Relation Algebras of Intervals

Ph. D. Thesis of

Robin David Hirsch,

Imperial College of Science, Technology and Medicine

January 25, 1994

Acknowledgements

I am grateful to my supervisor, Dov Gabbay, for his constant advice to try the non-standard approach: I hope I have lived up to that. I am particularly indebted to Ian Hodkinson for his detailed criticism of the text as well as original contributions to the work. Many thanks to Mark Reynolds for valuable conversations on a range of topics, some of which were relevant to this thesis. Most of all I would thank my three lovely children for their consideration in allowing me to enjoy at least some nights of uninterrupted sleep while preparing this work.

Abstract

A technique of game theory is used to analyse the representability of relation algebras. Representations are characterised by classes of *networks*. We prove that for atomic relation algebras the existence of a class of networks with the triangle addition property together with the Tarski axioms, exactly characterises those relation algebras which possess a *complete representation*. This extends Lyndon's result which is for finite relation algebras only.

Two very natural properties of representations are defined: *homogeneity* and *universality*. These types of representations can be characterised by network properties: the *joint embedding* and *amalgamation properties*. The familiar relation algebras from temporal reasoning are shown to have these properties. ω-categorical relation algebra are characterised exactly using a *network game*.

An alternative approach to the problem of axiomatising relation algebra is to consider *non-standard* representations. A relation algebra can always be interpreted in three dimensions (instead of the usual two).

A method of building an interval algebra from a normal representable relation algebra is given. In this way, the Allen interval algebra can be built from the point algebra. Starting with the metric system of Dechter, Meiri and Pearl an interval algebra is constructed which combines the qualitative expressive power of Allen's interval algebra with quantitative, metric expressivity.

The Allen and Koomen planner is studied in detail and extended by using *domain variables* and by allowing it to work forwards and backwards. Applications of the theoretical work on relation algebra include a method of decomposing intractable networks and a way of solving the *collapsing problem* in the Allen and Koomen planner. This is closely related to problems in other planning architectures (e.g. *stretching* and *clipping* in TMM) and more generally to the persistence problem.

Contents

0.1 Introduction and Background

In 1847 Boole[2] published "The mathematical analysis of logic, being an essay towards a calculus of deductive reasoning" [Boole, 1948]. He takes traditional logic (at that time the Aristotelian syllogism) and treats the semantics by the new algebraic methods. In modern terms we can say that Boolean algebra captures exactly the models of first-order logic with only unary relations. The success of this project can be judged, in part, by the fact that Boolean algebra is now generally included in the maths curriculum for twelve to fifteen year olds. It was not long before an attempt was made to handle the logic of binary relations in a similar manner [De Morgan, 1964] and after the invention of quantified first-order logic an algebra of the semantics of such a language was also constructed, namely *cylindric algebra* (see [Németi, 1991, Henkin *et al.*, 1971, Henkin *et al.*, 1985] for a comprehensive overview and bibliography of both problems).

The former problem, to find algebraic axioms for models of a first-order language with binary relations, turned out to be more problematic than Boolean algebra. This subject, known as *Relation Algebra* is one of the main themes of this dissertation.

What is a model of this *dyadic* language? The intention is that the binary predicates are interpreted as sets of pairs over some domain D. The Boolean connectives $\vee$ and $\neg$ are interpreted as union and complement[3] of sets. We can also make use of the binary structure to define the *identity* Id - consisting of all pairs (d, d); the *converse* of a relation R defined by $(d, e) \in R^{\smile} \Leftrightarrow (e, d) \in R$ and the *composition* of two relations $(d, e) \in R * S \Leftrightarrow \exists z[(d, z) \in R \wedge (z, e) \in S]$.

Definition A *Proper Relation Algebra* consists of sets of pairs from some domain, D, including the identity and closed under union, relative complement, converse and composition. The class RRA (representable relation algebras) consists of all structures isomorphic to a proper relation algebra.

So, to repeat Boole's success with his algebra of unary relations, it was required to find a finite set of axioms valid over RRA such that any algebraic structure satisfying these axioms had a model in RRA. This requirement would be the analogue of Stone's Theorem [Stone, 1936] and such an axiomatisation would algebraically capture exactly the class RRA. Tarski proposed such an axiomatisation ([Tarski, 1941], see section 1.2.1 below). A structure satisfying these axioms will be referred to as an *abstract relation algebra* or simply as a *relation algebra* . However Lyndon [Lyndon, 1950] showed that this axiomatisation was not complete and Monk [Monk, 1964] went on to show that the class RRA could not be axiomatised by any finite set of axioms.

Instead Lyndon gave an infinite set of axioms, in the form of a schema, which he showed were necessary and sufficient for a finite relation algebra to be representable as a proper relation algebra.

---

[2] Read [Maddux, 1991] for the history of the study of Relation Algebras.

[3] It would be more accurate to talk of complement relative to some universal binary relation 1.

More recently there has been a renewed interest in relation algebras, but this time from the point of view of temporal reasoning where certain relation algebras are widely used (e.g. [Allen, 1983, Dechter *et al.*, 1991]). Ladkin and Maddux were the first to examine these temporal reasoners as examples of relation algebra [Ladkin and Maddux, 1987a]. Here new problems were thrown up, in particular it was found that while testing the consistency of a *network* in the *point algebra* can be done in cubic time performing a similar operation in the *Allen interval algebra* was intractable (see 1.1.1 for an exposition of the point and interval algebras and 1.3.3 for the definition of a network). However the point algebra is far less expressive than the interval algebra and so there appears to be a trade off between expressive power and tractability.

These temporal reasoners have been widely used in artificial planning where further difficulties arise. On the border between temporal reasoning and planning we encounter the temporal aspect of the *frame problem* the so-called *persistence problem.* In the context of the Allen interval algebra this is expressed as the *collapsing problem* - roughly, given two intervals where the same property is true such that the two intervals might consistently be identical, under what conditions are we permitted to assume that the two are actually equal? This leads us to a more general problem in planning: the search problem. Given a number of choices for different actions and different temporal orderings we generate a very large search space. One of the approaches to dealing with this is known as *dependency directed back-tracking* where the search space is searched depth first but when back-tracking becomes necessary we know exactly which choices must be altered and which can be left alone.

0.2 Outline

This thesis is divided into three parts: the pure theory of relation algebras, relation algebras of intervals and applications to planning. The dividing lines between the parts are not entirely clear-cut and much of the work in the first two parts is closely related to the application in planning.

In part I the Tarski axioms for an abstract relation algebra are given along with a definition of a representation of such an algebra. A *network* of a relation algebra is then defined, an object widely used in temporal reasoning. Properties of the atomic, transitively closed networks of a relation algebra are used to characterise when an algebra is representable. It is proved that for relation algebras with only countably many atoms, the existence of a class of networks with the triangle addition property together with the Tarski axioms exactly characterises those relation algebra which possess a complete representation. This is equivalent to the infinite axiom schema given by Roger Lyndon, thus his result is extended from the finite case to relation algebras with countably many atoms. The analysis of the representability of relation algebras is taken further by defining two properties that a representation may also possess: *homogeneity* and *universality.* The relation algebras with this type of representation can also be characterised by network properties: the *joint embedding* and *amalgamation properties.* The familiar relation algebras from temporal reasoning are shown to have these properties. The *countably categorical* relation algebras are then characterised exactly by considering certain *network games.*

This part is concluded with another approach to the problem of axiomatising relation algebra. The usual Tarski axioms are taken but *non-standard* representations are considered. In particular it is demonstrated that a relation algebra can always be interpreted in three dimensions (instead of the usual two), where each element of the algebra is represented by a ternary instead of a binary relation. Composition and converse are given non-standard definitions.

Part II starts with an application of the previous work by showing how from any relation algebra with a normal representation it is possible to construct an algebra of *intervals.* In this way, the Allen interval algebra can be built from the point algebra. In a further application to temporal reasoning, the qualitative expressive power of Allen's interval algebra is combined with the quantitative power of the metric system of Dechter, Meiri and Pearl by building an interval algebra from the latter, metric system.

The final section includes several applications of this work to problems in artificial planning. The principal example taken is the Allen and Koomen planner, though it is shown that many of the issues raised here translate to equivalent problems in different planning architectures, for example the event calculus and the Temporal Database Management system of Dean and McDermott.

Part I

Representations of Relation Algebras

Chapter 1

Standard Representations

1.1 Introduction

A number of temporal reasoning devices can be classified as relation algebras (see [Ladkin and Maddux, 1987a]). These include the Allen Interval Algebra [Allen, 1983], a point-based version of Allen's Algebra and the metric point system of Dechter, Meiri and Pearl [Dechter *et al.*, 1991]. These have been used extensively in planning [Allen and Koomen, 1983, Pelavin, 1988, Dean and McDermott, 1987]. However, the point-based systems suffer from a lack of expressive power and Allen's Interval Algebra is intractable and does not represent metric information. In this section all of these are shown to belong to a certain sub-class of the relation algebras - namely the relation algebras possessing complete, fully universal, homogeneous representations - normal representations for short. A characterisation of this class is given in terms of certain properties pertaining to networks - the joint embedding and amalgamation properties.

1.1.1 Some temporal reasoners

In [Allen, 1983, Allen, 1984] Allen defines thirteen primitive relations which can hold between two intervals: before, meets, overlaps, starts, during, ends, equals and the converses of the first six. The composition of two primitive interval relations is given in a table e.g.

$$overlaps * during = \{during, starts, overlaps\}$$

An *interval network* consists of a complete directed graph on a set of nodes with each arc labelled by a disjunction of primitive interval relations. An interval network is *consistent* if it is modelled by a set of intervals lying on a linear flow. To determine whether an interval network is consistent has been shown to be NP-complete [Villain and Kautz, 1986, Villain *et al.*, 1989]. The point-based version $\mathcal{P}$ is similar but here there are only three primitive point relations: $<$, $=$, $>$. The composition table is much simpler, for example $< * <= \{<\}, < * >= \{<, =, >\}$. It has been shown that the consistency checking problem for the point algebra can be solved by a cubic-time algorithm [Villain *et al.*, 1989]. The metric system $\mathcal{M}$ in [Dechter *et al.*, 1991] also uses networks, but here the arcs are labelled by an interval of rational numbers. The intended meaning of the arc

is that $a \leq (y - x) \leq b$ i.e. the directed distance from x to y lies inside the interval[1]. Composition of two constraints is calculated by $[a_1, b_1] * [a_2, b_2] = [a_1 + a_2, b_1 + b_2]$. Here again, provided there is only a single interval on each arc the problem of checking consistency can be done by the Floyd-Walshall algorithm in cubic time, but for a network with finite unions of intervals labelling the arcs the problem is NP hard. Throughout this thesis $\mathcal{I}$ will stand for the Allen Interval Algebra, $\mathcal{P}$ the point algebra and $\mathcal{M}$ the metric point system of [Dechter *et al.*, 1991].

1.2 Relation algebras

All of the above are examples of proper relation algebras (see definition on page 4.) They consist of binary relations and are closed under the Boolean operations, converse and composition. They all contain the identity relation Id.

1.2.1 Abstract Relation Algebras

Definition An abstract relation algebra, or simply a relation algebra (RA) $\mathcal{A}$ is a tuple $(A, \vee, -, 0, 1, *, ^\smile, Id)$ which satisfies the following axioms, essentially due to Tarski. For all $a, b, c \in \mathcal{A}$

1. $(A, \vee, -, 0, 1)$ is a Boolean algebra (1 is the universal element)
2. $*$ is an associative binary operator on A
3. $(a^\smile)^\smile = a$
4. $Id * a = a * Id = a$
5. $a * (b \vee c) = a * b \vee a * c$
6. $(a \vee b)^\smile = a^\smile \vee b^\smile$
7. $(a - b)^\smile = a^\smile - b^\smile$
8. $(a * b)^\smile = b^\smile * a^\smile$
9. $a * b \wedge c^\smile = 0 \Leftrightarrow b * c \wedge a^\smile = 0$.

---

[1] Open or semi-open intervals can also be dealt with.

The last axiom is not an equation but it is equivalent (in the presence of the other axioms) to the equation

$$a^{\smile} * -(a * b) \leq -b$$

where $a \leq b$ is an abbreviation for $a \vee b = b$ and the unary operation $-a$ is defined as $1 - a$. We also use the abbreviation $a \wedge b =_{def} a - (a - b)$.

It is easy to see that all these axioms are valid over the class RRA. It is natural to ask whether these axioms are complete over that class. Lyndon [Lyndon, 1950] showed that the answer was negative[2]: there are (abstract) relation algebras i.e. structures satisfying axioms 1 to 9 above which are not isomorphic to any concrete algebra of binary relations. Lyndon gives an infinite set of conditions which are valid in any representable relation algebra and shows that they are complete for finite relation algebras. These conditions are rather complex and in section 1.3 we present an equivalent, though conceptually simpler, set of conditions in terms of networks and show that this is complete for the countable case too.

Definitions

- An *atom* is a minimal non-zero element of $\mathcal{A}$. $\mathcal{A}$ is *atomic* if every non-zero element contains an atom. The set of all atoms of an RA $\mathcal{A}$ is denoted by $atoms_{\mathcal{A}}$.
- A *unit* is an atom contained in the identity relation.
- A relation algebra is *complete* if arbitrary unions (and intersections) exist. Here, an infinite union of a set of elements is taken to mean the least upper bound of that set, if it exists and is undefined otherwise.
- A relation algebra which obeys the axiom

  $$a * b = 0 \Leftrightarrow a = 0 \text{ or } b = 0$$

  is called *integral*. It is not hard to show that a relation algebra is integral if and only if the identity element is an atom.
- A relation algebra is *simple* if for every non-zero element r it satisfies

  $$1 * r * 1 = 1$$

- An element r is called *non-singular* if the map $\theta_r : a \to a * r$ is invertible or, equivalently, if $r * r^{\smile} = Id$.
- An atomic relation algebra is a *complex group* if every atom is *non-singular*. For a complex group $\mathcal{A}$ it follows that $(atoms_{\mathcal{A}}, *)$ forms a group (hence the name).

---

[2] McKenzie [McKenzie, 1970] provides the simplest example of a relation algebra which does not possess a representation. Monk [Monk, 1964] proves that the class RRA cannot be finitely axiomatised at all.

Jónsson and Tarski showed that every relation algebra is isomorphic to a subalgebra of a complete relation algebra. For the most part the relation algebras that are considered here are atomic. Every relation algebra can be decomposed into simple relation algebras [Jónsson and Tarski, 1948].

Examples

1. The metric point system $\mathcal{M}$ is a complex group. An atomic relation consists of a singleton interval $[d, d]$ and asserts that the directed distance between two points is exactly d. Since $[d, d] * [-d, -d] = [0, 0]$ (the identity relation) each atom must be invertible. $\mathcal{M}$ has infinitely many atoms.
2. The point algebra $\mathcal{P}$ does not form a complex group. Of its three atoms $\{=, <, >\}$ only the identity is non-singular ($< * > = 1$).

1.3 Representations and Networks

1.3.1 Representations

Definitions

- A *representation* of a relation algebra $\mathcal{A}$ is a pair (X, D) where D is any set - the domain of the representation - and X is an isomorphism from $\mathcal{A}$ into a proper relation algebra (page 4) on the set D.
- Such a representation is called a *complete representation* if arbitrary unions are preserved wherever they are defined i.e. if $\bigvee_\lambda a_\lambda$ exists then

$$X(\bigvee_\lambda a_\lambda) = \bigcup_\lambda (X(a_\lambda)).$$

Notes

- While thinking about complete representations it is worth clarifying the notion of infinite unions in boolean algebras and their representations. In any boolean algebra we can define an infinite union - if it exists it is simply the least upper bound of a set of elements of the algebra. An isomorphism from one boolean algebra to another always sends least upper bounds to least upper bounds, so arbitrary unions are preserved, at least in this sense. The symbol $\bigvee$ is used for the algebraic notion of union.

  When a boolean algebra is represented as *sets* , or in our case when a relation algebra is represented as binary relations, there is a quite separate notion of infinite union. It is the set theoretic union of all the sets. The symbol $\bigcup$ is used for set-theoretic unions. Now in any representation X we have

$$X(\bigvee_\lambda a_\lambda) \supseteq \bigcup_\lambda X(a_\lambda)$$

but the two may not be equal as the set-theoretic union may not correspond to any element of the boolean algebra. In a complete representation we insist that the two are equal[3]. An example of a representation where this fails can be found on page 14.

- Where the context is unambiguous the representation may be referred to simply as X. Instead of saying that a point d belongs to the range of X or $d \in D$ we may just say that the point d is in X.

It might be helpful to take the algebraic properties given on page 10 and to see what they look like in a representation. Observe that for any representation (X, D) of $\mathcal{A}$ the universal relation 1 defines an equivalence relation on the domain D. This follows from the equations: $1 \geq Id$; $1 = 1^{\smile}$ and $1 = 1 * 1$. So D is partitioned into a disjoint union $D_1 \dot{\cup} D_2 \ldots$ and $X(1) = (D_1 \times D_1) \dot{\cup} (D_2 \times D_2) \ldots$.

- For a simple relation algebra each of these sub-domains defines a separate representation of $\mathcal{A}$ and a representation can always be chosen so that $X(1) = D \times D$ for a simple relation algebra, provided it is representable.
- A general relation algebra $\mathcal{A}$ can be decomposed into simple relation algebras and a representation of $\mathcal{A}$ consists of a disjoint union of representations of the simple subalgebras.
- Let (X, D) be a representation of $\mathcal{A}$. A representation of a general relation algebra partitions the domain into sets S_i $(i \in I)$ and each atom of the algebra has domain S_j and range S_k (some j, k). A representable relation algebra $\mathcal{A}$ is integral if and only if it has a representation (X, D) such that for every element $r \in \mathcal{A}$ the domain and the range of r is the whole of D.
- If $\mathcal{A}$ forms a complex group and (X, D) is a representation then each atom $a \in \mathcal{A}$ defines a permutation on D.

Let $\mathcal{A}$ be simple and let (X, D) be a representation with a single equivalence class, so $X(1) = D \times D$. (X, D) can be viewed as a complete directed graph with each edge $(d, e) \in D \times D$ labelled by the set, $X^{-1}(d, e)$, of all relations that hold on that edge[4]. So $a \in X^{-1}(d, e)$ iff $(d, e) \in X(a)$. Now, $X^{-1}(d, e)$ forms a filter[5] on $\mathcal{A}$ since

$$1 \quad \in \quad X^{-1}(d, e)$$

---

[3] Compare this with languages with equality where axioms are given for the binary relation symbol =, but in the end you have to insist that = is interpreted as true equality in a model. Notice that if a theory has a model then it has a model where equality is interpreted correctly. You might wonder whether a similar result can be obtained for complete representations - if a relation algebra has a representation must it have a complete representation? See problem 1

[4] There is an abuse of notation here though if instead of thinking of X as a *function* from $\mathcal{A}$ to $\mathcal{P}(D \times D)$ it is considered as a many : many *relation* between $\mathcal{A}$ and $D \times D$, then X^{-1} coincides with the inverse relation.

[5] Filters and ultrafilters are formally defined on page 32.

$$\begin{aligned} a \in X^{-1}(d,e) \;\&\; a \leq b &\Rightarrow b \in X^{-1}(d,e) \\ a,b \in X^{-1}(d,e) &\Rightarrow a \wedge b \in X^{-1}(d,e) \end{aligned}$$

$X^{-1}(d,e)$ is an ultrafilter since

$$a \in X^{-1}(d,e) \Leftrightarrow -a \notin X^{-1}(d,e)$$

Under favourable circumstances $X^{-1}(d,e)$ will form a principal ultrafilter and we will be able to define it by a single atomic relation.

Definition An *atomic representation* (X,D) of $\mathcal{A}$ is a representation such that for any d and e the ultrafilter $X^{-1}(d,e)$ is principal or, equivalently, there is a unique atom which holds on (d,e).

THEOREM 1 *Let $\mathcal{A}$ be a simple, atomic relation algebra. A representation X of $\mathcal{A}$ is an atomic representation if and only if it is a complete representation.*

PROOF:

First assume that X is a complete representation, i.e. arbitrary unions are preserved. Since $\mathcal{A}$ is atomic we have $1_{\mathcal{A}} = \bigvee atoms_{\mathcal{A}}$. So

$$X(1) = X(\bigvee atoms_{\mathcal{A}}) = \bigcup_{\text{atoms } \alpha} X(\alpha)$$

Therefore for any two points x and y in the representation, as we are assuming that representations consist of single components, the two points must be related by $X(1)$ and hence by some atom $X(\alpha)$. Thus X is an atomic representation.

