algebra $\mathcal{A}$ with base A and relations $R_1^{\mathcal{A}}, \ldots, R_n^{\mathcal{A}}$ of $\mathcal{A}$ define the graph $g_{\mathcal{A},\bar{R}}$ to have the set of vertices A and an edge (a, b) with label j for each $(a, b) \in R_j^{\mathcal{A}}$. Observe that $g_{\mathcal{A},\bar{R}}$ is not necessarily in $\mathcal{D}$. By an induction on terms it can be proved that for each term t in $T_\Sigma(R_1, \ldots, R_n)$ and elements $a, b \in A$ it holds $(a, b) \in t^{\mathcal{A}}[\bar{R}]$ if and only if there is a $\mathcal{G}$-morphism $g_t \longrightarrow g_{\mathcal{A},\bar{R}}$ which takes s to a and f to b.

The statement of the theorems follows now easily. ///

The strength of Theorem 1 is to reduce equational reasoning in binary relations to reasoning about graph theoretical morphisms. In particular since diagrams are finite one can check whether or not there are morphisms between two given ones, so we have a decision procedure for equality (in $E_{\mathcal{R}}$).

This result can be improved in at least two directions from a computational point of view:

- Refine the morphisms in order to stratify the equations, hence possibly getting better computational tools for interesting fragments.
- Investigate rewrite systems and normal forms in this new representation.

We pursue these directions in the following two sections. In fact the developments are independent of one another, so that the reader interested primarily in diagram-rewriting can on a first reading proceed directly to Section 4.

3 Terms and Equations as Diagrams and Morphisms

Theorem 1 shows that morphisms between diagrams reflect equations valid in the theory of binary relations. Unfortunately it puts all these valid equations in one sack. Experience shows that certain equations appear more often than others in practice and are in some sense are more fundamental. Our program in this section is to classify equations by their operational meaning.

3.1 Equational Characterisation of $\mathcal{D}$

$$
\begin{array}{rr}
x1 = x & 1^\circ = 1 \\
x(yz) = (xy)z & 1 \cap 1 = 1 \\
x \cap y = y \cap x & (1 \cap x)(1 \cap y) = (1 \cap x \cap y) \\
x \cap (y \cap z) = (x \cap y) \cap z & x \cap y(1 \cap z) = (x \cap y)(1 \cap z) \\
x^{\circ\circ} = x & 1 \cap x(y \cap z) = 1 \cap (x \cap y^\circ)z \\
(xy)^\circ = y^\circ x^\circ & \\
(x \cap y)^\circ = x^\circ \cap y^\circ & \mathrm{dom}\, 1 = 1 \\
 & (\mathrm{dom}\, x)^\circ = \mathrm{dom}\, x \\
 & \mathrm{dom}((x \cap y)z) = 1 \cap x(\mathrm{dom}\, z)y^\circ \\
 & \mathrm{dom}((\mathrm{dom}\, x)y) = (\mathrm{dom}\, x) \cap (\mathrm{dom}\, y)
\end{array}
$$

Table 1. The equations $E_{\mathcal{D}}$.

To start with, the class $\mathcal{D}$ itself embodies certain equations in the sense that each graph in $\mathcal{D}$ can come from several different terms. It is interesting that these identifications can be axiomatised by a finite set of equations, $E_{\mathcal{D}}$, shown in Table 1. The equations in the left column capture the essential properties of