Conversely, let X be any atomic representation. For any $a \in \mathcal{A}$ we have

$$X(a) = \bigcup \{X(\alpha) : \text{atoms } \alpha \leq a\}.$$

So, for any set of elements $a_\lambda \in \mathcal{A}$

$$\begin{aligned} & X(\bigvee_\lambda a_\lambda) \\ = \; & \bigcup \{X(\alpha) : \text{atoms } \alpha : \alpha \leq \bigvee_\lambda a_\lambda\} \\ = \; & \bigcup_\lambda \bigcup \{X(\alpha) : \text{atoms } \alpha \leq a_\lambda\} \\ = \; & \bigcup_\lambda X(a_\lambda) \end{aligned}$$

and so X is a complete representation.

□

In view of this theorem, herearter we use the term "complete representation".

Note If a relation algebra $\mathcal{A}$ has a complete representation X it follows that $\mathcal{A}$ is an atomic relation algebra. The converse doesn't follow at all: an atomic (even finite) relation algebra might not have any representations and even if it is representable it is not clear that it should have a complete representation.

Example A proper, atomic relation algebra need not be a complete representation of itself. Consider the proper relation algebra with domain $\mathbb{R}$. For each interval I with rational endpoints (closed, open or semi-open) define a basic relation $r(I)$ as $\{(x, y) \in \mathbb{R} \times \mathbb{R} : y - x \in I\}$. The identity relation is $r[0, 0]$; the converse of $r(p, q)$ is $r(-q, -p)$; $r(p_1, q_1) * r(p_2, q_2) = r(p_1 + p_2, q_1 + q_2)$. Let the relation algebra consist of all finite unions of these basic relations. The atomic relations are of the form $r[q, q]$ for some rational number q and because the rationals are dense in the reals the relation algebra is atomic.

But it is not a complete representation since the pair $(0, \pi)$ are not related by any atomic relation.

This relation algebra has another representation, though. Consider instead the proper relation algebra with domain $\mathbb{Q}$ and all the other definitions being as before. This relation algebra is isomorphic to the first, hence a representation. Furthermore, for every pair of rationals (p, q) it *is* the case that there is an atom $(r[q - p, q - p])$ which relates (p, q) so the representation is complete.

PROBLEM 1 *For atomic, representable $\mathcal{A}$ with only countably many atoms, must $\mathcal{A}$ have a complete representation?*[6]

Notation

- Let (X, D) be a complete representation of $\mathcal{A}$ and let $d, e \in D$. We write $\overset{\leftarrow}{X}(d, e)$ for the unique atomic relation which holds between d and e. So $\overset{\leftarrow}{X}(d, e) = \bigwedge X^{-1}(d, e)$.
- Let (X, D) be a complete representation and $\Delta \subseteq D$. $\overset{\leftarrow}{X}(\Delta)$ is defined to be the complete graph (or atomic network) with Δ as its set of nodes and for all $\delta_1, \delta_2 \in \Delta$, the edge from δ_1 to δ_2 is labelled by the atomic relation which holds on (δ_1, δ_2) i.e. $\overset{\leftarrow}{X}(\Delta)(\delta_1, \delta_2) = \overset{\leftarrow}{X}(\delta_1, \delta_2)$.
- Where the context is clear we may write simply X instead of (X, D).

1.3.2 First-order theory

We obtain some results about the size of a representation by converting to a first-order theory. Let $\mathcal{A}$ be any simple relation algebra, and let $L = L(\mathcal{A})$ be the first-order language

---

[6] In a private communication, Roger Maddux suggests that the answer is no. He refers to an error in [Lyndon, 1950] where a relation algebra M' is said to be unrepresentable. Maddux argues that instead M' has a representation but in every representation there are points d, e such that the ultrafilter $X^{-1}(d, e)$ is non-principal.

with one binary predicate symbol for each element of $\mathcal{A}$. Define an L-theory, $Th(\mathcal{A})$ to consist of all of the following:

$$\begin{array}{lcll}
\sigma_{Id} & = & \forall x,y[Id(x,y) \leftrightarrow (x=y)] & \\
\sigma_{\vee}(R,S,T) & = & \forall x,y[R(x,y) \leftrightarrow S(x,y) \vee T(x,y)] & : \text{ for each } R,S,T \text{ with} \\
 & & & \qquad R = S \vee T \\
\sigma_{\neg}(R) & = & \forall x,y[1(x,y) \rightarrow (R(x,y) \leftrightarrow \neg(-R)(x,y))] & \\
\sigma_{\mathrm{conv}}(R,S) & = & \forall x,y[R(x,y) \leftrightarrow S(y,x)] & : \qquad R = S^{\smile} \\
\sigma_{*}(R,S,T) & = & \forall x,y[R(x,y) \leftrightarrow \exists z(S(x,z) \wedge T(z,y))] & : \qquad R = S * T
\end{array}$$

It is clear that an L-structure is a representation of $\mathcal{A}$ if and only if it is a model of $Th(\mathcal{A})$. The following theorem follows immediately from the Löwenheim - Skolem theorem.

THEOREM 2 *If the cardinality of $\mathcal{A}$ is λ and if $\mathcal{A}$ is representable then it has representations of size μ for each cardinal $\mu \geq \lambda + \omega$.*

$\mathcal{A}$ may have smaller representations than $\lambda + \omega$ though. A lower bound on the size of a representation is the number of atoms - fix one point x_0 in the representation, for each atom a_i there must be a point x_i in the representation such that (x_0, x_i) are related by a_i. Clearly the x_i are distinct.

PROBLEM 2 *Let $\mathcal{A}$ be a representable, atomic relation algebra and suppose $|Atoms_{\mathcal{A}}| = \lambda \geq \omega$. Let L_{ats} be the first-order language with binary predicate symbols for the atoms of $\mathcal{A}$ only, so $|L_{ats}| = \lambda$. Can the representations of $\mathcal{A}$ be axiomatised in the language L_{ats} and must $\mathcal{A}$ have a representation of size λ? Theorem 10, later, will tell us that if $\mathcal{A}$ has a complete representation then is has a complete representation of size λ^+, but a similar result for arbitrary representations is not clear.*

1.3.3 Networks

Definitions

1. Let $\mathcal{A}$ be a relation algebra. A *network* N over $\mathcal{A}$ is a square anti-symmetric matrix[7] whose entries are elements of $\mathcal{A}$. N is usually represented as a complete, directed graph with each arc labelled by an element of $\mathcal{A}$. We write $n \in N$ if n is a node in the graph and $N(n,m)$ for the label on the arc from n to m. In this thesis we always adopt the convention that a network is transitively closed (or locally consistent)[8] i.e. for any $l, n, m \in N$
$$N(n,m) * N(m,l) \geq N(n,l).$$
2. A *network isomorphism* $f : N \rightarrow M$ is a bijection from the nodes of N to the nodes of M such that
$$N(m,n) = M(f(m), f(n)).$$

---

[7] Anti-symmetric means that $N(i,j) = N(j,i)^{\smile}$ for each i and j.

[8] In some of the literature this property is referred to as 'path consistency'.

3. An $\mathcal{A}$-network N is *consistent* (or satisfiable) if $\mathcal{A}$ has a representation (X, D) and there is an isomorphism h from N into D i.e. for all $n, m \in N$
$$(h(n), h(m)) \in X(N(n, m)).$$
4. A network is *atomic* if each arc is labelled by an atom.
5. N is a *subnetwork* of M if N is the restriction of M to a subset of its nodes.

Notation Let $Net(\mathcal{A})$ be the class of all finite, atomic, transitively closed $\mathcal{A}$-networks.

Comments

1. In the temporal reasoning literature [Allen, 1983, Villain *et al.*, 1989, Dechter *et al.*, 1991] a network is considered as an example of a constraint satisfaction problem - given a (transitively closed) network, usually with non-atomic relations on the arcs, the problem is to discover whether the network is globally consistent. Global consistency has been taken to mean that the network can be tightened[9] to an atomic network which is still transitively closed. The problem is to choose one atom from the relation on each arc in such a way that the resulting atomic network is transitively closed. We shall see that an important assumption has been made, namely that a transitively closed, atomic network is globally consistent, that is, it has a model. This assumption turns out to be valid only on a certain subclass of the class of all relation algebras - those where $Net(\mathcal{A})$ forms a TAP-class (defined later).
2. Another interpretation of networks will be developed in the following section. A network will be used as a kind of bridge between the purely algebraic properties of the relation algebra and the properties of a representation. We have seen that any finite substructure of a complete representation defines a network. We shall consider the converse - when does a given network embed in some representation of the relation algebra?

1.3.4 Games

An elegant method of determining whether a relation algebra has a representation and, if so, what properties its representations might have is to try to construct a representation using a game (see [Hodges, 1985]). The crucial *conjunction lemma* (below) will allow us to avoid lengthy induction arguments in the proofs.

---

[9] Network N is a *tightening* of M if N and M have exactly the same nodes and for each i and j, $N(i, j) \leq M(i, j)$.

Let $\mathcal{K}$ be a class of finite atomic $\mathcal{A}$-networks downward closed and closed under isomorphism. Our games will be played by two players, $\forall$ and $\exists$ who alternately pick some network from $\mathcal{K}$ extending the network of the previous player. $\forall$ plays first and the game is of countable or finite length. So a *play* of the game $G(\mathcal{K})$ might be

$$N_0 \subseteq N_1 \subseteq \ldots$$

where the even turns are chosen by $\forall$ and the odd ones by $\exists$. The class of all possible plays of the game is partitioned into two: wins for $\forall$ and wins for $\exists$ - there are no draws. The wins for $\exists$ are determined by some property P of the *limit* of the play $\bigcup_{i<\omega} N_i$. $\exists$ wins this play of the game G_P if and only if the limit satisfies the property P.

Definition Let P be a property, consider the game G_P. Let

$$\sigma : \mathcal{K}^{<\omega} \to \mathcal{K}$$

be such that $\sigma(N_0, N_1, \ldots, N_n) \supseteq N_n$. $\exists$ can use σ to play the game by responding to the initial play $N_0 \subseteq N_1 \subseteq \ldots N_{2k}$ with $N_{2k+1} = \sigma(N_0, N_1, \ldots, N_{2k})$. σ is a *winning strategy* for $\exists$ if in any play of G_P in which $\exists$ always uses σ, $\exists$ is guaranteed to win the game.

Example If every $N \in \mathcal{K}$ has a proper extension in $\mathcal{K}$ then $\exists$ has a winning strategy to ensure that the limit is infinite: for each turn she must choose a *proper* extension of $\forall$'s previous move.

LEMMA 3 (Conjunction Lemma) *Let $\{P_i : i < \omega\}$ be a countable set of properties and suppose σ_i is a winning strategy for G_{P_i}. Then $\exists$ has a winning strategy in the game G_P where $P = \bigwedge_{i<\omega} P_i$, that is, she can achieve all the properties at once.*

PROOF:

See [Hodges, 1985] page 30 or the proof of theorem 9 herein. $\square$

1.3.5 Using Games to Build Representations

The finite relation algebras possessing representations can be characterised by an infinite axiom schema [Lyndon, 1950] but by no finite set of axioms. In this section a representation of a relation algebra is further classified as *universal* or *homogeneous.* Relation algebras with such representations are characterised by certain network properties. These properties are equivalent to infinite axiom schemas but they are conceptually clearer.

Definitions Let (X, D) be a complete representation of $\mathcal{A}$ and let $\mathcal{K}$ be any class of finite atomic networks over $\mathcal{A}$, closed under isomorphism and downward closed (i.e. if $N \in \mathcal{K}$ and M is a subnetwork of N then $M \in \mathcal{K}$). So $\mathcal{K}$ could be $Net(\mathcal{A})$ for example.

1. A *local isomorphism* θ of (X, D) is a finite partial map from X to itself say $\theta : \overline{d} \to \overline{e}$ ($\overline{d}, \overline{e}$ k-tuples from D) such that $\overset{\leftarrow}{X}(d_i, d_j) = \overset{\leftarrow}{X}(e_i, e_j)$. So θ preserves each edge relation in $\overline{d}$.

2. X is *homogeneous* if every local isomorphism of X extends to a full automorphism of X.

3. The *age* of a representation ($age(X)$) is defined to be the class of all finite atomic networks that embed into X (The range of X is D. Instead of saying a network N embeds into the range of X we will simply say N embeds in X).
   X is *universal* over $\mathcal{K}$ if every member of $\mathcal{K}$ embeds into X
   X is a *universal representation* of $\mathcal{A}$ if X is universal over $age(Y)$ for every representation Y of $\mathcal{A}$
   X is a *fully universal representation* of $\mathcal{A}$ if X is universal over $Net(\mathcal{A})$ the class of *all* finite atomic $\mathcal{A}$-networks.

4. $\mathcal{K}$ has the *joint embedding property* (JEP) if any two networks N and M in $\mathcal{K}$ jointly embed into a network L of $\mathcal{K}$.

5. $\mathcal{K}$ has the *Amalgamation Property* (AP) if given any networks L, M_1 and M_2 from $\mathcal{K}$ and embeddings $f_i : L \to M_i$ $(i = 1, 2)$ there is a network $N \in \mathcal{K}$ and embeddings $g_i : M_i \to N$ such that $f_1 g_1 = f_2 g_2$. Less formally, given M_1 and M_2 containing common isomorphic copies of a network L, there is a way of gluing M_1 and M_2 together along L and forming a larger network N, still in $\mathcal{K}$. If we allow L to be empty then the amalgamation property implies the joint embedding property.

6. An *amalgamation class* is a class of finite atomic networks with JEP and AP which is closed under isomorphism and downward closed.

7. $\mathcal{K}$ has the *triangle addition property* (TAP) if given any atomic network $N \in \mathcal{K}$ and any atomic network of size three T (a triangle) such that N and T have one edge each labelled by the same relation then it is possible to embed N and T into an atomic network $L \in \mathcal{K}$ such that N and T share that same edge in L. If all the triangles belong to $\mathcal{K}$ then TAP becomes a special case of AP.

Notation Write $\tau(\mathcal{A})$ for the union of all the TAP-classes of $\mathcal{A}$, clearly this is still a TAP-class.

Note As we are not assuming that the relation algebra is integral it may be that the identity Id is not an atom. So we should worry, a little, about which unit relation holds on (n, n) for each node n in a network. However, this is not a problem. Given any atom r there are unique units $e_{\text{dom}}, e_{\text{range}}$ such that

$$\begin{aligned} e_{\text{dom}} * r &= r \\ r * e_{\text{range}} &= r \end{aligned}$$

Therefore, if we have a consistent atomic $\mathcal{A}$-network where only the edges between distinct nodes are labelled, the unit relation that holds between a node and itself is implicitly definable from the other relations. Effectively, this means that we do not need to worry about the unit relations.

Notation Let N be an $\mathcal{A}$-network. Define $N^{\uparrow} : \mathcal{A} \to \mathcal{P}(N \times N)$ by

$$N^{\uparrow}(a) = \{(m, n) : N(m, n) \leq a\}.$$

Effectively, $N^{\uparrow}$ is formed from N by 'closing upwards' on each edge.

LEMMA 4 *Let $\mathcal{A}$ be a simple, atomic relation algebra and let N be a transitively closed atomic network. $N^{\uparrow}$ is a complete representation of $\mathcal{A}$ iff for every edge (n, m) labelled by a and every pair of atoms b, c with $b * c \geq a$, there is a node $l \in N$ such that (n, l) is labelled b and (l, m) is labelled c.*

PROOF:

($\Rightarrow$) Suppose $N^{\uparrow}$ is a complete representation. Now suppose $N(n, m) = a$ and $b * c \geq a$ (atoms a, b, c). So $(n, m) \in N^{\uparrow}(b * c)$ and from the definition of $*$ there must be a point l with $(n, l) \in N^{\uparrow}(b)$ and $(l, m) \in N^{\uparrow}(c)$, or $N(n, l) = b, N(l, m) = c$.

($\Leftarrow$) Assume that the edge property holds. We must show that $N^{\uparrow}$ preserves all the operators. It is easy to verify that the Boolean operations[10] and converse are preserved. The edge property guarantees that composition is preserved as well. $\square$

THEOREM 5 *Let $\mathcal{A}$ be an atomic relation algebra with at most countably many atoms and let $\mathcal{K}$ be some downward closed ($N \in \mathcal{K}, M \subseteq N \to M \in \mathcal{K}$) isomorphism class of finite, atomic, transitively closed $\mathcal{A}$-networks which contains every $\mathcal{A}$-triangle.*

1. *$\mathcal{A}$ is a completely representable relation algebra if and only if there is some $\mathcal{K}$ with the triangle addition property (if and only if $\tau(\mathcal{A})$ is non-empty).*
2. *$\mathcal{A}$ has a complete representation X with age(X) $= \mathcal{K}$ iff $\mathcal{K}$ has TAP and JEP. $\mathcal{A}$ has a complete, universal representation iff $\tau(\mathcal{A})$ is non-empty and has JEP.*

---

[10] You need $\mathcal{A}$ to be simple in order that $I(-a) = -I(a)$.

3. *$\mathcal{A}$ has a homogeneous, complete representation X, with age$(X) = \mathcal{K}$ iff $\mathcal{K}$ has JEP and AP. For the countable case, such a representation will be unique up to isomorphism.*

4. *$\mathcal{A}$ has a complete, countable, homogeneous, universal representation iff $\tau(\mathcal{A})$ is non-empty and has JEP and AP (i.e. $\tau(\mathcal{A})$ forms a non-empty amalgamation class). Such a representation is unique up to isomorphism.*

PROOF:

1. $\Rightarrow$ If $\mathcal{A}$ has a complete representation X, then let $\mathcal{K}$ be the class of all finite atomic networks which embed in the range of X (the finite networks that embed in X, for short). Clearly $\mathcal{K}$ is downward closed. Also, let N be any atomic network and let Ξ embed N into X. If $N(n, m) = a$ and $a \leq b * c$ then it is necessary to show that the triangle T

can be joined onto the edge (n, m) of N and form an atomic network that still belongs to $\mathcal{K}$. But since X is a representation there must be some point l say with $\overleftarrow{X}(\Xi(n), \Xi(l)) = b$ and $\overleftarrow{X}(\Xi(l), \Xi(m)) = c$ (use lemma 4). So the complete atomic network on $N \cup \{l\}$ is the required extension. Therefore $\mathcal{K}$ has TAP.

$\Leftarrow$

First suppose that $\mathcal{A}$ is simple.
Assume that some TAP-class $\mathcal{K}$ exists. Let $\exists$ win the game $G_R(\mathcal{K})$ if and only if the limit is a complete representation of $\mathcal{A}$.
Claim: $\exists$ has a winning strategy for $G_R(\mathcal{K})$ and therefore $\mathcal{A}$ has an complete representation, moreover each network $N \in \mathcal{K}$ embeds into some complete representation[11] (if $\forall$'s first move is N then the limit is a complete representation embedding N).
Proof of claim: $\exists$ needs to ensure that the conditions of lemma 4 are met viz. for each edge (n, m) which appears in the play labelled by the atom a say, and each pair of atoms b, c such that $b * c \geq a$ there is a node l such that (n, l) is labelled by b and (l, m) is labelled by c. Now since $\mathcal{K}$ has the triangle addition property, for any *one* edge (n, m) and any

---

[11] But note that the networks in $\mathcal{K}$ do not necessarily all embed into the same representation.

single pair of atoms b, c, $\exists$ can ensure that the required node is added. After the edge has been included in the play she simply chooses from $\mathcal{K}$ some network with the relevant triangle added on. Since there are only countably many edges which appear and at most countably many pairs of suitable atoms, we can use the conjunction lemma to conclude that for *every* edge and *every* suitable pair of atoms the extra node exists in the limit. This proves the claim.

Now if $\mathcal{A}$ is not simple it can be shown that the TAP-class $\mathcal{K}$ decomposes into disjoint classes, with each component being a TAP-class over a simple subalgebra of $\mathcal{A}$. Therefore each simple component of $\mathcal{A}$ has a complete representation, so the disjoint union of complete representations of each of the simple subalgebras will form a complete representation of $\mathcal{A}$. Henceforth we assume $\mathcal{A}$ to be simple as it is always possible to perform this type of decomposition.