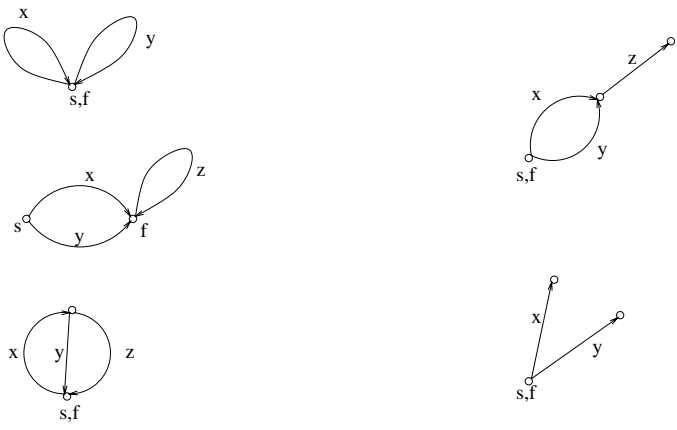

Fig. 4. Graphs representing some terms in the equations in Table 1. The ones in the left correspond to the last three equations about $1 \cap x$. The ones on the right to the last two equations about dom. Observe that in each of these equations the left- and right- hand side terms are represented by the same graph.

the operators (associativity, commutativity, and the involutive laws for converse), those in the upper right hand deal with identifications among terms of the form $1 \cap x$, and the rest take care of the identification of terms which contain the operator dom. All of these equations are trivially valid in $\mathcal{D}$: their left- and right-hand sides compile to the same diagram.

Theorem 2. *$\mathcal{D}$ is the free algebra over the set of labels for the set of equations $E_{\mathcal{D}}$.*

Proof. For the non-trivial direction we have to show that if r, t are not provably equal under $E_{\mathcal{D}}$, then $g_r \neq g_t$. This is done by proving that $E_{\mathcal{D}}$ can be completed by a finite number of equations into a confluent and terminating rewrite system modulo the equations for associativity of composition, AC of intersection and $1 \cap x(y \cap z) = 1 \cap (x \cap y^\circ)z$. A complete proof is in [14]. ///

3.2 Equations Capturing Morphisms

Freyd and Scedrov in [12] made the observation (without proof) that $\mathcal{D}$-morphisms which collapse at most two vertices at a time correspond to a simple and natural equational theory, an abstract theory of relations, the theory of *allegories*. Motivated by this idea we introduce a proper hierarchy of equational theories, stratifying $E_{\mathcal{R}}$ in terms of complexity of the morphisms acting on the data, each of which has a geometric as well as algebraic aspect.

Definition 6. *Let $\varphi : g_1 \longrightarrow g_2$ be an arrow in $\mathcal{D}$. We call φ an* n-arrow *if* $|V_{g_1}| \leq |V\varphi(g_1)| + n$.

$$
\begin{aligned}
x \cap x &= x \\
x(y \cap z) &= x(y \cap z) \cap xy \\
xy \cap z &= (x \cap zy^{\circ})y \cap z.
\end{aligned}
$$

Table 2. The operational equations

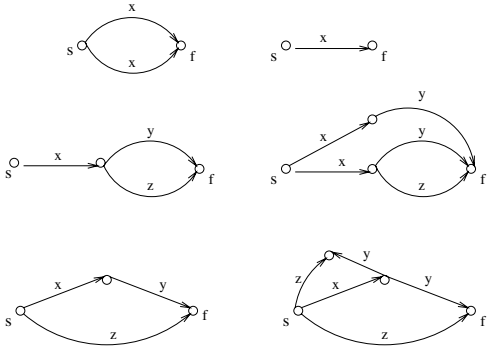

Fig. 5. The graph representations of the equations in Table 2.

Note that in general the composition of n-arrows is not an n-arrow. This motivates the following definition.

Definition 7. *Let $n \geq 0$ be a natural number.*

1. *The category $\mathcal{D}^n$ is the category whose objects are the graphs in $\mathcal{D}$, and whose arrows are finite compositions of n-arrows in $\mathcal{D}$.*
2. *The theory $E^n_{\mathcal{R}}$ is the set of equations $r = t$ between Σ-terms such that $g_r \longrightarrow g_t$ and $g_t \longrightarrow g_r$ in $\mathcal{D}^n$.*

We have the following chain of inclusions of categories: $\mathcal{D}^0 \subseteq \mathcal{D}^1 \subseteq \cdots \subseteq \mathcal{D}$. It can be shown that if $n \geq 1$ then $E^n_{\mathcal{R}}$ is closed under deduction. So we have a hierarchy of equational theories $E^0_{\mathcal{R}} \subseteq E^1_{\mathcal{R}} \subseteq \cdots \subseteq \bigcup_i E^i_{\mathcal{R}}$.