2. (a) $\Rightarrow$

    Suppose $age(X) = \mathcal{K}$. As earlier, it follows that $\mathcal{K}$ has TAP. Also $\mathcal{K}$ has JEP since if $N, M \in \mathcal{K}$ they must embed in X, say as N_x and M_x. Let L be the atomic network formed by the nodes of $N_x \cup M_x$ with arcs labelled by the atomic relation that holds in X. $L \in \mathcal{K}$.

    $\Leftarrow$

    Let $\mathcal{K}$ have TAP and JEP. Define a win for $\exists$ in the game $G_U(\mathcal{K})$ by requiring that the limit is a complete representation with age $\mathcal{K}$. It is required to prove that $\exists$ has a winning strategy for this game. Since $\mathcal{K}$ has TAP, $\exists$ can certainly ensure that the limit is a complete representation. Also, given any one network $N \in \mathcal{K}$ she can ensure that N embeds in the limit and she can do this in one move: if $\forall$'s last turn was M then JEP means that she can choose some network $L \in \mathcal{K}$ extending M and embedding N. By the conjunction lemma she can enforce that the limit is a complete representation and *each* network $N \in \mathcal{K}$ embeds in the limit.

    (b) Let $\mathcal{K} = \tau(\mathcal{A})$.

3. $\Rightarrow$

    Suppose $\mathcal{A}$ has a homogeneous complete representation X with $age(X) = \mathcal{K}$. Let $L, M_1, M_2 \in \mathcal{K}$ with embeddings $f_i : L \to M_i$. Let $M_1\theta_1$ and $M_2\theta_2$ be respective appearances of M_1 and M_2 in X (θ_i are embeddings from M_i into X). There is a local isomorphism from $Lf_1\theta_1$ to $Lf_2\theta_2$ (namely ${\theta_1}^{-1} * {f_1}^{-1} * f_2 * \theta_2$). By homogeneity, this extends to a full automorphism α of X. To construct N embedding M_1 and M_2 consider the points $M_1\theta_1\alpha \cup M_2\theta_2$ in X. For each pair of nodes n, m from this set label the arc (n, m) by the atomic relation that actually holds on the two points in X, namely $\overset{\leftarrow}{X}(n, m)$. This labelling agrees with the arcs in M_1 and M_2 and makes them intersect correctly on L. This defines N which embeds in X and therefore belongs to $\mathcal{K}$. The case where M_1 and M_2 do not intersect covers the JEP and the case where they do covers AP.

$\Leftarrow$

Suppose $\mathcal{K}$ has JEP and AP. Let $\exists$ win $G_A(\mathcal{K})$ if the limit has age $\mathcal{K}$ and is homogeneous. Since $\mathcal{K}$ contains every triangle of $\mathcal{A}$ the triangle addition property follows from AP. Thus $\exists$ can certainly ensure that the limit is an complete representation with age $\mathcal{K}$. She must also make sure that any local isomorphism extends to a full automorphism i.e. that the limit is homogenous. Suppose a local isomorphism $\theta : \bar{n} \rightarrow \bar{m}$ eventually appears in the play. She must extend θ to an automorphism of the limit. So for any n which appears in the play she must ensure that θ can be extended to a bigger local isomorphism in such a way that n is in the domain and range of θ. She can achieve this in one move. Let $\forall$'s last move be N.

To define $n\theta$, take a copy N' of N and identify the tuple $\bar{n'}$ of N' with $\bar{m}$ of N (this can be done since θ preserves all edge relations from $\bar{n}$ to $\bar{m}$). The amalgamation property allows us to extend the result to a structure in $\mathcal{K}$ containing N and N'. Let n' be the element of N' that corresponds to n and set $n\theta = n'$.

Similarly we can extend θ and N so that $n\theta^{-1}$ is defined. Since there are only countably many local isomorphisms that appear in the play and countably many nodes n to extend the local isomorphism to, the conjunction lemma allows us to conclude that $\exists$ can extend every local isomorphism to a full automorphism. Thus the limit is a complete, homogeneous representation of $\mathcal{A}$.

Let X and Y be any two countable, homogeneous, complete representations of $\mathcal{A}$ with age $\mathcal{K}$. We get an isomorphism by a back and forth construction. Start with an arbitrary singleton network in X and any in Y labelled by the same unit. These two networks are isomorphic. At each stage we have two locally isomorphic finite atomic networks M and N in X and Y respectively and we extend one of them (say M in X) by choosing the next point, x in X. Let M^* be the atomic network formed by the points $M \cup \{x\}$. To find the corresponding point in Y use the fact that $age(X) = age(Y)$ to show that an atomic network isomorphic to M^* exists in Y and then homogeneity to show that such a network extends N to $N \cup \{y\}$ for some y. Provided we ensure that each point of X and Y gets chosen eventually, the limit of this process will be an isomorphism between X and Y.

4. Set $\mathcal{K} = Net(\mathcal{A})$ in the previous part to show that $\mathcal{A}$ has a complete, homogeneous, universal representation if and only if $Net(\mathcal{A})$ has JEP and AP. That $\mathcal{A}$ has a countable representation of this kind follows from the way the representation is constructed.

□

Definition The countable, complete, fully universal, homogeneous representation constructed in the last part of the preceding theorem is called the *Fraïssé* limit of the amalgamation class.

Definition A *normal representation* is a complete, fully universal, homogeneous representation. So the theorem tells us that a representation has a normal representation if and only if its networks form an amalgamation class.

PROBLEM 3 *Apply this method to the analysis of representations of cylindric algebras*[12].

COROLLARY 6 *Let $\mathcal{A}$ be an atomic relation algebra with countably many atoms. $\mathcal{A}$ satisfies the Lyndon conditions [Lyndon, 1950] if and only if $\mathcal{A}$ has a complete representation.*

PROOF:

[Sketch] We show here that a relation algebra has a TAP-class if and only if it satisfies the Lyndon conditions[13]. The previous theorem gives the equivalence with completely representable relation algebras.

Notation Let τ and κ be any classes of atomic $\mathcal{A}$-networks, both closed under isomorphism and downward closed. We say that κ is an *extension* of τ and write $Ext(\tau,\kappa)$ if $\forall N \in \tau\ \forall m, n \in N\ \forall$ atoms b, c with $b{*}c \geq N(m,n)\ \exists N^+ \in \kappa$ such that

$$N^+ = N \cup \{l\} \ \wedge\ N^+(m,l) = b \ \wedge\ N^+(l,n) = c$$

So, for example, τ is a TAP-class if and only if $Ext(\tau,\tau)$ holds. Now for any non-empty TAP-class it is easy to see that every triangle

---

[12] [Andréka *et al.*, 1991] includes some important work on the inter-relations between cylindric algebra and relation algebra.

[13] These conditions may be found in full in [Lyndon, 1950], though the less studious reader will find a very rough description of them in the current proof.

must belong to the class $(y * z \geq x)$. Let τ_0 be the class of all such triangles. If $\mathcal{A}$ has a TAP-class then the following chain of classes must exist

$$\tau_0 \subseteq \tau_1 \subseteq \ldots$$

where for each $k < \omega$ we have $Ext(\tau_k, \tau_{k+1})$. Conversely if such a chain exists then the union will form a TAP-class. Thus $\mathcal{A}$ has a TAP-class if and only if

$$\phi_k =_{\text{def}} \exists\tau_0 \ldots \exists\tau_{k+1}[Ext(\tau_0, \tau_1) \wedge [Ext(\tau_1, \tau_2) \wedge \ldots [Ext(\tau_k, \tau_{k+1})] \ldots]]$$

holds on $\mathcal{A}$ for each $k < \omega$. ϕ_0 is equivalent to saying that for any triangle occuring in a network in τ_0 and any triangle sharing a common edge

it is possible to glue them together and form a new network in τ_1. Thus we derive the first Lyndon condition

$$\forall x, y, z, u, v\ [x \leq ((y * z) \wedge (u * v)) \rightarrow \exists w \leq ((y^{\smile} * u) \wedge (z^{\smile} * v))]$$

There are two other axioms which correspond to gluing a triangle onto the edges labelled y, z respectively. The conjunction of these three axioms is ϕ_0.

To obtain ϕ_1 we require that if $s * t \geq w$ then it is possible to glue the triangle onto the quadrilateral above to get

in τ_2. Thus the axiom given above extends to

$$\begin{gathered}\forall x, y, z, u, v\, [x \leq ((y * z) \wedge (u * v)) \rightarrow \\ \exists w [w \leq ((y^{\smile} * u) \wedge (z^{\smile} * v)) \wedge \forall s, t (s * t \geq w \rightarrow \\ \exists p, q ((p \leq s^{\smile} * z \wedge t * v) \wedge (q \leq t * u^{\smile} \wedge s^{\smile} * y^{\smile} \wedge p * x^{\smile})))]]\end{gathered}$$

ϕ_1 consists of this plus five other conditions - one for each of the other edges of the quadrilateral - plus another six for each of the other two conjuncts in ϕ_0. The validity of ϕ_1 guarantees that for every edge on every quadrilateral there is a pentagon formed by gluing the quadrilateral to a triangle.

Continuing this way, we can derive all the Lyndon conditions which jointly ensure the existence of a TAP-class.

□

PROBLEM 4 *The formula ϕ_k in the corollary uses about $k + 2$ variables and is of the form $\forall_k$ (alternating universal and existential quantifier blocks to a depth of k). Show that the completely representable relation algebras cannot be axiomatised using a bounded number of variables and that they cannot be axiomatised by $\forall_k$ (for any fixed k)*[14].

PROBLEM 5 *The homogeneous representations appear to be the* natural *representations and they must offer some advantage for computation, in that the* context *of a network*

---

[14] In [Lyndon, 1956] it is shown that this class cannot be axiomatised using universal sentences and [Andréka, 1989] has shown further that the representable relation algebras cannot be characterised by any set of axioms using a fixed number of variables. See also [Maddux, 1989] for a strengthening of the basic result.

embedding in some representation makes no difference. All of the representable relation algebras encountered up to here possess homogeneous representations. Is this always the case? Equivalently, is there a representable relation algebra with no amalgamation class of atomic networks? (Solution on page 30).

Note We will show later that the atomic networks of the Allen interval algebra ($\mathcal{I}$) form an amalgamation class and the last part of the preceding theorem tells us that all the countable, universal, homogeneous representations are isomorphic. It is tempting to draw the conclusion that $\mathcal{A}$ is ω-categorical (that all its countable representations are isomorphic (a theorem first proved for $\mathcal{I}$ in [Ladkin and Maddux, 1987a]) if its atomic networks form an amalgamation class. However, this conclusion would be far too hasty[15] and to illustrate this we define a relation algebra whose atomic networks form an amalgamation class and which therefore has only one countable, universal, homogeneous representation (up to isomorphism) but which has non-isomorphic representations which are either not universal or inhomogeneous (or neither).

Let $\mathcal{B}$ have atomic elements $\{Id, e, d\}$ where Id is the identity, $e^{\smile} = e$ and $d^{\smile} = d$ and the composition is defined by

$$\begin{aligned} e * e &= Id \vee e \\ e * d = d * e &= d \\ d * d &= 1 \end{aligned}$$

Put simply, the relation $Id \vee e$ is an equivalence relation which partitions the domain of a representation into at least three clusters. The countable, universal, homogeneous representation consists of a countable collection of countable 'clusters'. Within one cluster two distinct elements are related by e. Two elements from different clusters are related by d.

However, the representation defined similarly but with always three elements in a cluster is homogeneous but not universal. A representation with clusters of size n for all $n > 2$ is universal but not homogeneous. If there are infinitely many clusters of size 7 and 11, but no other clusters, then the representation is neither universal nor homogeneous. Thus $\mathcal{B}$ has many non-isomorphic, countable representations.

1.3.6 Examples of relation algebras

- The point algebra $\mathcal{P}$ and the interval algebra $\mathcal{I}$. $Net(\mathcal{P})$ and $Net(\mathcal{I})$ both form amalgamation classes. In fact the universal, homogeneous representation constructed by the previous theorem (the Fraïssé limit) is in each case the only countable representation of each algebra (proved on page 41).

---

[15] In the section on intervals we will prove a generalisation of Ladkin and Maddux's theorem.

- The relation algebra $\mathcal{B}$ (above). $Net(\mathcal{B})$ forms an amalgamation class so $\mathcal{B}$ has a unique countable, homogeneous, universal representation. However $\mathcal{B}$ has uncountably many other non-isomorphic countable representations.

- Let $\mathcal{C}$ have atoms Id and $\#$ (both self converse) and composition defined by $\# * \# = Id$. $\mathcal{C}$ has only one representation, which has exactly two elements, and looks like this

- Let $\mathcal{D}$ have atoms Id and $\#$ (both self converse) and composition defined by $\# * \# = Id \vee \#$. $Net(\mathcal{D})$ forms an amalgamation class and the Fraïssé limit consists of a countable set of points and each pair of distinct points is related by $\#$. However, for each $n \geq 3$, it also has finite representations consisting of a set of size n with each pair of distinct points related by $\#$. Still, $\mathcal{D}$ has only one countable representation (up to isomorphism) and thus is weakly ω-categorical (see definition of ω-categorical on page 36).

- **Lyndon Algebras** A Lyndon Algebra [Lyndon, 1961] is a relation algebra $\mathcal{A}$ satisfying

  $$(\forall r, s \in \mathcal{A}) Id \leq r * s \Leftrightarrow r \wedge s \neq 0$$

  Lyndon algebras are all integral and every element is self-converse. These algebras are very useful as counter-examples to a number of conjectures, as we will see later. Here we find a representable Lyndon algebra with no representation universal over $Net(\mathcal{A})$. It has five atoms: Id, a_1, a_2, a_3, a_4 and composition defined by $a_i * a_i = Id \vee a_i$ $(i = 1, 2, 3, 4)$ and $a_i * a_j = \bigvee_{k \neq i,j} a_k$ $(i \neq j, k \leq 4)$, e.g. $a_1 * a_2 = a_3 \vee a_4$.

  A representation of $\mathcal{A}$ can (in fact *must*) be constructed as follows. Take a projective geometry G of order three (so four points on a line). Form an affine geometry by picking one line l (the "line at infinity") and removing it, leaving $A = G \setminus l$. Enumerate[16] the four points of l in some way $l = \{l_1, l_2, l_3, l_4\}$. We use A as the domain for the representation, in fact there are nine points in A. For any pair of distinct points $x, y \in A$ there is a unique line $\overline{xy}$ through them and this line meets l at a unique point. Define $X(a_i)$ to be the set of pairs (x, y) such that $x \neq y$ and $\overline{xy}$ meets l at l_i (for each $i \leq 4$). Of course $X(Id) = \{(x, x) : x \in A\}$. It is not hard to see that this is a representation of $\mathcal{A}$.

  But it is not universal over $Net(\mathcal{A})$ as the following network does not embed in X.

---

[16] A different enumeration may give rise to a non-isomorphic representation.

This is because there are only three distinct points in A on the line through l_1.

It can be shown (see [Lyndon, 1961]) that all representations of $\mathcal{A}$ are affine representations, hence the quadrangle above does not embed in any representation of $\mathcal{A}$.

Notation

- A relation algebra $\mathcal{A}$ such that $Net(\mathcal{A})$ forms a TAP-class with JEP is called *fully universal.*
- A representation of a relation algebra $\mathcal{A}$ which is universal over $Net(\mathcal{A})$ is *fully universal.*

Equivalently, for a fully universal representation every atomic $\mathcal{A}$-network embeds in the representation. A fully universal relation algebra always has a fully universal representation. Since $Net(\mathcal{A})$ is the largest possible TAP-class, a fully universal representation is automatically universal, though the previous example shows that the converse is false.

1.3.7 The Containment Algebra ($\mathcal{C}$)

Ladkin and Maddux [Ladkin and Maddux, 1987a] presented the containment algebra as a subalgebra of the Allen interval algebra. In their version it has five atoms $=, \subset, \supset, Out, Lap$ where the last four are non-atomic elements of the Allen interval algebra:

$$\begin{aligned} \subset &= \text{start} \vee \text{end} \vee \text{during} \\ \supset &= \subset^{\smile} \\ Out &= < \vee > \vee \text{meet} \vee \text{met-by} \\ Lap &= \text{overlaps} \vee \text{overlapped-by} \end{aligned}$$

Ladkin and Maddux construct an atomic network (below) which is transitively closed but is inconsistent *over representations of* $\mathcal{I}$.

To use the terminology of this thesis, no representation of $\mathcal{I}$ can be fully universal over the containment algebra.

The full composition table for the containment algebra, $\mathcal{C}$, is given by

$*$	$\subset$	$\supset$	Out	Lap
$\subset$	$\subset$	1	Out	$\subset, Lap, Out$
$\supset$	$=, Lap, \subset, \supset$	$\supset$	$Lap, \supset, Out$	$\supset, Lap$
Out	$Lap, \subset, Out$	Out	1	$Out, Lap, \subset$
Lap	$\subset, Lap$	$\supset, Lap, Out$	$Out, Lap, \supset$	1

Now, although the network (above) is unsatisfiable in any representation of $\mathcal{I}$, a fully universal representation of $\mathcal{C}$ can be constructed.

The atomless Boolean algebra An obvious representation of the containment algebra ($\mathcal{C}$) might be to take the power set of some set and interpret each relation as a *set relation* e.g. $a \subset b$ means that a is a proper subset of b. The problem with this is that we have the composition $\subset * \subset = \subset$ but the power set includes the singleton set $\{x\}$. If the full power set were a representation then from $\emptyset \subset \{x\}$ we should be able to find some set S with $\emptyset \subset S \subset \{x\}$, but none exists. So, instead, the so-called atomless Boolean algebra is used.

Take any infinite set X and define an equivalence relation $\sim$ on $\mathcal{P}(X)$ by $S \sim T$ if the symmetric difference $S \bigtriangledown T$ is finite. Denote the equivalence class of S by $[S]$ and the set of equivalence classes by $\mathcal{P}(X)/\sim$. The domain for the representation will be $(\mathcal{P}(X)/\sim) \setminus \{[\emptyset], [X]\}$. The atomic relations are interpreted in the natural way so, for example $[S]Out[T]$ holds if and only if there are only finitely many points common to S and T.

It is not hard to check that this representation is fully universal but note that it will be an uncountable representation and it is not homogeneous. To see that it is not homogeneous, partition X into three infinite sets $X = S_1 \dot{\cup} S_2 \dot{\cup} S_3$ and then re-partition it into four - $X = T_1 \dot{\cup} T_2 \dot{\cup} T_3 \dot{\cup} T_4$. The local isomorphism which maps $[S_i] \rightarrow [T_i]$ $(i = 1, 2, 3)$ does not extend to a full isomorphism since T_4 cannot be the image of any element.

It can be shown that the networks of the containment algebra form an amalgamation class and therefore there is a unique, countable, homogeneous, universal representation of $\mathcal{C}$ (details omitted). It is also clear that there are other countable, inhomogeneous representations and so it is not ω-categorical. As to whether there are representations possessing a different age (i.e. whether the only TAP-class is Net($\mathcal{C}$)) is unclear at the time of writing. This is relevant to problem 7 on page 41.

PROBLEM 6 *Theorem 5 exactly characterises the RAs with complete, homogeneous, universal representations. Can it be proved that no finite set of axioms is capable of doing this?*

1.3.8 A Relation Algebra with no Homogeneous Representation

Whilst we are giving examples of relation algebras and their representations we answer problem 5 negatively by presenting in outline a representable relation algebra with no homogeneous representation. This was discovered by Ian Hodkinson, following a paper of Roger Lyndon [Lyndon, 1961].

Lyndon gives a very general definition of a geometry G, which we quote. It consists of a set of *points*, together with certain subsets, called *lines*. The geometry must satisfy three axioms:

- there exists at least one line, and each line contains at least four points
- each pair of distinct points p and q lies on a unique line, $\overline{pq}$
- if p, q and r are distinct points, and a line meets $\overline{pq}$ and $\overline{pr}$ in distinct points, then it meets $\overline{qr}$.

Given a *geometry* G Lyndon defines $A(G)$. The elements of $A(G)$ are all subsets of $G \cup \{Id\}$ where Id is a new element, disjoint from G. Id is the neutral element and composition is defined on the other atoms by

$$p * p = Id \vee p$$

$$p * q = \overline{pq} \setminus \{p, q\} \text{ for } p \neq q$$

where $\overline{pq}$ is the unique line containing p and q. $A(G)$ forms a Lyndon algebra (see page 27) which is a special case of a relation algebra.

If G can be embedded as a hyperplane in a geometry H of one higher dimension then we can define the *affine* space $D = H \setminus G$. An *affine representation* Δ of G is defined on the atoms by letting $\Delta(Id)$ be the identity relation on D and for a point $p \in G$

$$\Delta(p) = \{(x, y) : x \neq y \text{ and } p \in \overline{xy}\}.$$

This extends uniquely to a complete representation of $A(G)$. We now quote a theorem from [Lyndon, 1961] page 24

THEOREM 7 *Each affine representation of $A(G)$ is a representation, and each completely additive [complete - RH] representation of $A(G)$ is equivalent to some affine representation.*

For finite relation algebras the distinction between complete representations and general representations does not exist. So if we can prove the lemma that follows we will have shown that, at least for finite G, the algebra $A(G)$ is representable but has no homogeneous representation.

LEMMA 8 *An affine representation of dimension greater than one and order greater than three is not homogeneous.*

PROOF:

> Let (D, Δ) be an affine representation of $A(G)$ as defined above. First we show that any automorphism α of D that fixes two distinct points d and e must be the identity. Let $f \in D$ be any point and suppose first that f is not on the line $\overline{de}$. Let $\overline{df}$ meet G at p i.e. $(d, f) \in \Delta(p)$. Since α is an automorphism $\overline{(d\alpha)(f\alpha)} = \overline{d(f\alpha)} \in \Delta(p)$. So $f\alpha$ is on the line $\overline{df}$. Similarly it is on the line $\overline{ef}$ and as these two lines meet only once, it follows that f is fixed. The case where f is on the line $\overline{de}$ can be covered by two iterations of this argument: once to fix some other point off the line and a second time to fix f.
>
> Now in an affine space the order of the geometry is the same as the number of points on each line. By assumption the number of points on a line is at least four. Let l be any line and let a_1, a_2, a_3, a_4 be any distinct points on the line. If l meets G at p then each pair $(a_i, a_j) \in \Delta(p)$ $(i \neq j)$. So the mapping
>
> $$(a_1, a_2, a_3, a_4) \to (a_2, a_1, a_3, a_4)$$
>
> preserves the relations between the four points and is therefore a local isomorphism. It cannot extend to a full automorphism since it fixes two points and it is not the identity. Consequently this representation is not homogeneous.
>
> □

So an example of a representable relation algebra with no homogeneous representation can be formed by taking a field of order four and embedding a two-dimensional vector space in a three-dimensional one over that field. These can be turned into a one-dimensional projective space G embedded in a two-dimensional projective space H. Let D be the affine space $H \setminus G$. G has five points so the relation algebra $A(G)$ has six atoms: $Id, g_1, \ldots, g_5$. Id is the neutral element and other compositions are defined by

$$g_i * g_i = Id \vee g_i \text{ (for } i = 1 \ldots 5)$$

$$g_i * g_j = \bigvee_{k \neq i,j} g_k \; (i \neq j)$$

All of its representations must be equivalent to an affine representation into D, though they are not all necessarily isomorphic as permutations of the atoms can induce non-isomorphic representations.

1.3.9 More general representations and transfinite games

The previous section has dealt with complete representations and the corresponding networks have therefore been labelled by atoms. To generalise this work it will be necessary to use ultrafilters instead of atoms as the labels on the edges of the networks. Unfortunately, for infinite $\mathcal{A}$ the number of ultrafilters will be uncountable and so it is necessary to generalise the rules of the *games* and the application of the conjunction lemma[17].

Definitions

- A *regular* cardinal κ has the following property: for any sequence of sets $(S_\lambda)_{\lambda<\mu}$ where the length of the sequence (μ) is less that κ and the size of each S_λ is also strictly less than κ, the size of the union $\bigcup_{\lambda<\mu} S_\lambda$ is strictly less than κ.

  For example ω is the first regular cardinal (a finite union of finite sets is finite) and it is known that regular cardinals are arbitrarily large. Let κ^+ be the least regular cardinal at least as big as κ.

- An $\mathcal{A}$-*filter* (or just filter) is a non-empty set of elements of $\mathcal{A}$, closed upwards and closed under finite intersections. The *principal* filter $F(a) =_{\text{def}} \{b \in \mathcal{A} : a \leq b\}$.

- A filter F holds on pair of points (x, y) (in a representation X) and we write $F(x, y)$ if $\forall f \in F$ we have $f(x, y)$ i.e. all relations f hold on the pair.

- The *product* and *converse* of filters F and G are defined by

  $$\begin{aligned} F * G &= \{a \in \mathcal{A} : \exists f \in F, g \in G, f * g \leq a\} \\ F^\smile &= \{f^\smile : f \in F\} \end{aligned}$$

  $F^\smile, F * G$ are also filters. Beware of the following reversal: we write $F \leq G$ if and only if $F \supseteq G$.

- An *ultrafilter* U is a filter with the property that for each $a \in \mathcal{A}$ exactly one of $a, -a$ belongs to U.

- Let $\mathcal{A}$ have κ ultrafilters. A *U-network* N (over $\mathcal{A}$) is a directed graph with less than κ^+ nodes with each edge labelled by an ultrafilter such that for every triangle $(l, m, n) \in N$

  $$N(l, m) * N(m, n) \geq N(l, n).$$

- Given a class $\mathcal{K}$ of U-networks downward-closed and closed under isomorphism a *play* of a game will be a sequence

  $$N_0 \subseteq N_1 \subseteq \ldots N_\omega \subseteq \ldots N_\lambda \ldots$$

---

[17] The suggestion to generalise to the transfinite case was made by Hajnal Andreka.

where the length of the play is κ^+ and each network N_λ has size strictly less that κ^+. $\forall$ goes first and $\forall$ is always the first to play at a limit ordinal. The limit of the play is of course $\bigcup_{\lambda<\kappa^+} N_\lambda$. A win for $\exists$ can be defined by any property of the limit. A winning strategy for $\exists$ can be defined as the obvious generalisation of the countable version.

LEMMA 9 (Transfinite Conjunction Lemma) *Let $\exists$ win the game G_{P_λ} if and only if the limit has the property P_λ. Let σ_λ be a winning strategy for $\exists$ in the game G_{P_λ} (each $\lambda < \kappa^+$). Then $\exists$ has a winning strategy in the game G_P where $P = \bigwedge_{\lambda<\kappa^+} P_\lambda$, in other words, she can achieve all the properties at once.*

PROOF:

Suppose it is the λ'th round of the play, let $\forall$ choose N_λ. $\exists$ will make a sequence of λ extensions and since κ^+ is regular the union of these extensions will be smaller than κ^+ and so a legitimate move. First she pretends that her previous moves using σ_0 were her only moves and that all intervening moves were made by $\forall$ (again use regularity). Considering the play this way she uses σ_0 to make her first extension N^0_λ. For $\mu \leq \lambda$, her μ'th extension will be made by considering the play so far and identifying only her previous moves using σ_μ as her own moves and all others as belonging to $\forall$. She uses σ_μ to extend the network to N^μ_λ. After her λ'th extension her turn is finished. $\forall$ continues with the $\lambda + 1$'th round. Thus σ_λ is employed by $\exists$ in all rounds of the game from λ onwards, in other words κ^+ times. So the limit must satisfy the property P_λ for each λ.

$\Box$

THEOREM 10 *An atomic relation algebra $\mathcal{A}$ with $|Atoms_{\mathcal{A}}| = \lambda$ is completely representable if and only if there is a class of atomic networks*[18] *with the triangle addition property. In this case $\mathcal{A}$ has representations of size μ for each $\mu \geq \lambda^+$ and none of size less than λ. All the other results in theorem 5 go over to the uncountable case too.*

PROOF:

Same as theorem 5. $\Box$

Important Note If we follow corollary 6 and attempt to produce a set of logical axioms which are necessary and sufficient for an atomic relation algebra to be completely representable we must confront the following difficulty: the statement that the existence of a network of size λ implies the existence of an extension of size $\lambda+1$ cannot be translated into first order logic. Indeed it would appear to require λ different variables to state it and the

---

[18] Here we must take an atomic network to be a transitively closed graph, labelled by atoms of $\mathcal{A}$ and whose size is strictly less than λ^+. Note that in [Lyndon, 1950] with a correction in [Lyndon, 1956] it is shown that the class of completely representable relation algebras is not definable by universal sentences.

quantifiers could be nested to the same depth. This does not prove that it is impossible to characterise the completely representable relation algebras using less variables or smaller quantifier depth, but it does suggest that this is, in fact, the case.

Contrast this with Tarski's result [Tarski, 1954 and 1955] that the representable relation algebras form a variety and can therefore be axiomatised by equational axioms. This does not contradict the previous suggestion which referred to *completely* representable relation algebras.

LEMMA 11 *cf. Lemma 4. A U-network N is a representation of $\mathcal{A}$ if and only if for every pair of nodes $m, n \in N$ for every $a \in N(m, n)$ and for all relations (not necessarily atomic) $b, c \in \mathcal{A}$ such that $b * c \geq a$ there is some node $l \in N$ with $N(m, l) = b$ and $N(l, n) = c$.*

PROOF:

Suppose we have some U-network N that satisfies the condition. In order to show that it is a representation it is necessary to show that the mapping $I : \mathcal{A} \to \mathcal{P}(N \times N)$ defined by

$$I(a) = \{(m, n) \in N \times N : a \in N(m, n)\}$$

respects all the operators of the relation algebra. Since there are ultrafilters on each edge it follows that complement and finite unions are preserved, though arbitrary unions may not be. We show that composition is preserved, the results for converse and identity follow easily. Let $m, n \in N$ and $b, c \in \mathcal{A}$.

$(m, n) \models b * c$

$\Leftrightarrow \quad b * c \in N(m, n)$

$\Leftrightarrow \quad \exists l[N(m, l) = b \wedge N(l, n) = c]$

($\Leftarrow$ by transitive closure of N, $\Rightarrow$ by the condition in statement of lemma)

$\Leftrightarrow \quad \exists l[(m, l) \models b \wedge (l, n) \models c.$

It follows that $I(b * c) = I(b) * I(c)$ as required.

Conversely if N is any representation we have already seen that the relations between any two points in N forms an ultrafilter and the edge condition follows from the definition of composition. □

Definition A class $\mathcal{K}$ of U-networks is *good* if it is closed under isomorphism, downward closed and for any $N \in \mathcal{K}$ $\forall m, n \in N$

$$\forall a \in N(m, n)\, \forall b * c \geq a\, \exists N^+ \supset N$$
$$[N^+ \in \mathcal{K} \;\wedge\; \exists l \in N^+ \text{ such that } b \in N^+(m, l) \;\wedge\; c \in N^+(l, n)]$$

THEOREM 12 *A relation algebra $\mathcal{A}$ of arbitrary size is representable if and only there is a good class of U-networks.*

PROOF:

As before $\exists$ and $\forall$ play a (transfinite) game by choosing a nested sequence of U-networks from $\mathcal{K}$. $\exists$ wins the game if the limit is a representation of $\mathcal{A}$. If $\mathcal{K}$ is *good* then it can be seen that $\exists$ has a winning strategy and therefore $\mathcal{A}$ has a representation.

Conversely, given any representation X the class of all the U-networks formed by substructures of X of cardinality less than κ^+ is good.

□

THEOREM 13 *A relation algebra $\mathcal{A}$ is representable if and only if its ultrafilter extension $\mathcal{A}^+$ is.*

PROOF:

$\Rightarrow$ Let X be a representation of $\mathcal{A}$. For each ultrafilter U_λ of $\mathcal{A}$ add two new constant symbols c_λ, d_λ to the language. Consider the first-order theory

$$\Sigma = Th(\mathcal{A}) \cup \bigcup_\lambda \{r_\lambda(c_\lambda, d_\lambda) : r_\lambda \in U_\lambda\}$$

We use compactness to show that Σ is consistent. A finite subset of $\{r_\lambda c : r_\lambda \in U_\lambda\}$ is equivalent to the single formula $R_\lambda(c_\lambda, d_\lambda)$ where R_λ is the conjunction of the r_λs appearing in the finite set. So we have to show that a set of formulas

$$Th(\mathcal{A}) \cup \{R_1(c_1, d_1), R_2(c_2, d_2) \ldots R_n(c_n, d_n)\}$$

is consistent. But this set of formulas all hold in the model X, given appropriate interpretations of the constant symbols. So Σ has a model.

This model Y is a representation of $\mathcal{A}$ embedding all the ultrafilters of $\mathcal{A}$. To be a representation of the ultrafilter extension we require more: by lemma 4, Y is a representation of $\mathcal{A}^+$ if for all ultrafilters $V, W \in \mathcal{A}^+$ and for all $y_1, y_2 \in Y$

$$V * W(y_1, y_2) \to (\exists z \in Y)\ V(y_1, z) \wedge W(z, y_2) \qquad (*)$$

where $F(x, y)$ means that all the relations in the filter F hold on the pair (x, y).

Our strategy is to embed the representation Y in a larger representation Y_1 such that for all (V, W, y_1, y_2) where equation $(*)$ fails in Y there is a new point $x \in Y_1$ making that equation true. This extension Y_1 may contain new pairs of points where the equation fails, but these get repaired in the next extension. After ω iterations we take the limit and then show that it is a representation of $\mathcal{A}^+$.

Now in more detail, let $L(Y)$ be the first-order language $L(\mathcal{A})$ augmented with a new constant symbol c_y for each element $y \in Y$ and let $\triangle(Y)$ be the *diagram* of Y consisting of all quantifier-free $L(Y)$-formulas that hold in Y. Clearly Y is a model of $\triangle(Y)$ and Y embeds in any model of $\triangle(Y)$.

For each tuple (V, W, c_1, c_2) such that $(*)$ fails on the points named by c_1, c_2 add a new constant c to $L(Y)$ and define a new theory Γ to include $\Sigma \cup \triangle(Y)$ and all formulas $v(c_1, c)$ and $w(c, c_2)$ for all $v \in V, w \in W$ and this should be done for all tuples (V, W, c_1, c_2). As before, we can show that for any finite subset of Γ, Y is already a model hence Γ is consistent. Let Y_1 be a model of Γ. Since $Y_1 \models \Sigma$ it follows that Y_1 is a representation of $\mathcal{A}$ and $Y_1 \models \triangle(Y)$ implies that Y embeds in Y_1.

By iterating this procedure we get a chain of representations

$$Y = Y_0 \subseteq Y_1 \subseteq Y_2 \subseteq \ldots$$

(Since Y_n embeds isomorphically in Y_{n+1} we can assume without loss of generality that Y_{n+1} is actually an extension of Y_n.) Notice that Σ is an $\forall\exists$ theory and each Y_i is a model of Σ. Therefore the limit of the chain $Z = \bigcup_{i<\omega} Y_i$ is also a model of Σ and hence a representation of $\mathcal{A}$. Also, for all possible tuples (V, W, z_1, z_2) equation $(*)$ is satisfied in Z so Z is a representation of $\mathcal{A}^+$.

$\Leftarrow$ Since $\mathcal{A}$ is a subalgebra of its ultrafilter extension, any representation of the extension is automatically a representation of $\mathcal{A}$.

$\square$

Note The representation Z of $\mathcal{A}^+$ constructed in the proof is a complete representation. This is because Z is a representation of $\mathcal{A}$ so any pair of points (z_1, z_2) from Z are related by an ultrafilter of $\mathcal{A}$, that is, an atom of $\mathcal{A}^+$. By theorem 1 this must be a complete representation. So an arbitrary representable relation algebra $\mathcal{A}$ embeds in its ultrafilter extension $\mathcal{A}^+$ which has a complete representation. We cannot conclude, however, that $\mathcal{A}$ has a complete representation as the embedding from $\mathcal{A}$ to $\mathcal{A}^+$ might not respect arbitrary unions.

1.3.10 ω-categorical relation algebras

Definitions

1. A relation algebra $\mathcal{A}$ is *weakly ω-categorical* if it has a countably infinite representation and all such representations are isomorphic.
2. $\mathcal{A}$ is *strongly ω-categorical* if it has exactly one representation (up to isomorphism) of size ω or less.

3. Let $\mathcal{A}$ be a relation algebra and $\mathcal{K}$ a TAP-class of atomic $\mathcal{A}$-networks. Let $N^+ \in \mathcal{K}$ and let $N \in \mathcal{K}$ be obtained from N^+ be the deletion of a single point n (alternatively N^+ is a *one point extension* of N). The *Type* of n over N, written $Type(n, N)$ is defined to be the pair (N^+, n). We may also write this type as $Type(N^+, N)$ without any ambiguity. Intuitively, the type of a one point extension to N tells you exactly how the new point relates to all the old ones.

4. A one point extension N^+ of N (with nodes $N \cup \{n\}$) has the *d-support property* if there is a subnetwork L of N of size at most d such that the type of n over L *determines* the type of n over N. More precisely, if M is another one point extension of N with extra node m and if $Type(m, L) \cong Type(n, L)$ then $Type(m, N) \cong Type(n, N)$.

5. $\mathcal{K}$ has the *d-chain condition* if for all $N, M \in \mathcal{K}$ with $N \subseteq M$ there is a chain of networks
$$N = N_0 \subset N_1 \subset \ldots \subset N_k$$
such that $M \subseteq N_k$ and for each $i < k$, N_{i+1} has the d-support property over N_i.

6. $\mathcal{A}$ has the d-chain condition if every TAP-class of atomic $\mathcal{A}$-networks has the d-chain condition or equivalently if the maximal TAP-class $\tau(\mathcal{A})$ has the d-chain condition.

Examples

- The point algebra $\mathcal{P}$ has only one TAP-class, namely $\mathcal{K} = Net(\mathcal{P})$. An atomic network $N \in \mathcal{K}$ is effectively a finite linear order. If N^+ is a one point extension of N then the type of N^+ over N looks like one of the three diagrams below.

In the first and last case the extension N^+ has the 1-support property over N and in the second case it has the 2-support property. So $\mathcal{P}$ has the 2-chain condition.

- The Allen Interval Algebra $\mathcal{I}$. Again there is only one TAP-class $\mathcal{K} = Net(\mathcal{I})$. The one point extension, below, has the 4-support property but not the 3-support property.

It is not hard to see that any one point extension has the 4-support property. Although here the 2-support property fails we still get the 2-chain condition. In the diagram below there is a chain

$$N \subset N + \{l\} \subset N + \{l, m\} \subset N + \{l, m, n\}$$

which covers $N + \{n\}$ and where each network has the 2-support property over the previous.

$\mathcal{I}$ has the 2-chain condition.

- Let $\mathcal{D}$ be the algebra with two atoms: Id and $\sharp$ with composition $\sharp * \sharp = Id \vee \sharp$. Representations of this are just unstructured sets of size at least 3 where Id is interpreted as equality and $\sharp$ is interpreted as inequality. If N is a network with d unequal elements (all related to each other by $\sharp$) and N^+ is the one point extension with the extra point not equal to any of the others then N^+ does not have the $d-1$-support property over N. $\mathcal{A}$ does not have the d-support for any d. This is an example of a relation algebra which is weakly ω-categorical but not strongly ω-categorical - it has exactly one countably infinite representation but also non-isomorphic finite representations of size n for each $n \geq 3$.

Equivalence Games Strongly ω-categorical relation algebras can be characterised by properties of their networks: this is most elegantly done using a second variety of game[19]. Let $\mathcal{A}$ be a finite relation algebra and let τ be a TAP-class. Again there are two players, $\forall$ and $\exists$, who build a nested sequence of networks from τ of finite or countable length

$$N_0 \subseteq N_1 \subseteq N_2 \subseteq \ldots$$

Suppose N_i has been built ($i < \omega$). First, $\exists$ chooses $n, m \in N_i$ and atoms $a, b \in \mathcal{A}$ such that $N_i(n,m) \leq a * b$. Then, $\forall$ chooses from τ an extension N_{i+1} of N_i with at most one extra point so that there is some l in N_{i+1} with $N_{i+1}(n,l) = a$ and $N_{i+1}(l,m) = b$. $\forall$ can always do this since τ has the triangle addition property. This ends the round, the game continues from N_{i+1}.