Theorem 1 can now be rephrased as $E_{\mathcal{R}} = \bigcup_i E^i_{\mathcal{R}}$.

The equational theory of *allegories* is presented by the axioms in the left column of Table 1 plus the ones shown in Table 2. These last three equations correspond (in the sense of Theorem 1) to 1-arrows over graphs in $\mathcal{D}$, as can be checked from the graphical representation in Figure 5. The next theorem formalises the converse statement i.e., that they are sufficient to axiomatise the equations obtained from 1-arrows.

Theorem 3. *$E^1_{\mathcal{R}}$ is exactly the equational theory of allegories.*

Proof. The proof is delicate and we do not present it here; a complete proof can be found in [14]. ///

Theorem 4. *For each $n \geq 2$ it holds $E^{2n-1}_{\mathcal{R}} \subset E^{2n}_{\mathcal{R}}$ (the inclusion is proper).*

Proof. An adaptation of the technique in [12, 2.158] showing that $E_{\mathcal{R}}$ is not finitely axiomatisable. ///

Unfortunately we do not know yet know much more about each of the steps in the hierarchy. But we conjecture that $E^1_{\mathcal{R}} = E^2_{\mathcal{R}} = E^3_{\mathcal{R}}$. We also conjecture that for every $n \geq 2$, the equational theory $E^n_{\mathcal{R}}$ is finitely axiomatisable. We do not know the answer to the question: "is $E^{2n}_{\mathcal{R}} = E^{2n+1}_{\mathcal{R}}$ for $n \geq 1$?"

4 Normalisation

This section presents a very general combinatorial lemma concerning the set of functions over finite structures viewed as an abstract reduction system. It is most conveniently presented using the language of categories, but no more than rudimentary category theory is required for the presentation. The material in this section is condensed from [13].

In this section juxtaposition denotes composition of arrows in a category, and is to be read in the standard way, so that fg means "g first." We use $A \cong B$ to indicate that A and B are isomorphic.

Definition 8. *Let $\mathcal{C}$ be any category, and A, B objects of $\mathcal{C}$. Define the relation $\rightleftarrows$ between objects of $\mathcal{C}$ as follows:*

$$A \rightleftarrows B \text{ if and only if there are arrows } A \longrightarrow B \text{ and } B \longrightarrow A.$$

Clearly $\rightleftarrows$ is an equivalence relation. Our goal is to find simple conditions which make the relationship $\rightleftarrows$ decidable.

The following notion is motivated by the observation that a (set-theoretic) function f between sets A and B can be seen as an map onto its image $f(A)$ followed by the inclusion of $f(A)$ into B.

Definition 9. *An arrow m is* mono *if whenever $ma = mb$ then $a = b$. An arrow e is* epi *if whenever $ae = be$ then $a = b$.*

An arrow $A \xrightarrow{f} B$ has an epi-mono factorisation *if there exist arrows e epi and m mono such that $f = me$. A category $\mathcal{C}$* has epi-mono factorisation *if every arrow in $\mathcal{A}$ has such a factorisation.*

Definition 10. *An object A is* hom-finite *if the set $Hom(A, A)$ of maps from A to A is finite. A category $\mathcal{C}$ is* hom-finite *if each object of $\mathcal{C}$ is hom-finite.*

Of course any concrete category of sets of finite objects will be hom-finite. In particular, the categories $\mathcal{G}$, $\mathcal{D}$ and $\mathcal{D}^n$ for each n are each hom-finite.

Lemma 1. *Suppose that A is hom-finite. If $m : A \longrightarrow A$ is a monomorphism, then m is an isomorphism. Also, If there are monomorphisms $m_1 : A \longrightarrow B$ and $m_2 : B \longrightarrow A$, then A and B are isomorphic.*

Proof. For the first assertion: consider the monomorphisms $m, m^2, \cdots : A \longrightarrow A$. Because A is hom-finite, there are integers $i < j$ such that $m^i = m^j$. Now, using the fact that m is mono, $1_A = m^k$ for $k = j - i$. Thus m is a monomorphism with a right inverse; this implies that m is an isomorphism.

For the second claim: we have that $m_2 m_1 : A \longrightarrow A$ is mono, hence, by the first part it is an isomorphism. Hence there exists f with $m_2 m_1 f = 1_A$. Thus m_2 is a monomorphism with a right inverse; this implies that it is an isomorphism. ///

Definition 11. *Let A, B be objects in a category $\mathcal{C}$. Define $A \Longrightarrow B$ if and only if B is both (the target of) a quotient- and (the source of) a sub-object of A; that is, there is an epimorphism e and a monomorphism m such that*

$$A \xrightarrow{e} B \xrightarrow{m} A.$$

We also require that e not be an isomorphism.

By $\overset{*}{\Longrightarrow}$ we will denote the reflexive-transitive closure of $\Longrightarrow$, where reflexivity is defined up to isomorphism. Thus $A \overset{*}{\Longrightarrow} B$ means either $A \cong B$ or there is a finite sequence $A \Longrightarrow C_1 \Longrightarrow \cdots \Longrightarrow C_n \Longrightarrow B$.

Lemma 2. *If A is hom-finite then there are no infinite $\Longrightarrow$-reductions out of A.*

Proof. For sake of contradiction suppose there were such a sequence. For each i we have maps $A_i \xrightarrow{e_i} A_{i+1} \xrightarrow{m_i} A_i$, and so we may define $a_i = m_1 \cdots m_i e_i \cdots e_1 : A \to A$. Since A is hom-finite there are $i < j$ with $a_i = a_j$. Cancelling the monos on the left and the epis on the right, we have $1_A = m_{i+1} \cdots m_j e_j \cdots e_{i+1}$. This implies that e_{i+1} is iso, a contradiction. ///

Observe that $A \overset{*}{\Longleftrightarrow} B$ implies that $A \rightleftarrows B$. The converse need not be true in general, but the next result provides a strong converse in certain categories.

Proposition 1. *Suppose $\mathcal{C}$ is a hom-finite category with epi-mono factorisation. Then if $A \rightleftarrows B$, then there exists C such that $A \overset{*}{\Longrightarrow} C \overset{*}{\Longleftarrow} B$.*

Proof. The proof is by Noetherian induction over $\Longrightarrow$, out of the multiset $\{A, B\}$. We are given $A \xrightarrow{f} B \xrightarrow{g} A$. If both f and g are mono then by Lemma 1 A and B are isomorphic and we may take C to be A. Otherwise, by symmetry we may suppose f is not mono without loss of generality. Factor the arrow gf as epi-mono, obtaining: $A \xrightarrow{e} X \overset{m}{\hookrightarrow} A$. Now, e is not mono, otherwise gf would be, contradicting the assumption that f is not mono. In particular e is not iso, and so $A \Longrightarrow X$. Since $X \rightleftarrows B$ we may apply the induction hypothesis to $\{X, B\}$, obtaining C with $A \Longrightarrow X \overset{*}{\Longrightarrow} C \overset{*}{\Longleftarrow} B$ as desired. ///

The previous results imply that the relation $\Longrightarrow$ is a terminating and confluent abstract reduction system capturing the equivalence relation $\rightleftarrows$:

Corollary 1 ((Normal Forms for $\rightleftarrows$)). *Suppose $\mathcal{C}$ is a hom-finite category with epi-mono factorisation.*

If $A \rightleftarrows B$, then there is a C, unique up to isomorphism, such that $A \overset{*}{\Longrightarrow} C$ and $B \overset{*}{\Longrightarrow} C$ and C is $\Longrightarrow$-irreducible.