Now let $N^+ \supseteq N$ belong to τ and suppose $|N^+ \setminus N| = 1$. The game $G(N^+, N)$ is played as above, starting with N. Let the play be $N = N_0 \subseteq N_1 \subseteq N_2 \subseteq \ldots$. $\exists$ wins the play if there is some $i < \omega$ and a point $n \in N_i$ such that $N_i \downarrow_{N_0 \cup \{n\}} \cong N^+$ and the isomorphism fixes N_0 pointwise. Otherwise, if there is no such i and n, $\forall$ wins the play.

THEOREM 14 *Let $\mathcal{A}$ be any finite relation algebra. The following are equivalent.*

1. *$\mathcal{A}$ has a countable, homogeneous representation and $\mathcal{A}$ is strongly ω-categorical*
2. *$\mathcal{A}$ has a unique TAP-class τ and for all one point extensions $N \subset N^+ \in \tau$ $\exists$ has a winning strategy for $G(N^+, N)$.*

PROOF:

$2 \Rightarrow 1$.

Assume the hypothesis of 2. Let X, Y be countable representations of $\mathcal{A}$. By assumption, X and Y have age τ. We are not assuming that $\mathcal{A}$ is integral and so let e be some atom contained in the identity and choose $x \in X$, $y \in Y$ such that the pairs (x,x) and (y,y) both satisfy the same relation e. Then the map $x \rightarrow y$ is a partial isomorphism from X to Y. So it suffices (by back and

---

[19] This idea and the theorem which follows are due to Ian Hodkinson.

forth) to show that if $\theta : X \to Y$ is a finite partial isomorphism and $x \in X$, then θ extends to x.

Let $\overleftarrow{X}\,(dom(\theta)) = N$ and $\overleftarrow{X}\,(dom(\theta) \cup \{x\}) = N^+$. We can assume that $N^+ \supset N$. Starting with $\overleftarrow{Y}\,(Range(\theta))$, $\forall, \exists$ play a game on Y. In the i'th round, if the play so far has been $\overleftarrow{Y}\,(Range(\theta)) = M_0 \subset M_1 \subset \ldots \subset M_i$ ($\subseteq Y$) (where each M_{k+1} is a one point extension of M_k), $\exists$ chooses $n, m \in M_i$ and atoms $a, b \in \mathcal{A}$ such that $Y \models (a * b)(n, m)$. $\forall$ then chooses $y \in Y$ with $Y \models a(n, y) \wedge b(y, m)$. This is always possible for $\forall$ as Y is a representation of $\mathcal{A}$. Letting $M_{i+1} = \overleftarrow{Y}\,(M_i \cup \{y\})$ the game continues.

This gives a play of the game $G(N^+, N)$ so $\exists$ can use her winning strategy. If she does so, then for some $i < \omega$ we find a point $z \in M_i$ such that $N^+ \cong \overleftarrow{Y}\,(M_0 \cup \{z\})$ and the isomorphism extends θ. Thus θ extends to $\theta^+ : X \to Y$ where $\theta^+ = \theta \cup \{(x, z)\}$. The same argument can be used to show that the countable representation is homogeneous.

$1 \Rightarrow 2$.

Clearly if $\mathcal{A}$ is ω-categorical, there is a unique TAP-class, τ say. Suppose, for contradiction, that $N, N^+ \in \tau$ are such that $\exists$ has no winning strategy for $G(N^+, N)$. Now let $\forall, \exists$ play $G(N^+, N)$ and let the play be $N = N_0 \subset N_1 \subset \ldots$. Let $\exists$ play so that $Y = \bigcup_{i<\omega} N_i$ is a representation of $\mathcal{A}$ and let $\forall$ play so that $\exists$ loses the game. $\exists$ can achieve her aim by making sure she eventually chooses every edge (n, m) which occurs in a network and every pair of atoms $a, b \in \mathcal{A}$ such that $Y(n, m) \leq a * b$. $\forall$ can achieve his goal because $\exists$ does not have a winning strategy.

Since $N^+ \in \tau$, N^+ embeds into Y and since Y is homogeneous we can assume the embedding takes N to $N_0 \subseteq Y$. But then $N^+ \subseteq N_i$ for some i, a contradiction.

Hence $\exists$ has a winning strategy as required.

□

COROLLARY 15 *If $\mathcal{A}$ is a finite relation algebra with the 2-chain condition then $\mathcal{A}$ is strongly ω-categorical.*

PROOF:

Let N be any network in $\tau(\mathcal{A})$ and let $N^+ = N \cup \{n\}$ be any one point extension. Assuming that $\mathcal{A}$ has the 2-chain condition we must prove that $\exists$ has a winning strategy for the game $G(N^+, N)$. Now the 2-chain condition ensures that there is a chain of networks

$$N = N_0 \subset N_1 \subset \ldots \subset N_k$$

such that $N^+ \subseteq N_k$ and for each $i < k$, N_{i+1} has the 2-support property over N_i.

This will mean that $\exists$ can force the play of the game $G(N^+, N)$ to be the chain $N = N_0 \subset N_1 \subset \ldots \subset N_k$, at least up to isomorphism and thus she has a strategy which guarantees she wins the game. To show that she can enforce this, suppose the chain has been constructed correctly up to N_i. N_{i+1} has the 2-support property over N_i so there are two nodes $a, b \in N_i$ which uniquely determine the type of the new node n over N_i. Let $N_{i+1}(a, n) = r$ and $N_{i+1}(b, n) = s$. If $\exists$ chooses nodes a and b and atoms r, s then $\forall$ can only respond with a network isomorphic to N_{i+1} and note that the isomorphism fixes N_i.

Thus $\exists$ wins the game and by the previous theorem, $\mathcal{A}$ is strongly ω-categorical.

□

COROLLARY 16 *If $\mathcal{A}$ has the 2-chain condition then $\mathcal{A}$ has at most one TAP-class.*

PROOF:

If $\mathcal{A}$ has two TAP-classes then it is possible to construct two non-isomorphic, countable representations and so $\mathcal{A}$ is not ω-categorical and therefore $\mathcal{A}$ does not have the 2-chain condition. □

The following corollary was first proved in [Ladkin and Maddux, 1987b].

COROLLARY 17 *The point algebra $\mathcal{P}$ is strongly ω-categorical.*

PROBLEM 7 *Find a finite relation algebra $\mathcal{A}$ with a unique TAP-class but which is not strongly ω-categorical. Such an algebra must have at least one inhomogeneous representation.*

PROBLEM 8 *Does the converse of corollary 15 hold? If a finite relation algebra has a TAP-class which does not have the 2-chain condition, does it follow that it has two countable non-isomorphic representations?*

Infinite relation algebras Theorem 14 applies only to finite relation algebras. When extended to countably infinite relation algebras we get the following, weaker, result.

THEOREM 18 *Let $\mathcal{A}$ be a countably infinite relation algebra. The following are equivalent.*

1. *$\mathcal{A}$ has a countable, homogeneous, complete representation and all its countable, complete representations are isomorphic*
2. *$\mathcal{A}$ has a unique TAP-class τ and for all one point extensions $N \subset N^+ \in \tau$ $\exists$ has a winning strategy for $G(N^+, N)$.*

But what about other (non-complete) representations? This theorem falls short of characterising the ω-categorical relation algebras.

PROBLEM 9 *Prove the following conjecture: an infinite, atomic relation algebra cannot be categorical in any cardinal.*

1.3.11 The networks of Allen's Point Algebra and the Metric Point System form Amalgamation Classes

The result for the point algebra is easily established[20], but it will be useful for the section on metric interval algebra to give an alternative proof, using the extension property (defined below), to show that both systems have a normal representation and then use theorem 5 to prove that they form an amalgamation class.

THEOREM 19 *Any complex group $\mathcal{A}$ has a normal representation. The complete representations of a complex group are all isomorphic (and so they have the same size).*

PROOF:

The first part was first proved in [Dechter *et al.*, 1991] (theorem 3.3). Here, we use the so-called *regular representation* which is defined like this. The domain of the representation is the set of atoms of $\mathcal{A}$ itself. Each relation a of $\mathcal{A}$ is interpreted as a binary relation on $atoms_{\mathcal{A}}$ by

$$X(a) = \{(a_1, a_2) \in atoms_{\mathcal{A}} : a_1^{\smile} * a_2 \leq_{\mathcal{A}} a\}$$

(Since the atoms of $\mathcal{A}$ form a group $a_1^{\smile} * a_2$ is itself an atom of $\mathcal{A}$, so every pair of elements of the domain are related by exactly one atomic element of $\mathcal{A}$.) It is a routine exercise to check that this does give a normal representation of $\mathcal{A}$.

Now let X and Y be any complete representations of $\mathcal{A}$. We show they are isomorphic by a back-and-forth construction which should proceed transfinitely if $\mathcal{A}$ has uncountably many atoms. Suppose $\theta : \bar{x} \to \bar{y}$ is a local isomorphism from X to Y and $x \in X$. As usual, we must extend θ to x. Pick any x_0 from $\bar{x}$ and let $y_0 = \theta(x_0)$. The relation $\overleftarrow{X}(x_0, x)$ is an atom of $\mathcal{A}$. Since a complex group is integral there must be some point $y \in Y$ such that $\overleftarrow{Y}(y_0, y) = \overleftarrow{X}(x_0, x)$. Define an extension to θ by $\theta^+ = \theta \cup (x, y)$. θ^+ is still a local isomorphism for the following reason: if x_i is any point from $\bar{x}$ and $y_i = \theta(x_i)$ then the atom $\overleftarrow{X}(x_i, x) = \overleftarrow{X}(x_i, x_0) * \overleftarrow{X}(x_0, x)$ (since the product of two atoms is an atom). Similarly $\overleftarrow{Y}(y_i, y) = \overleftarrow{Y}(y_i, y_0) * \overleftarrow{Y}(y_0, y)$. Therefore $\overleftarrow{Y}(y_i, y) = \overleftarrow{X}(x_i, x)$ which implies that θ^+ preserves all the edge relations and is a local isomorphism. $\Box$

COROLLARY 20 *The metric point system $\mathcal{M}$ has such a representation. This representation turns out to be $\mathbb{Q}$.*

PROBLEM 10 *Under what conditions can a relation algebra be extended to a complex group? McKenzie [McKenzie, 1970] showed that the subalgebras of complex groups cannot be finitely axiomatised relative to the integral, representable ones. The class of relation*

---

[20] Indeed we already know that $\mathcal{P}$ is ω-categorical and $\mathbb{Q}$ forms a universal, homogeneous representation, so $Net(\mathcal{P})$ forms an amalgamation class.

algebras whose networks form an amalgamation class is, similarly, not finitely axiomatisable relative to the integral, representable relation algebras (see problem 6). Question: can an integral relation algebra whose networks form an amalgamation class be extended to a complex group? Does the converse hold?

Definition Let $\mathcal{A}$ be a representable relation algebra and let N be a (not necessarily atomic) $\mathcal{A}$-network. A *model* of N is a homomorphism I from N into some representation X of $\mathcal{A}$. So for all $n, m \in N$ the relation $N(n,m)$ must hold between the pair $I(n), I(m)$. A (not necessarily atomic) finite network N has the *extension property*[21] if for any subnetwork M of N every model of M extends to a model of N.

Definition A *simple* $\mathcal{M}$-network is a transitively closed $\mathcal{M}$-network with a single interval on each arc. (In the terminology of [Dechter *et al.*, 1991] this is called an STP.)

THEOREM 21 *Any (transitively closed) simple $\mathcal{M}$-network has the extension property.*

PROOF:

Let N be such a network and let M be any subnetwork. Let I map the nodes of M to $\mathbb{Q}$ so that $I(m_2) - I(m_1) \in M(m_1, m_2)$. We must show that I can be extended to all the nodes of N. So, let $n \in N \backslash M$. In order to be consistent with the relation $N(m,n), n$ must be assigned a value from $S_m = I(m) + N(m,n)$. So the problem reduces to finding a value which belongs to S_m for all $m \in M$, in other words we must show that $\bigcap_{m \in M} S_m$ is non-empty. Now, each pair S_m, S_{m^*} intersect, otherwise the triangle (m, m^*, n) would not be transitively closed. So $\{S_m : m \in M\}$ is a finite set of convex sets of rationals which intersect pairwise. It is not hard to show by induction on $|M|$ (using convexity and linearity) that such a set must have a common intersection, and this allows us to assign a value to n and in a similar way, all the remaining nodes in N. $\square$

COROLLARY 22 *The point algebra $\mathcal{P}$ has a normal representation.*

PROOF:

The domain of the representation will be the regular representation of $\mathcal{M}$, namely $\mathbb{Q}$.

Fully universal. Let N be an atomic $\mathcal{P}$-network. Translate N into a simple $\mathcal{M}$-network by replacing '=' by $[0,0]$, '<' by $(0,\infty)$ and '>' by $(-\infty,0)$. The previous theorem allows us to find a model of this network in $\mathbb{Q}$.

---

[21] In [Dechter *et al.*, 1991] this property is called *decomposability.*

Homogeneous. Let $\bar{x}$ and $\bar{y}$ be locally isomorphic tuples in $\mathbb{Q}$. A full automorphism is constructed by a back and forth argument, each step using the extension property.

□

COROLLARY 23 *Both* $Net(\mathcal{P})$ *and* $Net(\mathcal{M})$ *form amalgamation classes.*

COROLLARY 24

1. *Any (transitively closed, not necessarily atomic)* $\mathcal{P}$*-network (N say) is* nearly minimal *i.e. for each arc and each relation, other than equality, on that arc there is a model of the network which realises that relation. (Note: this is weaker than the extension property, which does not hold for general* $\mathcal{P}$*-networks, see diagram below). Hence, any transitively closed* $\mathcal{P}$*-network is satisfiable.*
2. N nearly has the extension property *that is, for every subnetwork* M *of* N *and every model of* M *such that every node of* M *is given a* different value *there is an extension of the model to* N*.*

PROOF:

1. The following sketch is an alternative to the original proofs in [Villain and Kautz, 1986] and [van Beek and Cohen, 1990]. A corrected proof is in [Villain *et al.*, 1989]. First note that if any of the arcs has only the atomic relation '=' on it we may reduce the network by 'collapsing' the nodes together. So assume every arc has at least one of '<' and '>' on it. The proof is similar to the proof of theorem 21. Here we assign values in $\mathbb{Q}$ to nodes n of N, but ensuring that no two nodes are assigned the same value. As before it is necessary to show that $\bigcap_{m \in M} S_m$ is non-empty. But here the S_m may not be convex (consider $\{<, >\}$). Nevertheless, each S_m can be turned into a convex set S^*_m by adding at most a single point. Therefore $\bigcap_{m \in M} S^*_m$ is non-empty and it is not hard to see that it is, in fact, an infinite intersection. So removing the finite number of points added to the S_m will still leave an infinite intersection. Therefore it is possible to assign n a value, moreover this can be chosen to be different to all the previous values.
2. Similar

□

The following network, from [Villain *et al.*, 1989], is transitively closed but does not have the extension property. The model of $\{A, D\}$ where $A = D$ does not extend to the whole network. Thus the network is not complete, though it is satisfiable.

Chapter 2

Non-standard Representations

The axioms[1] for a relation algebra were intended to characterise the concrete algebras of binary relations [Tarski, 1941]. This failed [Lyndon, 1950] because there is no finite set of axioms which characterises this class correctly. The Tarski axioms are obviously sound over the class of concrete binary relation algebras but they are not complete: there are (finite) structures satisfying all the Tarski axioms which possess no such representation. McKenzie found the smallest of these with a relation algebra with only four atoms possessing no representation as a proper relation algebra [McKenzie, 1970].

Let us stand back, for a moment, from the immediate issue of representing relation algebras and consider more widely the area of general algebras and representations. Given some class of algebraic structures what is a representation? Certainly it will be some kind of isomorphism from an algebraic structure to some other structure, but there may be other properties that representations should have. Three classic examples of representations are

- the Stone representation of boolean algebras. Here a boolean algebra $\mathcal{B} = (B, \vee, -, 1, 0)$, considered purely as an abstract algebraic structure, is mapped (using an ultrafilter extension) to a field of sets together with set-theoretic operations $\mathcal{S} = (S, \cup, \setminus, 1, \emptyset)$.
- the representation of groups by matrices. In this case each group element is represented as an $n \times n$ matrix (some fixed n) over some field and the group operation is interpreted as matrix multiplication.
- The relation between a purely syntactical, logical theory and a model. Here a model will be a function from logical symbols to some set (the domain of the model) and the logical connectives are interpreted as set-theoretic operations. (But note that this means that the concept of a 'model of set-theory' is quite problematic as the universal class of set-theory is represented as a set.)

In each case we have a mapping from one structure to another preserving the operations. In the first two cases this map is an isomorphism and, further, the algebraic operations get a uniform interpretation for the whole class of structures: for boolean algebras $\vee$ is mapped

---

[1] The work on non-standard representations is joint work with Dov Gabbay.

to $\cup$ and for groups there is a fixed method of calculating the matrix product (though admittedly this involves the field operations and the field does not have to be the same for each group representation). Moreover the representation should offer some advantage either by mapping a difficult structure to one that is well-understood or, perhaps, by replacing an abstract algebra by a more concrete structure. It should always be the case that the operations in the representation are independent of the algebra to be represented. In this chapter we also require that the operations have first-order definitions.

Returning to relation algebras, we can get round the representation problem by generalising the definition of a representation and the dimension of the representation. Fix the dimension as n (in the standard case $n = 2$). We assume throughout that a relation algebra is *finite* and *integral* (no proper divisors of zero) and under the latter condition it follows that the identity, Id is an atom.

2.1 Definitions

- The atomic formulas of n-adic first-order logic are of the form

  $$x = y$$

  $$R(\bar{x}) \text{ or}$$

  $$C(\bar{x})$$

  for some n-ary predicate symbol R or constant symbol C and n-tuple of variables $\bar{x}$. Other formulas of the language are built from atoms using Boolean connectives and first order quantifiers. Quantification over predicates ($\exists R[\ldots]$) is not allowed.

- An (n-dimensional) *interpretation* $\mathcal{H}$ consists of five n-adic, first-order formulas: $Disj(R, S, \bar{x})$, $Neg(R, \bar{x})$, $Eq(\bar{x}), Conv(R, \bar{x})$ and $Comp(R, S, \bar{x})$.

  So the standard interpretation is a 2 dimensional example of this where

  $$\begin{array}{rcl} Disj(R, S, x, y) & \equiv & R(x, y) \vee S(x, y) \\ Neg(R, x, y) & \equiv & 1(x, y) \backslash R(x, y)) \\ Eq(x, y) & \equiv & (x = y) \\ Conv(R, x, y) & \equiv & R(y, x) \\ Comp(R, S, x, y) & \equiv & \exists z[R(x, z) \wedge S(z, y)] \end{array}$$

  Note that even with standard representations the top element of the Boolean algebra 1 is not necessarily the universal relation $D \times D$ (see [Lyndon, 1950]) so the negation formula uses 1 in order to define relative complement. When dealing with non-standard definitions we can use the constant symbol 1 to effectively give us a limited form of quantification over predicates. Instead of the illegal formula

  $$\exists R[R(\bar{x})]$$

we write the legal one

$$1(\bar{x}).$$

Provided $1 = \bigcup \mathcal{A}$ (which is ensured by the axioms of boolean algebra) the two formulas are equivalent. This does not permit arbitrary quantification, only existential quantification over atomic formulas.

- Given an interpretation $\mathcal{H}$ a *representation* of a relation algebra $\mathcal{A}$ is a structure $\mathcal{M} = (X, D)$ where D is the domain and X interprets all the predicates and constants i.e.

  $$X : \mathcal{A} \to \mathcal{P}(D^n)$$

  such that

  $$\begin{aligned} X(R \vee S) &= \{\bar{x} : \mathcal{M} \models Disj(R, S, \bar{x})\} \\ X(-R) &= \{\bar{x} : \mathcal{M} \models Neg(R, \bar{x})\} \\ X(Id) &= \{\bar{x} : \mathcal{M} \models Eq(\bar{x})\} \\ X(R^{\smile}) &= \{\bar{x} : \mathcal{M} \models Conv(R, \bar{x})\} \\ X(R * S) &= \{\bar{x} : \mathcal{M} \models Comp(R, S, \bar{x})\}. \end{aligned}$$

  In other words X is an isomorphism.

2.2 Ternary Representations

2.2.1 The Idea

In this section a non-standard interpretation is given: abstract relation algebras are represented as *ternary* relations.