Proof. By Lemma 2 we may let C and C' be any $\Longrightarrow$-irreducible objects such that $A \overset{*}{\Longrightarrow} C$ and $B \overset{*}{\Longrightarrow} C'$ respectively. Then $C \rightleftarrows C'$. But Proposition 1 and the irreducibility of C and C' imply that C and C' are isomorphic. ///

Observe that in the preceding Corollary we have $A \rightleftarrows C$ and $B \rightleftarrows C$. Also note that by taking B to be A we may conclude that for each A there is a C, unique up to isomorphism such that $A \overset{*}{\Longrightarrow} C$ and C is $\Longrightarrow$-irreducible. We refer to such a C as a"$\Longrightarrow$-normal form for A".

5 Normal Forms for Diagrams

We want to apply the results of the previous section to the categories $\mathcal{D}$ and $\mathcal{D}^n$. The following facts about $\mathcal{D}$ and each $\mathcal{D}^n$ are easy to check: (i) a map is epi if and only if it is surjective on vertices and on edges, and (ii) a map is an isomorphism if it is bijective on vertices and on edges. The next result is deeper; a proof can be found in [14].

Theorem 5. *Let $g \in \mathcal{G}$. Then g is in $\mathcal{D}$ if and only if the underlying undirected graph of g does not have diamond and diamond ring (see Figure 3) as minors.*

In particular, the set of graphs $\mathcal{D}$ is closed under the formation of subgraphs, i.e., if $g \in \mathcal{D}$ and h is a connected subgraph containing s_g and f_g then $h \in \mathcal{D}$.

Observe that the categories $\mathcal{D}^n$ have the same objects as $\mathcal{D}$ so the above theorem applies immediately to the $\mathcal{D}^n$. Theorem 5 is crucial in verifying that $\mathcal{D}^n$ supports the techniques of the previous section.

Proposition 2. *The categories $\mathcal{D}$ and each $\mathcal{D}^n$ are hom-finite and has epi-mono factorisation.*

Proof. The first assertion is easy to see. For the second, let $\varphi : g_1 \longrightarrow g_2$ be an arrow in $\mathcal{D}^n$. Then the graph $\varphi(g_1)$ is a subgraph of $g_2 \in \mathcal{D}$, hence by Theorem 5 it is also in $\mathcal{D}$. So we have $g_1 \xrightarrow{\varphi'} \varphi(g_1) \xrightarrow{i} g_2$, where φ' and i are epi and mono respectively. ///

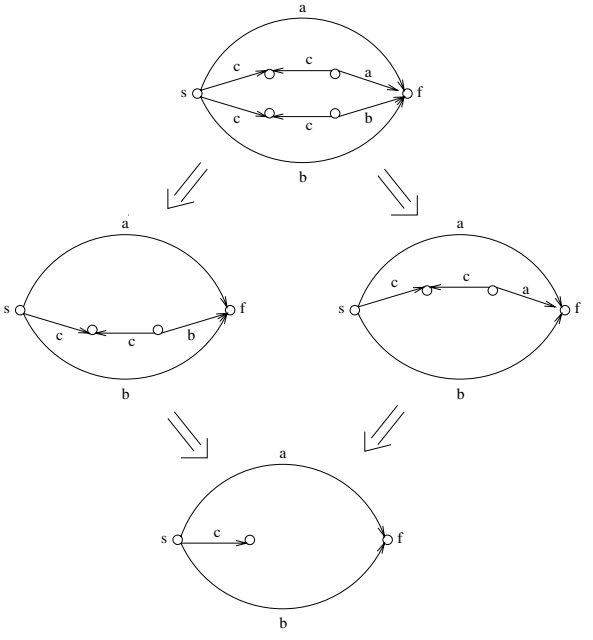

Fig. 6. A graph in $\mathcal{D}$ (top) and possible reductions. Observe that the reduction of the graphs in the middle to a common graph (bottom) depends on the existence of the operator br (dom).