Given an atomic, integral relation algebra $\mathcal{A}$ the elements of the algebra will be represented as *sets of triples of atoms* of $\mathcal{A}$. For any $R \in \mathcal{A}$ define the representation X by

$$X(R) = \{\text{atoms}(a, b, c) : a \leq R \,\&\, a \leq b * c\}$$

The first element of a triple acts as a label for an atom contained in the relation.

2.2.2 A ternary interpretation

We define five triadic formulas (see section 2.1) thus.

$$\begin{aligned} Disj(R, S, x, y, z) &\equiv R(x, y, z) \vee S(x, y, z) \\ Neg(R, x, y, z) &\equiv 1(x, y, z) - R(x, y, z) \\ Eq(e, y, z) &\equiv 1(e, y, z) \wedge \forall w[1(w, e, w) \wedge 1(w, w, e)] \\ Conv(R, x, y, z) &\equiv 1(x, y, z) \wedge \exists e, x'[Eq(e, x, x') \wedge Eq(e, x', x) \wedge R(x', e, x')] \\ Comp(R, S, x, y, z) &\equiv 1(x, y, z) \wedge \exists e, r, s[R(r, e, r) \wedge S(s, e, s) \wedge 1(x, r, s)] \end{aligned}$$

This interpretation will be referred to as $\mathcal{H}$.

2.2.3 Completeness

Given this ternary interpretation, every relation algebra $\mathcal{A}$ has a representation $\mathcal{M} = (X, atoms_{\mathcal{A}})$ where

$$\begin{aligned} X(R) &=_{def} \{\text{atoms}(a,b,c) : a \leq R \,\&\, a \leq b * c\} \text{ and} \\ X(1) &=_{def} \{\text{atoms}(a,b,c) : a \leq b * c\}. \end{aligned}$$

For

- $$\begin{aligned} (a,b,c) \in X(R \vee S) &\Leftrightarrow a \leq R \vee S, a \leq b * c \\ &\Leftrightarrow a \leq R \text{ or } a \leq S \text{ (since } a \text{ is an atom,) } a \leq b * c \\ &\Leftrightarrow (a,b,c) \in X(R) \cup X(S) \\ &\Leftrightarrow \mathcal{M} \models Disj(R,S,a,b,c) \end{aligned}$$

- $$\begin{aligned} (a,b,c) \in X(-R) &\Leftrightarrow a \leq (-R), a \leq b * c \\ &\Leftrightarrow (a,b,c) \in X(1) \wedge a \not\leq R \\ &\Leftrightarrow \mathcal{M} \models Neg(R,a,b,c) \end{aligned}$$

- $$\begin{aligned} (a,b,c) \in X(Id) &\Leftrightarrow a = Id \wedge a \leq b * c \text{ (since } Id \text{ is an atom)} \\ &\Leftrightarrow \forall x[a * x = x * a = x] \wedge a \leq b * c \\ &\Leftrightarrow \mathcal{M} \models Eq(a,b,c) \end{aligned}$$

- $$\begin{aligned} (a,b,c) \in X(R^{\smile}) &\Leftrightarrow a \leq R^{\smile}, a \leq b * c \\ &\Leftrightarrow a^{\smile} \leq R, a * a^{\smile} \geq Id, a \leq b * c \\ &\Leftrightarrow \exists e, a'[\mathcal{M} \models Eq(e,a,a') \wedge Eq(e,a',a) \wedge R(a',e,a')], a \leq b * c \\ &\Leftrightarrow \mathcal{M} \models Conv(R,a,b,c) \end{aligned}$$

- $$\begin{aligned} &(a,b,c) \in X(R * S) \\ &\quad \Leftrightarrow a \leq R * S, a \leq b * c \\ &\quad \Leftrightarrow \exists atoms\, r \leq R, s \leq S : a \leq r * s \\ &\qquad \text{(since } R * S = \textstyle\bigcup_{atoms\, r \leq R, s \leq S} r * s\text{) for finite } \mathcal{A} \\ &\quad \Leftrightarrow \exists r, s : (r,e,r) \in X(R), (s,e,s) \in X(S) \text{ and } (a,r,s) \in X(1) \\ &\quad \Leftrightarrow \mathcal{M} \models Comp(R,S,a,b,c) \end{aligned}$$

2.2.4 The class $RA(3)$

Every atomic, integral relation algebra can thus be represented by sets of triples, but not every collection of sets of triples closed under the operations is necessarily isomorphic to a relation algebra. For that purpose we define the class $RA(3)$ of algebras of triples, so that the completeness result still holds (it is straightforward to check that a representation of a relation algebra, as defined in section 2.2.3, will be in $RA(3)$) and we get soundness too: every member of $RA(3)$ forms a relation algebra.

Definition An *algebra of triples* $\mathcal{T}$ over a domain D has the form $(T, 1)$ where $T \subseteq \mathcal{P}(D \times D \times D)$ (i.e. the elements of T are ternary relations on D), $1 \in T$ such that all the axioms, below, are valid. $\mathcal{T}$ may be considered as a relation algebra by defining

$$R * S =_{def} \{(x, y, z) : \mathcal{T} \models Comp(R, S, x, y, z)\}$$

and making similar definitions for the other operators (so $\vee$ and $-$ are just union and relative complement of sets).

$\forall R, S \in T, \forall a, b, c \in D$

1. T is closed under the five operations.
2. $\exists! e \in D, \forall x \in D[(x, x, e), (x, e, x) \in 1]$ (write $e = Id$)
3. $\forall x \exists! x'[1(Id, x, x') \wedge 1(Id, x', x)]$
4. $1(a, b, c) \wedge R(a, x, y) \rightarrow R(a, b, c)$
5. $1(Id, a, a') \wedge 1(Id, b, b') \wedge 1(Id, c, c') \rightarrow (1(a, b, c) \leftrightarrow 1(a', c', b')$ $((R * S)^{\smile} = S^{\smile} * R^{\smile})$
6. $1(a, b, c) \wedge 1(c, d, e) \Rightarrow \exists f : 1(f, b, d) \wedge 1(a, f, e)$ and
   $1(f, b, d) \wedge 1(a, f, e) \Rightarrow \exists c : 1(a, b, c) \wedge 1(c, d, e)$ (Associativity)
7. $1(Id, b, b') \wedge 1(Id, a, a') \rightarrow (1(d', a, b) \leftrightarrow 1(b', d, a))$ (triangle rule)

Let $RA(3)$ be the class of all algebras of triples.

2.2.5 Soundness of the Relation Algebra Axioms over $RA(3)$

.

It is necessary to check that the nine Tarski axioms of relation algebras hold in any algebra of triples. This is done for three cases, the others being left as exercises.

Axiom: $\forall R, S((R * S)^{\smile} = S^{\smile} * R^{\smile})$

Let R, S belong to an algebra of triples $\mathcal{T}$, let $(a,b,c) \in 1$.

$$\begin{aligned}
&(a,b,c) \in S^{\smile} * R^{\smile} \\
&\quad \Leftrightarrow \quad \mathcal{T} \models Comp(S^{\smile}, R^{\smile}, a, b, c) \\
&\quad \Leftrightarrow \quad \exists e, \sigma, \rho[(a,b,c) \in 1, (\sigma, e, \sigma) \in S^{\smile}, (\rho, e, \rho) \in R^{\smile} \;\&\; (a, \sigma, \rho) \in 1] \\
&\qquad\quad\; \text{(by definition of } Comp\text{)} \\
&\quad \Leftrightarrow \quad \exists e, r, s, \rho, \sigma, \alpha[(a,b,c) \in 1, (e,\sigma,s), (e,\rho,r), (e,\alpha,a) \in Id, (s,e,s) \in S, (r,e,r) \in R \\
&\qquad\quad\; \text{and } (a,\sigma,\rho) \in 1] \text{ (definition of converse, axiom 2)} \\
&\quad \Leftrightarrow \quad \exists \alpha, r, s[(e,\alpha,a) \in Id, (\alpha, r, s) \in 1, (r,e,r) \in R, (s,e,s) \in S] \text{ (axiom 5)} \\
&\quad \Leftrightarrow \quad (a,b,c) \in (R * S)^{\smile} \text{ (definition of converse).}
\end{aligned}$$

Axiom: $*$ is associative.

$$\begin{aligned}
(a,b,c) \in (R*S)*T \quad &\Leftrightarrow \quad \exists x, t : (x,e,x) \in R*S, (t,e,t) \in T \;\&\; (a,x,t) \in 1 \\
&\Leftrightarrow \quad \exists r, s, t, x : (r,e,r) \in R, (s,e,s) \in S (t,e,t) \in T, \\
&\qquad\quad (x,r,s) \in 1 \;\&\; (a,x,t) \in 1 \\
&\Leftrightarrow \quad \exists r, s, t, y : (a,r,y), (y,s,t) \in 1 \text{ (axiom 6)} \\
&\Leftrightarrow \quad (a,b,c) \in R * (S * T)
\end{aligned}$$

Axiom: (triangle rule): $R * S \cap T^{\smile} = 0 \leftrightarrow S * T \cap R^{\smile} = 0$.

$$\begin{aligned}
(a,b,c) \in R*S \cap T^{\smile} \quad &\Leftrightarrow \quad \exists r, s, t : (r,e,r) \in R, (s,e,s) \in S, (t,e,t) \in T \\
&\qquad\quad (a,r,s) \in 1 \;\&\; (e,t,a) \in Id \\
&\Leftrightarrow \quad \exists r, s, t, \rho : (e,t,a), (e,\rho,r) \in Id \;\&\; (a,r,s) \in 1 \\
&\Leftrightarrow \quad \exists r, s, t, \rho : (e,t,a), (e,\rho,r) \in Id, \;\&\; (\rho,s,t) \in 1 \text{ (axiom 7)} \\
&\Leftrightarrow \quad \exists (\rho, s, t) \in R^{\smile} \cap S * T
\end{aligned}$$

2.3 Generalising the method

We have shown that every atomic, integral relation algebra can be represented as an algebra of triples and conversely that every algebra of triples is isomorphic to a relation algebra. One may be tempted to ask whether there is any advantage to be obtained from this representation, or indeed, whether there is any essential difference between a relation algebra (which is effectively a table of converses and compositions of atoms) and the ternary representation defined here. It is worth observing that each relation algebra must come equipped with an explicit definition of the operators $*$ and $^{\smile}$. This contrasts with its ternary representation where we have a *uniform* definition of the operators ($\mathcal{H}$) which is the same for every algebra of triples. This works because we code the operations in the representation of each element of the relation algebra.

This idea generalises to representations of Boolean algebras with operators, provided we have distribution laws of the operators over $\vee$. This allows us to represent the algebra by considering the atoms only. We need a standard representation of the Boolean operations in order that 1 is interpreted as the union of all the elements. This allows us to sneak in (effective) existential quantification of atomic formulas without actually writing a formula $\exists R[R(\bar{x})]$.

Further example - groups Given a group $(G, *)$ we can convert it to a Boolean algebra with operators by considering $(\mathcal{P}(G), 1, \cup, \setminus, *)$. Now we can represent any such structure in three dimensions by defining an interpretation of the operators.

$$\begin{aligned} Disj(S,T,x,y,z) &\equiv S(x,y,z) \vee T(x,y,z) \\ Neg(S,x,y,z) &\equiv 1(x,y,z) - S(x,y,z) \\ Comp(S,T,x,y,z) &\equiv 1(x,y,z) \wedge \exists e, \bar{s}, \bar{t}[S(\bar{s}) \wedge T(\bar{t}) \wedge 1(x, s_0, t_0)] \end{aligned}$$

where $\bar{s} = (s_0, s_1, s_2)$.

With this interpretation, we can represent $(\mathcal{P}(G), 1, \cup, \setminus, *)$ by mapping

$$\begin{aligned} I(1) &=_{\text{def}} \{(f,g,h) : g * h = f\} \\ I(S) &=_{\text{def}} \{(f,g,h) : f \in S \text{ and } g * h = f\} \end{aligned}$$

for each $S \subseteq G$.

As with relation algebras it would be necessary to put further restrictions on structures of triples in order to guarantee that each one was the representation of some group.

2.4 Reducing the dimension

Under certain circumstances it is possible to reduce the dimension of the representation. Some (but not all) of the relation algebras possess representations as *binary relations* i.e. standard representations. Characterising this class by a finite set of axioms cannot be done.

A group can always be represented in two dimensions by

$$I(S) = \{(f,g) : \exists s \in S, f * s = g\}$$

and interpreting $Comp$ as

$$Comp(S,T,x,y) = \exists z[S(x,z) \wedge T(z,y)]$$

Recalling that the atoms of a complex group form a group we see that any subalgebra of a complex group has a standard representation. We cannot claim this as one of the more significant breakthroughs in the study of representable relation algebras as this representation is the same regular representation defined on page 42. Nevertheless, the representability of relation algebras is part of a more general problem of identifying conditions when the dimension of a representation can be reduced.

Note on Tarski-Johnsson Non-Standard Representation In [Jónsson and Tarski, 1948] it is shown that every relation algebra has a two-dimensional non-standard representation. They show first that every relation algebra embeds into a complete, atomic relation algebra (their theorem 4.21) and so they are able to restrict to the atomic case without loss of generality.

Given an atomic relation algebra $\mathcal{A}$ the domain of their representation is the set of atoms of $\mathcal{A}$. They represent an element $x \in \mathcal{A}$ by

$$F(x) = \{(a, b) : a, b \text{ are atoms and } a \leq x * b\}.$$

Under this mapping all the operations of the relation algebra get interpreted as the standard set operations except complement[2] which has a non-standard definition. Let R be a binary relation on $atoms_{\mathcal{A}}$ and suppose $R = F(x)$ for some $x \in \mathcal{A}$. Let $\sim R =_{\text{def}} F(-(x))$. Under this definition they show (theorem 4.22) that every relation algebra has such a representation.

This representation has the advantage that all the operations, except complement, are given natural definitions, but it suffers one draw-back. The complement operator $\sim$ is not independent of the relation algebra: in order to calculate the complement of a binary relation it is necessary to look it up in the original relation algebra. Allowing this kind of definition is far too powerful, indeed take any algebraic structure B and take *any* map h from B to some domain D. If (say) $\sharp$ is a binary operation of the algebra and if $h(b_1), h(b_2)$ are two elements of the domain then let $\sharp_D(h(b_1), h(b_2)) =_{\text{def}} h(\sharp(b_1, b_2))$. This gives a non-standard representation, in fact the trivial identity map from an algebra to itself is a non-standard representation in which every operator is interpreted as itself!

No, this will not do. The operations in the representation must *not* depend on the relation algebra: as in the ternary representations (above) we must have a *uniform* representation in which the operators have a fixed definition independent of the relation algebra.

The Tarski-Jonsson representation has one fortunate property which might help though. Under their definition the non-standard complement of a set is always the smallest set in the representation such that the union of a set and its complement is the universal relation 1. This follows from the fact that they have a standard interpretation of union. This non-standard complement might not be the true complement because the complement of a binary relation might not be the representation of any element of the relation algebra.

Now if we *define* $\sim R$ to be the smallest relation S such that $R \cup S = 1$ then we do appear to have a uniform definition independent of the relation algebra but it is, unfortunately, a second-order definition. This disqualifies it as a representation according to the definitions given on page 47.

PROBLEM 11 *Relation algebras can always be represented in three dimensions. Using the standard interpretation it is not always possible to represent a relation algebra in two dimensions. Is there a two dimensional non-standard interpretation of the operators which is*

---

[2] In [Jónsson and Tarski, 1952] the boolean operations are, in fact, union and intersection and it is intersection which gets a non-standard treatment.

capable of representing all relation algebras? Such a representation must have a first-order definition, uniform for the class of all relation algebras and independent of the particular relation algebra to be represented. (Probably not).

Part II

Intervals

Chapter 3

From Points to Intervals

In[1] chapter I it was shown that certain standard representations could be classified as universal or homogeneous or both. These properties were characterised by properties of the networks of the relation algebra. Here we demonstrate the value of this classification by showing that from any relation algebra with a normal representation it is possible to construct a relation algebra of intervals. In particular the Allen Interval Algebra can be built from the Point Algebra and starting with the metric system of Dechter, Meiri and Pearl, it is possible to construct an interval system which represents metric information. This latter is preferable to other attempts to combine intervals and metrics [Ladkin, 1989, Meiri, 1991] because it is done in an entirely uniform way while the complexity of the propagation algorithm is the same as that of its rivals. Our system is likely to offer some efficiency advantage because of a simple rule for removing some disjuncts.

3.1 Intervals from Points

In this section we devise a construction which builds an 'interval' algebra from a 'point' algebra. The 'point' algebras which are suitable are those with normal representations - far more general than the usual three atomic relations on a linearly ordered set. Let $\mathcal{A}$ be an atomic relation algebra with a normal representation X. Fix one atomic relation r from $\mathcal{A}$ (exemplary case: $<$ in $\mathcal{P}$). For $x, y \in X$ let $\overleftarrow{X}(x, y)$ be the atomic relation which holds between x and y in X.

Definitions

1. The set of r-intervals in X, denoted X_r^2 is the set of ordered pairs (x, y) such that $(x, y) \models r$.

   Let α and β be r-intervals in X, respectively (α_1, α_2) and (β_1, β_2).

---

[1] Much of this section appears in [Hirsch, 1994]

2. Define the *atomic interval relation* $I(\alpha,\beta)$ between α and β as the 2 by 2 matrix $\left[\begin{array}{cc} \overset{\leftarrow}{X}(\alpha_1,\beta_1) & \overset{\leftarrow}{X}(\alpha_1,\beta_2) \\ \overset{\leftarrow}{X}(\alpha_2,\beta_1) & \overset{\leftarrow}{X}(\alpha_2,\beta_2) \end{array}\right]$.

3. The set of atomic interval relations over $\mathcal{A}$ is the set of all atomic 2 by 2 matrices $\left[\begin{array}{cc} a & b \\ c & d \end{array}\right]$ such that

$$a * c^{\smile} \geq r$$
$$b * d^{\smile} \geq r$$
$$a^{\smile} * b \geq r$$
$$c^{\smile} * d \geq r.$$

In other words the network below must be transitively closed.

We have now defined the atoms of the new interval algebra. Define the non-atomic interval relation

$$\left[\begin{array}{cc} a & b \\ c & d \end{array}\right] =_{\text{def}} \bigcup\{\text{atomic interval relations } \left[\begin{array}{cc} \alpha & \beta \\ \gamma & \delta \end{array}\right] : \mathit{atoms}\alpha \leq a, \beta \leq b, \gamma \leq c, \delta \leq d\}$$

Given a set of atoms there are various ways of extending them to form a relation algebra, or indeed a Boolean algebra. We could take all finite and co-finite unions of atomic interval relations, or the full power set, but here we define our algebra $\mathcal{A}_r^2$ as the set of finite unions of *basic interval relations* of the form $\left[\begin{array}{cc} a & b \\ c & d \end{array}\right]$ where a, b, c, d are not necessarily atomic. From this definition it follows for infinite $\mathcal{A}$ that $\mathcal{A}$ and $\mathcal{A}_r^2$ have the same cardinality[2].

Now given any two r-intervals there is exactly one atomic interval relation which holds between them[3] in X_r^2. The main problem in showing that $\mathcal{A}_r^2$ is a relation algebra arises because it is not automatic that the composition of two atomic interval relations will be a union of atomic interval relations.

---

[2] But note that $\mathcal{A}_r^2$ is not complete.

[3] In order that two intervals are related by an atomic interval relation it is necessary that X is a complete representation.

THEOREM 25 *Let $\mathcal{A}$ be a relation algebra with at most countably many atoms and let X be any normal representation of $\mathcal{A}$.*

1. *$\mathcal{A}_r^2$ forms a relation algebras. $\vee$ is of course interpreted as union and the complement is calculated by*

$$-\begin{bmatrix} a & b \\ c & d \end{bmatrix} =_{def} \begin{bmatrix} -a & 1 \\ 1 & 1 \end{bmatrix} \vee \begin{bmatrix} 1 & -b \\ 1 & 1 \end{bmatrix} \vee \begin{bmatrix} 1 & 1 \\ -c & 1 \end{bmatrix} \vee \begin{bmatrix} 1 & 1 \\ 1 & -d \end{bmatrix}.$$

   The identity relation is $\begin{bmatrix} Id_{\mathcal{A}} & r \\ r^{\smile} & Id_{\mathcal{A}} \end{bmatrix}$

   The inverse of an atomic interval relation is given by

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{\smile} = \begin{bmatrix} a_{11}{}^{\smile} & a_{21}{}^{\smile} \\ a_{12}{}^{\smile} & a_{22}{}^{\smile} \end{bmatrix}.$$

   The composition of any two atomic interval relations is calculated thus:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} * \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11} * b_{11} \cap a_{12} * b_{21} & a_{11} * b_{12} \cap a_{12} * b_{22} \\ a_{21} * b_{11} \cap a_{22} * b_{21} & a_{21} * b_{12} \cap a_{22} * b_{22} \end{bmatrix}.$$

   in other words, ordinary matrix multiplication with $\cap, *$ in place of $+, \times$ respectively.