Remark. It is precisely here that we see the benefit of our extended signature. The class of diagrams built over the traditional signature — without dom — does **not** have epi-mono factorisation, due to the fact that it is not closed under subgraphs. Figure 6 shows this by example.

We can now present our main results.

Theorem 6. *Let $\mathcal{C}$ be either $\mathcal{D}$ or one of the $\mathcal{D}^n$.*

If $g \rightleftarrows h$ then there is a k, unique up to isomorphism, such that $g \Longrightarrow k$ and $h \Longrightarrow k$ and k is $\Longrightarrow$-irreducible.

For each graph g, there is a graph $\mathrm{nf}(g)$ *such that* $\mathrm{nf}(g)$ *is $\Longrightarrow$-irreducible and such that for any h, $g \rightleftarrows h$ in $\mathcal{C}$ if and only if* $\mathrm{nf}(g) \cong \mathrm{nf}(h)$. *The graph* $\mathrm{nf}(g)$ *is unique up to isomorphism.*

Proof. This is an immediate consequence of Corollary 1 and Proposition 2. ///

It is important to note that the notions $\rightleftarrows$ and $\Longrightarrow$ in the previous Theorem are taken relative to the category ($\mathcal{D}$ or $\mathcal{D}^n$) one has chosen to work in.

Theorem 7. *For $\mathcal{D}$ and for each $\mathcal{D}^n$ the relation $\rightleftarrows$ is decidable in non-deterministic polynomial time. Each theory $E^n_{\mathcal{R}}$ is decidable in non-deterministic polynomial time.*

Proof. Decidability follows from the previous results and the fact that $\Longrightarrow$ is computable. To get the NP upper bound we examine the complexity of reduction to normal form. If $A \Longrightarrow A'$ then the sum of the number of vertices and edges of A must exceed that of A', since epimorphisms are surjections and bijections are isomorphisms (and we know the map from A to A' is not an isomorphism by definition). So any reduction of a diagram A to normal form takes a number of steps bounded by the size of A. So to test whether $A \rightleftarrows B$ we can generate sequences of morphisms reducing each of A and B — not necessarily to normal form — and test that the results are isomorphic. The latter test is of course itself in NP.

The second assertion follows immediately from the definition of $E^n_{\mathcal{R}}$. ///

6 Conclusion

We have examined the equational theory $E_{\mathcal{R}}$ of binary relations over sets and a family $E^n_{\mathcal{R}}$ of approximations to this theory. The theory $E^1_{\mathcal{R}}$ is Freyd and Scedrov's theory of allegories. By working with a natural notion of diagram for a relation-expression we have defined a notion of reduction of a diagram which yields an analysis of the theories above. A surprisingly important aspect was the inclusion of a "domain" operator in the signature: the corresponding operation is definable in terms of the traditional operations, but the class of diagrams for the enriched signature has better closure properties.

Since each notion of reduction of diagrams is terminating and confluent, we may compute unique normal forms for each of the theories. Each theory is therefore decidable and in fact normal forms can be computed in non-deterministic polynomial time.

The decidability of $E^n_{\mathcal{R}}$ is reminiscent of the decidability of $E_{\mathcal{R}}$, but has been more difficult to establish. This is because although equality in $E_{\mathcal{R}}$ is witnessed by any pair of graph morphisms between diagrams, equality in $E^n_{\mathcal{R}}$ is witnessed by a *sequence* of restricted morphisms. The length of this sequence could be bounded only after our work relating $\rightleftarrows$ and $\Longrightarrow$.

References

[1] H. Andréka and D.A. Bredikhin. The equational theory of union-free algebras of relations. *Algebra Universalis*, 33:516–532, 1995.

[2] R. C. Backhouse and P. F. Hoogendijk. Elements of a relational theory of datatypes. In B. Möller, H. Partsch, and S. Schuman, editors, *Formal Program Development*, volume 755 of *Lecture Notes in Computer Science*, pages 7–42. Springer-Verlag, 1993.