2. *X_r^2 forms a normal representation of $\mathcal{A}_r^2$.*

PROOF:

1. There is a potential problem because there is no distribution law for intersection ($a*(b\cap c) = a*b\cap a*c$ is not generally true in relation algebras) and consequently the matrix multiplication may not be associative. But for the restricted case where we have a normal representation the matrix multiplication turns out to be associative. We show this by proving that the matrix product $R*S$ calculates the composition of the interval relations R and S in the representation correctly. This will also prove that the composition of two interval relations is a union of atomic interval relations. To do this we need a lemma.

LEMMA 26 *Let $\mathcal{A}$ be an atomic relation algebra with a normal representation X, let R and S be the matrices of atomic interval relations and let α and β be any r-intervals from X_r^2. Then $(\alpha, \beta) \in X_r^2(R*S)$ (the matrix product) iff there is some r-interval $\gamma \in X_r^2$ such that $(\alpha, \gamma) \in X_r^2(R)$ and $(\gamma, \beta) \in X_r^2(S)$.*

Proof of Lemma

$\Leftarrow$

This part of the lemma always holds for any representation X. Suppose $(\alpha, \gamma) \in X_r^2(R)$ and $(\gamma, \beta) \in X_r^2(S)$. $(\alpha, \gamma) \in X_r^2(R)$ means that $\overset{\leftarrow}{X}(\alpha_i, \gamma_j) \in R_{ij}(i, j = 1, 2)$. Similarly $\overset{\leftarrow}{X}(\gamma_j, \beta_k) \in S_{jk}(j, k = 1, 2)$. Considering the triangle $(\alpha_i, \gamma_j, \beta_k)$ in the representation X, it must be the case that $\overset{\leftarrow}{X}(\alpha_i, \beta_k) \in R_{ij} * S_{jk}(i, j, k = 1, 2)$.

Therefore $\overset{\leftarrow}{X}(\alpha_i, \beta_k) \in R_{ij} * S_{1k} \cap R_{i2} * S_{2k}$ which is the entry calculated by the matrix product. Hence $(\alpha, \beta) \in X_r^2(R * S)$.

$\Rightarrow$

Suppose; $(\alpha, \beta) \in X_r^2(R * S)$. Let the atomic interval relation between α and β in X_r^2 be T $(T \subseteq R * S)$. Since R is an interval relation, the network

is consistent. By considering each triangle, the following network is transitively closed.

Since the representation is fully universal[4] it follows that this network occurs in the representation and homogeneity ensures that the local isomorphism from this occurrence to

extends to a full automorphism of the representation. Therefore there is an r-interval $\gamma = (\gamma_1, \gamma_2)$ such that $(\alpha, \gamma) \in X_r^2(R)$ and $(\gamma, \beta) \in X_r^2(S)$ as required.

Proof of Theorem (continued).

Let $R, S, T \in \mathcal{A}_r^2$. Consider the interval relation $(R * S) * T$. In X_r^2 this is represented as the composition $(X_r^2(R) \circ X_r^2(S)) \circ X_r^2(T)$ where $\circ$ is just composition of relations (use previous lemma). Composition of binary relations is always associative, so this is equal to $X_r^2(R) \circ (X_r^2(S) \circ X_r^2(T))$ which is the representation of $R * (S * T)$ in X_r^2. Hence matrix multiplication in $\mathcal{A}_r^2$ is associative. The other axioms are straightforward to check.

2. We have already seen that X_r^2 is a representation of $\mathcal{A}_r^2$. It is easy to show that it is fully universal by
    - taking any interval network

---

[4] In fact we do not need full universality, only that networks of size six embed in the representation. Similarly, instead of homogeneity we need only require that local isomorphisms whose domain has size four extend to local isomorphisms of size six.

- converting to a network of points
- using the full universality of X to find the network in X and
- convert back to X_r^2 thus showing that the original network appears in X_r^2.

Homogeneity is handled in much the same way.

□

Important note

- If $\mathcal{A}$ has a normal representation X then $\mathcal{A}_r^2$ is isomorphic to the proper relation algebra formed on X_r^2 by the boolean operations, identity, converse and ordinary composition of relations.

- If X fails to be fully universal or homogeneous then the proper relation algebra on X_r^2 may not be isomorphic to $\mathcal{A}_r^2$: it could be that some atomic interval relation may not exist at all in X_r^2; the matrix multiplication may not calculate the correct composition of two interval relations and the interpretations of the 'atomic interval relations' may not in reality be atomic. X_r^2 may not be a representation of $\mathcal{A}_r^2$.

- If $\mathcal{A}$ does not have *any* normal representation then $\mathcal{A}_r^2$ might not form a relation algebra as the matrix composition may not be associative.

COROLLARY 27

1. *The Allen Interval Algebra can be constructed from the point algebra $\mathcal{P}$ and has a normal representation as ordered pairs of rationals. It is isomorphic to the relation algebra with atomic relations the set of two by two matrices* $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ *with elements taken from* $\{<,>,=\}$ *such that* $a * c^{\smile} \geq$ *'<'*; $b * d^{\smile} \geq$ *'<' etc. The atomic interval relations and their corresponding matrices are*

$$\text{'equals'} \begin{bmatrix} = & < \\ > & = \end{bmatrix};$$

$$\text{'before'} \begin{bmatrix} < & < \\ < & < \end{bmatrix};$$

$$\text{'meets'} \begin{bmatrix} < & < \\ = & < \end{bmatrix};$$

$$\text{'overlaps'} \begin{bmatrix} < & < \\ > & < \end{bmatrix};$$

$$\text{'starts'} \begin{bmatrix} = & < \\ > & < \end{bmatrix};$$

$$\text{`during'} \quad \begin{bmatrix} > & < \\ > & < \end{bmatrix};$$

$$\text{`ends'} \quad \begin{bmatrix} > & < \\ > & = \end{bmatrix}$$

plus the converses of the last six.

2. *It is possible to take the interval algebra, fix any one atomic relation say 'overlaps' and then define a relation algebra of 'intervals of intervals'. Here an interval will be any pair of intervals i, j such that i overlaps j.*

3.2 Points from Intervals

So far we have shown how to build an interval algebra from a point algebra. It will be useful if we can work backwards too: given a representation of an interval algebra we would like to retrieve the points from the representation. One way of doing this is to identify intervals of the form (x, x) with the point x, however, the definition of interval insists that the pair must be related by the atomic relation r whereas (x, x) is related by the identity relation. It may indeed be possible to get round this problem by extending the definition of interval, but here we recover points from intervals in a different way.

THEOREM 28 *Let Y be a representation of the interval algebra $\mathcal{A}^2_r$. There is a representation X of $\mathcal{A}$ such that $Y \cong X^2_r$. Furthermore X is unique up to isomorphism.*

PROOF:

The relation

$$same - start = \begin{bmatrix} Id_{\mathcal{A}} & r \\ r^{\smile} & r^{\smile} * r \end{bmatrix}$$

defines an equivalence relation on Y. Let the set of equivalence classes be D, these will form the points for the representation (X, D) of $\mathcal{A}$. Next we represent the atoms of $\mathcal{A}$. If a is an atom and $p, q \in D$, let

$$(p, q) \in X(a) \Leftrightarrow \forall i \in p, j \in q \ \ [i \begin{bmatrix} a & a * r \\ r^{\smile} * a & r^{\smile} * a * r \end{bmatrix} j]$$

It is a routine exercise to check that this defines a representation, (X, D)[5], of $\mathcal{A}$. To get the isomorphism take any interval $i \in Y$. Let Ξ map i to the pair $(p, q) \in X^2_r$ where p is the equivalence class of intervals to which i belongs and q is the set $\{j \in Y : i \text{ meets } j\}$ and 'meets' is the interval relation $\begin{bmatrix} r & r * r \\ Id_{\mathcal{A}} & r \end{bmatrix}$.

---

[5] Recall that we can drop the D and simply call the representation X.

Check that q is a point in X (i.e. an equivalence class under 'same-start') and that the pair (p, q) belongs to $X(r)$, in other words that (p, q) are related by r. It is not hard to show that Ξ respects all the operations.

Ξ is injective, for if $\Xi(i) = \Xi(j) = (p, q)$ then $i, j \in p$ so i and j are related by $\begin{bmatrix} Id_{\mathcal{A}} & r \\ r^{\smile} & r^{\smile} * r \end{bmatrix}$. By considering the definition of q we see that for all intervals k, i meets k if and only if j meets k. Hence i and j are also related by

$$\begin{bmatrix} r & r * r \\ Id_{\mathcal{A}} & r \end{bmatrix} * \begin{bmatrix} r & r * r \\ Id_{\mathcal{A}} & r \end{bmatrix}^{\smile} \leq \begin{bmatrix} r * r^{\smile} & r \\ r^{\smile} & Id_{\mathcal{A}} \end{bmatrix}.$$

Intersecting the two relations between i and j we deduce that i relates to j via

$$\begin{bmatrix} Id_{\mathcal{A}} & r \\ r^{\smile} & Id_{\mathcal{A}} \end{bmatrix}$$

which is the identity relation on intervals. Hence $i = j$.

To show that Ξ is surjective, let (p, q) be any pair of points from X related by r. Let i and j be any representative elements of p and q (respectively). i relates to j by $\begin{bmatrix} r & r * r \\ r^{\smile} * r & r^{\smile} * r * r \end{bmatrix}$ but since

$$\begin{bmatrix} r & r * r \\ r^{\smile} * r & r^{\smile} * r * r \end{bmatrix} = \begin{bmatrix} Id_{\mathcal{A}} & r \\ r^{\smile} & r^{\smile} * r \end{bmatrix} * \begin{bmatrix} r & r * r \\ Id_{\mathcal{A}} & r \end{bmatrix}$$

there must be an interval k such that the i and k are related by $\begin{bmatrix} Id_{\mathcal{A}} & r \\ r^{\smile} & r^{\smile} * r \end{bmatrix}$ and k and j are related by $\begin{bmatrix} r & r * r \\ Id_{\mathcal{A}} & r \end{bmatrix}$. This interval k satisfies $\Xi(k) = (p, q)$.

We have shown that the interval representation Y is isomorphic to X^2_r where X is the representation of $\mathcal{A}$ constructed above. To show that X is unique up to isomorphism, suppose $\Upsilon : Z^2_r \cong Y$ where Z is a representation of $\mathcal{A}$. For all $z \in Z$ map z to

$$\{i \in Y : \exists w \in Z[(z, w) \in Z^2_r \wedge \Upsilon(z, w) = i]\}.$$

Check that z maps to an element of X and that the map is an isomorphism.
□

COROLLARY 29 *If $\mathcal{A}$ is categorical in some infinite cardinality κ then $\mathcal{A}^2_r$ is κ-categorical.*

PROOF:

Follows from theorem 28. □

Note The converse does not always hold: the interval algebra $\mathcal{A}^2_r$ can be κ-categorical but the point algebra $\mathcal{A}$ may not be. It is true that $\mathcal{A}$ can have only one representation X of cardinality κ such that X^2_r is a representation of $\mathcal{A}^2_r$ but it may have other representations too (either not fully universal, inhomogeneous or not complete).

The next corollary was proved first in [Ladkin and Maddux, 1987a] but follows here from a more general result.

COROLLARY 30 *The Allen interval algebra $\mathcal{A}$ is ω-categorical.*

PROOF:

It has already been proved that $\mathcal{P}$ is ω-categorical (corollary 17). So corollary 29 gives the result.

□

3.3 Intervals with metrics

Since the metric point system $\mathcal{M}$ has a normal representation (corollary 20) the construction of theorem 25 produces a relation algebra of intervals with metrics. But this is a rather uninteresting algebra of intervals as an interval here is defined by a single, fixed atomic relation. That means that all intervals have to be of the same size, an over-restricted definition. An interval is more usually considered as any pair of points with the first one less than the second. In order to deal with these it is necessary to consider non-atomic networks. Recall that the regular representation of $\mathcal{M}$ is $\mathbb{Q}$.

3.3.1 Definition of $\mathcal{M}^2$

1. An *interval* is a pair of rationals (p, q) such that $p < q$.
2. An *elementary metric interval relation* R is a two by two matrix $\left[\begin{array}{cc} A & B \\ C & D \end{array}\right]$ where A, B, C and D are intervals of rationals such that the $\mathcal{M}$-network

   is transitively closed. $(p^-, p^+)R(q^-, q^+)$ asserts that $q^- - p^- \in A, q^+ - p^- \in B$ etc.

3. Such relations are composed according to the rule
$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} * \begin{bmatrix} E & F \\ G & H \end{bmatrix} = \begin{bmatrix} (A+E)\cap(B+G) & (A+F)\cap(B+H) \\ (C+E)\cap(D+G) & (C+F)\cap(D+H) \end{bmatrix}$$
in other words ordinary matrix multiplication with addition of intervals and intersection instead of multiplication and addition respectively.

4. The identity is $\begin{bmatrix} [0,0] & (0,\infty) \\ (-\infty,0) & [0,0] \end{bmatrix}$ (note that this is not atomic, so $\mathcal{M}^2$ is not integral) and the converse of a relation $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ is $\begin{bmatrix} -A & -C \\ -B & -D \end{bmatrix}$.

5. More general metric interval relations can be formed as disjuncts of elementary ones. The complement of an elementary relation will typically be a disjunction. Non-elementary matrices can be introduced e.g. $R \vee S$. The product
$$(R \vee S).(T \vee U) = RT \vee RU \vee ST \vee SU.$$
Thus the disjunct $i\{<,>\}j$ can be expressed as
$$i \begin{bmatrix} (-\infty,0) & (-\infty,0) \\ (-\infty,0) & (-\infty,0) \end{bmatrix} \vee \begin{bmatrix} (0,\infty) & (0,\infty) \\ (0,\infty) & (0,\infty) \end{bmatrix} j.$$
Note that this could not be expressed by a network in the point algebras $\mathcal{P}$ or $\mathcal{M}$.

6. The efficiency will be enhanced if a disjunction is removed by the rule: $M \vee N \Rightarrow N$ if $M \subseteq N$ (where $M \subseteq N$ means that each of the four entries of M is a subset of the corresponding entry of N). Also for matrices with three of the four entries equal
$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \vee \begin{bmatrix} A & B \\ C & D^* \end{bmatrix} \Rightarrow \begin{bmatrix} A & B \\ C & D \vee D^* \end{bmatrix}$$
provided $D \vee D^*$ is an interval.

As before, it is necessary to check that composition is associative and this is done by showing that matrix product is isomorphic to composition of relations. The critical section of the proof takes two intervals α and β related by the matrix product $R * S$. It is required to show that there exists a third interval γ such that α and γ are related by R and γ and β are related by S. But this follows from the fact that $\mathcal{M}$ has the extension property (theorem 21).

Expressive Power This system is capable of expressing all of Allen's interval relations e.g. 'overlaps' is written as $\begin{bmatrix} (0,\infty) & (0,\infty) \\ (-\infty,0) & (0,\infty) \end{bmatrix}$. A constraint on the duration of i can be expressed using this format as $i \begin{bmatrix} [0,0] & [d,e] \\ [-e,-d] & [0,0] \end{bmatrix} i$ where d and e are respectively lower and upper bounds on the duration of i. Thus, the qualitative expressive power of Allen's system is combined with the quantitative power of the metric system $\mathcal{M}$ of Dechter, Meiri and Pearl.

3.3.2 Complexity

It is possible to use any of the algorithms from the literature in conjunction with this language, but for the present suppose we use a fixpoint algorithm to calculate the transitive closure of a network. We have a network N and a new assertion $R(i,j)$ about the relation between the nodes i and j in N. Obtain N^* from N by letting $N^*(i,j) := N(i,j) \wedge R(i,j)$ and leave the other edges alone. Then find the transitive closure of N^* by

$$\text{Let } M := N^*$$

$$\text{Repeat } M := M^2$$
$$\text{Until } M = M^2.$$

The network M^2 is derived from M by

$$(\forall i,j,k \in M)\ M^2(i,k) := M(i,k) \wedge (M(i,j) * M(j,k)).$$

So every triangle (i,j,k) must be considered and the order they are taken will depend on the detail of the code. The Allen propagation algorithm is based on this specification. Transitive closure can be calculated in cubic time[6] certainly for any finite relation algebra and for infinite ones too provided they are finitely generated.

$\mathcal{M}^2$ is a highly expressive language and the worst case complexity of checking the consistency of a network will be at least as bad as its two sub-languages $\mathcal{M}$ and $\mathcal{I}$, i.e. it is NP-hard. In fact, a non-deterministic Turing machine could solve the problem in polynomial time since the non-disjunctive case can be solved in cubic time (below) so consistency checking for $\mathcal{M}^2$ is NP-complete.

But if we restrict to certain fragments of the full language we obtain the following results.

- A network with only elementary metric interval relations on the arcs (no disjuncts) can be checked in cubic time. This follows from the proof in [Dechter *et al.*, 1991] that computing the transitive closure of an $\mathcal{M}$-network with only one interval on each arc (in their terminology an STP), can be done in cubic time, and computes the minimal network.
- If all the relations on the arcs of the network are *pure Allen relations* i.e. equivalent to a union of some of the thirteen primitive interval relations, then the matrix product (which is calculated in constant time) will produce the same result as the Allen transitivity table. Therefore the same complexity and completeness results will hold i.e. consistency checking is NP-complete, but the Allen propagation algorithm provides a useful approximation in cubic time.

---

[6] See [Ladkin and Maddux, 1993] for a complexity analysis of the problem of calculating transitive closure of $\mathcal{A}$-networks.

- For general metric constraints with disjunctions, the problem is NP-complete. Dechter, Meiri and Pearl left open the problem of whether the fixpoint algorithm must terminate at all. If the metric values are commensurate (the ratios are rational) then without loss it may be assumed that all the metric constraints have integer bounds. In this case the fixpoint algorithm will certainly terminate and if the number of integers lying in any constraint has a fixed bound then the algorithm will terminate in cubic time. The argument is exactly the same when analysing $\mathcal{M}^2$ except the bound must apply to the number of atomic matrices with integer entries, within the constraint. In an appendix we answer an open question from [Dechter *et al.*, 1991] and show that the fixpoint always terminates, even when the intervals are allowed irrational endpoints.

  The remaining problem is that calculating the fixpoint (transitive closure) is not a *complete* deductive mechanism. It is easy to devise inconsistent but transitively closed networks in $\mathcal{M}$ (see diagram (c) page 73) and hence in $\mathcal{M}^2$. However, computing the transitive closure may give a useful first approximation for a consistency checker which then proceeds by brute force to test each choice of disjunctions for consistency (the constraint technology will clearly be useful here).

Examples

1. From $i\{\text{during}\}j$ and $j\{\text{overlaps}\}k$ we should deduce that $i\{<, \text{meets, overlaps, starts, during}\}k$.
   $i\begin{bmatrix}(-\infty,0) & (0,\infty)\\ (-\infty,0) & ((0,\infty)\end{bmatrix} j$ and $j\begin{bmatrix}(0,\infty) & (0,\infty)\\ (-\infty,0) & (0,\infty)\end{bmatrix} k$ gives
   $i\begin{bmatrix}(-\infty,0) & (0,\infty)\\ (-\infty,0) & ((0,\infty)\end{bmatrix}\begin{bmatrix}(0,\infty) & (0,\infty)\\ (-\infty,0) & (0,\infty)\end{bmatrix} k$ or
   $i\begin{bmatrix}(-\infty,\infty)\cap(-\infty,\infty) & (-\infty,\infty)\cap(0,\infty)\\ (-\infty,\infty)\cap(-\infty,\infty) & (-\infty,\infty)\cap(0,\infty)\end{bmatrix} k$ i.e.
   $i\begin{bmatrix}(-\infty,\infty) & (0,\infty)\\ (-\infty,\infty) & (0,\infty)\end{bmatrix} k$ which says that both endpoints of i must lie before the end of k, as required.