[3] J. Barwise and G. Allwein, editors. *Logical Reasoning with Diagrams.* Oxford University Press, 1996.

[4] R. Bird and O. de Moor. *Algebra of Programming.* Prentice Hall, 1997.

[5] C. Brink, W. Kahl, and G. Schmidt, editors. *Relational Methods in Computer Science.* Advances in Computing. Springer-Verlag, Wien, New York, 1997. ISBN 3-211-82971-7.

[6] P. Broome and J. Lipton. Combinatory logic programming: computing in relation calculi. In M. Bruynooghe, editor, *Logic Programming*. MIT Press, 1994.
[7] C. Brown and G. Hutton. Categories, allegories and circuit design. In *Logic in Computer Science*, pages 372–381. IEEE Computer Society Press, 1994.
[8] C.A.R.Hoare. An axiomatic basis for computer programming. *CACM*, 12(10):576–583, 1969.
[9] S Curtis and G Lowe. Proofs with graphs. *Science of Computer Programming*, 26:197–216, 1996.
[10] J. W. De Bakker and W. P. De Roever. A calculus for recursive program schemes. In M. Nivat, editor, *Automata, Languages and Programming*, pages 167–196. North-Holland, 1973.
[11] A. De Morgan. On the syllogism, no. IV, and on the logic of relations. *Transactions of the Cambridge Philosophical Society*, 10:331–358, 1860.
[12] P. Freyd and A. Scedrov. *Categories and Allegories*. North-Holland Mathematical Library, Vol. 39. North-Holland, 1990.
[13] C. Gutiérrez. Normal forms for connectedness in categories. *Annals of pure and applied logic*. Special issue devoted to the XI Simposio Latinoamericano de Logica Matematica, Venezuela, July 1998. To appear.
[14] C. Gutiérrez. *The Arithmetic and Geometry of Allegories: normal forms and complexity of a fragment of the theory of relations*. PhD thesis, Wesleyan University, 1999.
[15] M. Haiman. Arguesian lattics which are not linear. *Bull. Amer. Math. Soc.*, 16:121–123, 1987.
[16] G. Hutton, E. Meijer, and E. Voermans. A tool for relational programmers. Distributed on the mathematics of programming electronic mailing list, January 1994. `http://www.cs.nott.ac.uk/~gmh/allegories.gs`, 1994.
[17] Wolfram Kahl. Relational matching for graphical calculi of relations. *Journal of Information Science*, 119(3–4):253–273, December 1999.
[18] J. Lipton and E. Chapman. Some notes on logic programming with a relational machine. In Ali Jaoua, Peter Kempf, and Gunther Schmidt, editors, *Relational Methods in Computer Science*, Technical Report Nr. 1998-03, pages 1–34. Fakultät für Informatik, Universität der Bundeswehr München, July 1998.
[19] R.D. Maddux. Relation-algebraic semantics. *Theoretical Computer Science*, 160:1–85, 1996.
[20] C. S. Peirce. *Collected Papers*. Harvard University Press, 1933.
[21] V.R. Pratt. *Origins of the calculus of binary relations*, pages 248–254. IEEE Computer Soc. Press, 1992.
[22] J. G. Sanderson. *A Relational Theory of Computing*, volume 82 of *Lecture Notes in Computer Science*. Springer-Verlag, 1980.
[23] G. W. Schmidt and T. Ströhlein. *Relations and Graphs: Discrete Mathematics for Computer Scientists*. EATCS Monographs on Theoretical Computer Science. Springer-Verlag, 1993.
[24] E. Schröder. *Vorlesungen über der Algebra der Logik, Vol. 3, "Algebra und Logik der Relative"*. 1895.
[25] A. Tarski. On the calculus of relations. *Journal of Symbolic Logic*, 6(3):73–89, 1941.
[26] A. Tarski and S. Givant. *A formalization of set theory without variables*. AMS Colloquium Publications, Vol. 41. American Mathematical Society, 1988.