2. Let j start at least 5 second after i finishes and let j finish less than 10 seconds after i starts.

It should be possible to deduce that the duration of i is less than 5 seconds. Well, $i \begin{bmatrix} (0,\infty) & (0,10) \\ (5,\infty) & (5,\infty) \end{bmatrix} j$ and $j \begin{bmatrix} (-\infty,0) & (-\infty,-5) \\ (-10,0) & (-\infty,-5) \end{bmatrix} i$ so $i \begin{bmatrix} (-10,10) & (-\infty,5) \\ (-5,\infty) & (-\infty,\infty) \end{bmatrix} i$. Intersecting this with the initial constraint on i, gives $i \begin{bmatrix} [0,0] & (0,5) \\ (-5,0) & (0,0) \end{bmatrix} i$ as required.

3.3.3 Comparisons

A number of other attempts have been made to combine qualitative and quantitative reasoning [Kautz and Ladkin, 1991, Ladkin, 1989, Ligozat, 1990, Meiri, 1991]. The language $\mathcal{M}^2$ of this paper has two main advantages. Firstly, it uses the same uniform representation for all relations. There is no need to refer to a special table when dealing with an interval constraint and a separate table for metrics. All constraints are represented as matrices and compositions are calculated by matrix multiplication. By contrast, [Kautz and Ladkin, 1991], [Ladkin, 1989] and [Meiri, 1991] are all essentially hybrid systems which handle metric and interval information separately and translate from one to the other.

The other advantage of $\mathcal{M}^2$ is its expressive power. When disjunctive relations are allowed it is possible to express constraints which are neither point-based metric nor qualitative interval relations. For example to assert that there is a gap of at least 5 seconds between interval i and j, we use the disjunction

$$\begin{bmatrix} (-\infty,-5) & (-\infty,-5] \\ (-\infty,-5) & (-\infty,-5) \end{bmatrix} \vee \begin{bmatrix} (5,\infty) & (5,\infty) \\ [5,\infty) & (5,\infty) \end{bmatrix}.$$

Note that this could not be represented directly in any of the competing systems. It would be necessary to construct additional intervals and put constraints on these. This expressive power is achieved without additional complexity cost (the complexity of checking the consistency of a network in a sub-language of $\mathcal{M}^2$ is the same as that in competing systems).

More tentatively, there is one further advantage. When dealing with the disjunctive case, the simple algorithm for combining matching disjuncts (see page 64, Definition of $\mathcal{M}^2$, item 6) is very straightforward and will improve the efficiency considerably. Disjuncts like $before, meets$ in Allen's language translate to

$$\begin{bmatrix} (0,\infty) & (0,\infty) \\ (0,\infty) & (0,\infty) \end{bmatrix} \vee \begin{bmatrix} (0,\infty) & (0,\infty) \\ [0,0] & (0,\infty) \end{bmatrix}.$$

which gets rewritten as $\begin{bmatrix} (0,\infty) & (0,\infty) \\ [0,\infty) & (0,\infty) \end{bmatrix}$ thus eliminating a disjunct which could improve the efficiency. Theoretical results about average case performance are hard to provide in this area so this is most likely to be judged, eventually, by empirical results.

3.4 Conclusion

An interval relation algebra can be constructed from a point relation algebra provided it is normal. This allows us to construct a metric interval algebra from the metric system $\mathcal{M}$. This representation permits the expressing of Allen type disjuncts like $i\{<,>\}j$. It is thus more expressive than other systems that allow quantitative, metric information. As with these systems [Dechter *et al.*, 1991] the propagation algorithm will be complete and of cubic complexity if there are no disjuncts but its performance in general is intractable. We leave this section on the building of interval algebras from point algebras with an open problem.

PROBLEM 12 (When does transitive closure guarantee consistency?) *A transitively closed (not necessarily atomic) network in the point algebra $\mathcal{P}$ may not be complete though it nearly always is (see [Villain* et al.*, 1989]). Nevertheless it will always be consistent. For other algebras, like $\mathcal{I}$, this does not hold (see example (d) in section 4.3 for a case of a transitively closed network that is unsatisfiable). It would be useful to identify conditions on a relation algebra, $\mathcal{A}$, which ensure that a transitively closed $\mathcal{A}$-network is consistent.*

Negatively, we can already see that if $\mathcal{A}$ has a subalgebra $\mathcal{B}$ (i.e. a subset of the relations in $\mathcal{A}$ including $Id_{\mathcal{A}}, 0, 1$ and closed under $*$ and $\smile$) such that an atomic $\mathcal{B}$-network embeds in none of the representations of $\mathcal{A}$ then there are inconsistent transitively closed $\mathcal{A}$-networks. Simply take an atomic $\mathcal{B}$-network (say N) which does not appear in any representation of $\mathcal{A}$ and then consider N as an $\mathcal{A}$-network. This is exactly what happens with the Allen interval algebra and the containment algebra (see page 28). This algebra has a fully universal representation - the atomless Boolean algebra. However the containment algebra is not universal over any representation of $\mathcal{I}$.

So the problem, which remains open, is this: if every subalgebra of a relation algebra $\mathcal{A}$ is fully universal over a representation of $\mathcal{A}$, does it follow that every transitively closed $\mathcal{A}$-network is consistent? This result would clearly be of enormous benefit to computer science as for finite $\mathcal{A}$ the transitive closure can be calculated in cubic time (and for many infinite relation algebras, e.g. $\mathcal{M}$, too). Thus consistency checking for $\mathcal{A}$-networks becomes tractable if the condition holds. A counter example would also be useful as it might point the way to an alternative characterisation of the desired result.

Chapter 4

Dependence and Independence in the Allen Interval Algebra

4.1 Introduction

The Allen interval algebra has been used extensively for temporal reasoning and in planning [Allen, 1983, Allen and Koomen, 1983]. However, a number of problems have arisen. One is the incompleteness of the propagation algorithm, another is the problem of collapsing intervals in the Allen and Koomen planner - what do you do if each clause has a potential causal explanation but it is not consistent to collapse them all simultaneously? An attempt was made to deal with the intractability of the propagation algorithm [Koomen, 1989] by the use of reference intervals. A more general criterion for decomposing a network is presented here. As far as we know, no attempt has been made to deal with the latter problem (interval collapsing), and this is a serious problem for planning. In a typical planning problem, at the point when collapsing is tried there will be a large number of intervals which are to be collapsed. But, if it is inconsistent to collapse them all simultaneously, then at least one of the causal gaps cannot be removed by collapsing - an action must be introduced, but it is not known where. This could be solved by trying each one in turn - introduce a new action for a causal gap and continue with the plan algorithm. This involves back-tracking with an enormous branching factor and completely abandons the notion of least commitment. The groundwork to deal with this problem is laid here by defining dependence and independence. This theory is developed in section 5.7 to solve the collapsing problem. In the sections which follow networks are no longer assumed to be atomic, there may be non-atomic relations labelling an edge.

4.2 Basic Definitions and Results

- The Allen Interval Algebra $\mathcal{I}$ has a countable, normal representation which is unique up to isomorphism, namely ordered pairs of rational numbers (written $Int(\mathbb{Q})$).

- An *interval network* N is a not necessarily atomic network (as defined in section 1.3.3) over $\mathcal{I}$. N is equivalent to a first order formula $\exists_{(n\in N)}\, n[\bigwedge_{n,m\in N} N(n,m)]$ where $N(n,m)$ is the relation on the arc from n to m in N.
- Let $M \subseteq N$. Define the *restricted network* $N|_M$ by $N|_M(m_1, m_2) = N(m_1, m_2)$ (for all $m_1, m_2 \in M$).
- Let X be any representation of $\mathcal{I}$. A *model* of N is mapping I from N to X such that
  $$\overset{\leftarrow}{X}(I(n), I(m)) \leq N(n,m)$$
  where $\overset{\leftarrow}{X}(I(n), I(m))$ is the atomic relation which holds between $I(n)$ and $I(m)$ in X.

  Example The network

  has a model in $Int(\mathbb{Q})$ given by the map
  $$\begin{aligned} I : l &\rightarrow [1,5] \\ m &\rightarrow [0,3] \\ n &\rightarrow [1,9] \end{aligned}$$
- N has a model in *some* representation if and only if it has a model in $Int(\mathbb{Q})$ so we need only consider models in this representation. This result follows from the universality of $Int(\mathbb{Q})$ as a representation of $\mathcal{I}$.
- N is *complete* if for each arc (n,m) and each atomic relation $r \leq N(n,m)$ there is a model of N which realises r on (n,m).

Example The network below is not complete.

The relation ‘meets’ on side (a, b) cannot be realised in any model.

- N is *transitively closed* if

$$(\forall l, m, n \in N)\ N(l, m) * N(m, n) \geq N(l, n).$$

- N is *consistent* if it has a model.
- N' is a *tightening* of N if it has the same set of nodes and for all $n, m \in N, N'(n, m) \leq N(n, m)$.
- A *labelling* of N is a tightening N' such that $N'(n, m)$ is always an atom.
- A *point network* is just like an interval network, except we allow only three atomic relations between points: $<, >$ and $=$. The composition table is defined on these three in the obvious way. Here we always take the representation to be $\mathbb{Q}$.

Note Since $\mathcal{I}$ has a fully universal representation (namely $Int(\mathbb{Q})$) it follows that a network N is consistent if and only if it has a transitively closed labelling.

4.3 The Idea of Dependency

Even if a network is complete there are hidden dependencies.

The first two networks above are complete (every relation has a model) but not every combination of relations is allowed. In diagram (a) x and y cannot both be equality. In (b) you cannot have equality on w and $<$ on u (or v). The combination $(u : ‘ ='), (v : ‘ ='), (w : ‘ <')$ is not allowed either. (c) is a metric $\mathcal{M}$-network. Considering the triangle x, y, z we see that it is transitively closed but only two combinations are allowed: $x : 0, y : 1, z : -1$ or $x : 1, y : 0, z : 1$. The triangle u, v, z is similar. On y, v, w only $y : 0, v : 1, w : -1$ and $y : 1, v : 0, w : 1$ are permitted. (x, u, w) is similar. The network is transitively closed (each triangle is complete) but the network as a whole is inconsistent. There is no labelling of the network which is consistent with each triangle. (d) is the example taken from [Allen, 1984], similar to (c) in that each triangle is complete, but because of the restriction on allowed combinations, there is no consistent labelling of the whole.

An analysis of these networks requires a way of defining the inter-dependence of the choices of relations. A formal definition of independence will be used to see if a network can be decomposed into smaller subnetworks.

A further application of the concepts of dependence and independence will be given in part III. This is the so-called ‘collapsing problem’ and using independence we demonstrate an alternative to dependency directed backtracking to solve this problem.

4.4 Independence and Dependence

Let $\mathcal{A}$ be any atomic relation algebra such that $Net(\mathcal{A})$ has the amalgamation and joint embedding properties. Let N be an $\mathcal{A}$-network. Let $E, F \subseteq N \times N$ be sets of edges.

Definition E and F are *independent* in N, written $E \#_N F$, if for every pair of labellings L_E and L_F of E and F (respectively) consistent with N (i.e. they each extend to consistent labellings of N) there is a consistent labelling of N which extends both. In other words if L_E and L_F are each consistent with N then they are jointly consistent with N.

Examples

1.

$\{(i,j)\} \#\{(j,k)\}$ since the two labellings $i = j$ and $j = k$ are individually consistent with N but they do not jointly extend to N.

2. Reference intervals.

   We will show later that if i and j are disjoint reference intervals (see [Koomen, 1989]) then the networks of intervals contained respectively in i and j are independent of each other.

LEMMA 31 *Let $U \subseteq E$ and $V \subseteq F$. If $E \# F$ then $U \# V$.*

PROOF:

Assume $E \# F$. Let L_U, L_V be labellings of U, V consistent with N. That is, each one extends to N. Let L_E be the restriction to E of an extension of L_U (so L_E is an extension of L_U) and let L_F be a similar extension of L_V. Since $E \# F$, L_E and L_F have a joint extension L - a consistent labelling of N and an extension of L_U and L_V.

$\square$

Definition Let $E \subseteq N \times N$. E is *free* in N if *any* labelling extends consistently to N.

Note N is complete if every edge $\{(i,j)\}$ is free in N.

Definition A *dependency* is a minimal inconsistent labelling. Minimal means that if the labelling is defined on a set of edges E, then any proper subset of E yields a restricted labelling which can consistently extend to N.

Examples

1. In the diagram on page 74, the labelling which assigns equality to (i,j) and (j,k) is a dependency.

2\.

The following are dependencies: $(a \rightarrow =; b \rightarrow =; c \rightarrow =), (a \rightarrow =; v \rightarrow =)$ and $(a \rightarrow = ; b \rightarrow =; u \rightarrow <)$.

We now give the main theorem relating the concepts of dependency and independence.

THEOREM 32 *Let N be a network and $E, F \subseteq N \times N$.*

1. *$E \#_N F$ if and only if there is no dependency D contained in $E \cup F$ such that D uses at least one edge from E and at least one from F.*
2. *E is free in N if and only if there is no dependency D using edges only from E.*
3. *Let L be a labelling of some of the edges in N. L consistently extends to a labelling of N if and only if L is not an extension of some dependency.*

PROOF:

1\. $(\Rightarrow)$ Suppose there is a dependency D contained in $E \cup F$ and intersecting both. We must show that $E \not\#_N F$. D is minimal, so restricting D to the edges of E gives a consistent labelling i.e. one that consistently extends to N. Call this L_E and similarly define L_F. L_E and L_F are consistent labellings of E and F but with no joint extension (if there was then D would not be inconsistent).

($\Leftarrow$) Suppose $E \ \#_N F$ and we will show that there is a dependency D contained in $E \cup F$ and intersecting both. There are labellings L_E and L_F of E and F which are consistent with N but which do not jointly extend. So the labelling $L_E \cup L_F$ of $E \cup F$ is inconsistent. Delete edges from this labelling till you get a minimal inconsistent labelling (a dependency). This is clearly contained in $E \cup F$ but cannot be wholly contained in E (or F) since L_E (and L_F) are consistent.

2. Similar.
3. Straight-forward.

□

THEOREM 33 *Let $\mathcal{A}$ be a relation algebra with a fully universal, complete representation, and let N be a network. A labelling L of N is consistent iff it does not contain a dependency of size ≤ 3 which is contained in some triangle of the network.*

PROOF:

By full universality, if L is transitively closed then it is consistent. So, if L is inconsistent then it is not transitively closed which means that there are three elements i, j, k and $L_{\{i,j,k\}}$ is inconsistent. Therefore $L_{\{i,j,k\}}$ contains a dependency.

□

4.5 A Generalisation of Reference Intervals

We wish to decompose N into two (or more) independent parts E and F. That is $N \times N = E \cup F$ and $E \#_N F$. By theorem 32 this is the case iff there is no dependency D which intersects both E and F.

Note In determining when two sets of edges E and F form such a decomposition, theorem 33 informs us that we need only check that there are no dependencies involving E and F contained in some *triangle*.

Let i and j be disjoint intervals such that every interval in the network is contained (eq, d, st or ends) in one or other of i and j. i and j are called *reference intervals* [Koomen, 1989]. Every interval in the network is contained in exactly one of i and j.

- Suppose first that the relation between i and j is $\{<\}$. If $a \subseteq i, b \subseteq j$, then the relation between a and b can only be $\{<\}$. Claim: if L, M are respectively the network of intervals contained in i and j then $L\#M$. We prove a slightly more general result: if the network N splits into two sets of nodes A, B such that for any $a \in A, b \in B$ the relation between a and b is atomic then the networks $N|_{A\times A}$ and $N|_{B\times B}$ are independent. Note first that any dependency of N cannot involve an edge from A to B as these have only one atom on them, so an inconsistent labelling involving such an edge would remain inconsistent after the edge was deleted. Suppose for contradiction that $N|_{A\times A}$ and $N|_{B\times B}$ are not independent. By theorem 32 there is a dependency D involving only edges in $A \times A$ and $B \times B$. By theorem 33 there must be a dependency contained in some triangle (a, b, c). But it is not possible for such a dependency to involve no edges between A and B and yet intersect both A and B.

- If the relation between i and j is $\{<, >\}$ it can be shown that we get a decomposition into three mutually independent sets of edges: those relating two intervals contained in i; those relating two intervals contained in j and those relating an interval contained in i to an interval contained in j.

Note The previous theorems take a global view of the network. A dependency is inconsistent with the *whole* network. For intervals determining consistency is intractable. What is needed is a *local* criterion which gives the global result. Theorems 21 and 24 are such results: in order to tell whether a sub-network of a transitively closed $\mathcal{M}$-network is consistent it is only necessary to examine the sub-network; testing the consistency of a $\mathcal{P}$-network can be done by checking the completeness of each triangle i.e. by ensuring that the network is transitively closed.

The next theorem is a local *approximation* for finding independent subnetworks.

THEOREM 34 *Let $N \times N = E \cup F$. E and F are independent in N if it is not possible to find a partial labelling L defined on two or three edges of some triangle $\tau = \{i, j, k\}$ such that*

1. *L is inconsistent* with τ *(no need to consider the rest of the network),*

2. *L is minimal such and*

3. *L is defined on at least one edge of E and one of F.*

PROOF:

Suppose $E \#F$. There are consistent labellings L_E and L_F of E and F, but $L = L_E \cup L_F$ is an inconsistent labelling of N. Since the relation algebra has a fully universal representation, L is inconsistent on some triangle τ. Find a minimal set of edges from τ which is inconsistent *in* τ. Clearly these edges must involve both E and F since L_E and L_F are consistent with N, so certainly consistent with τ.

□

Note To decompose a network into independent sets of edges we could define an equivalence relation $\sim$ by taking the transitive closure of the relation defined by $i \sim j$ if i, j belong to some minimal inconsistent labelling of a triangle τ as in the previous theorem.

We can now partition N into independent $\sim$-classes. Because the theorem has only a one-way implication, we may fail to get the *finest* decomposition, but we are guaranteed that the equivalence classes are independent.

The concepts of independence and dependence will be explored further in part III where they are applied in artificial planning.

Part III

Applications to Planning

Chapter 5

The Allen and Koomen Planner

The Allen and Koomen planner [Allen and Koomen, 1983] is an elegant device based on the Allen interval algebra. The following description is based on the original but does contain some alterations and generalisations.

A *planning problem* is defined to be a *planning environment* (PE), a *goal state* (G) and some temporal constraints defined by an *interval network* (IN). The planning environment and goal state are sets of temporal assertions i.e. they consist of formulas of the form $p@i$ where p is a proposition and i is an interval name. Here, following [Pelavin, 1988] we generalise Allen and Koomen's concepts and do not require that the initial state comprises properties true at the initial time nor that the goal state has only properties true at the final time. There may be various temporal constraints on these intervals defined in the interval network. For the present we assume that the interval network is complete i.e. each relation which occurs on an edge is realised in some model of the network.

There is also a stock of available *actions* each of which consists of an action assertion $ac@x$, effect assertions (or post-conditions) and executability conditions (or pre-conditions) together with the temporal constraints between all of these.

Example An action to move block A onto B could be defined like this:-

action	$Move(A, B)@x$
preconditions:	$Clear(A)@y$ $Clear(B)@z$
post-conditions:	$On(B, A)@v$ $Clear(A)@w$
temporal constraints:	$x\{\text{during}\}y$, $x\{\text{equals, ends, ended-by}\}z$, $x\{\text{meets}\}v$, $x\{\text{meets}\}w$

Or in a diagram

Because an action may be used any number of times in a plan, the intervals in the definition of the actions should be considered as variables which get instantiated by Skolem constants if and when the action is introduced into the plan.

5.1 The Planning Algorithm

Assertions in the goal set are removed until there are none left. The resulting planning environment and the interval network constitute a *solution plan.* Goal assertions are removed by one of two methods: either an action is *introduced* which has the required assertion as one of its post-conditions or the goal assertion is *collapsed* with an assertion in the planning environment.

Introducing an action If $p@i$ is a goal assertion and the action Ac has $p@x$ as one of its post-conditions then the action may be introduced to remove the goal assertion. This is done by setting the variable x equal to the interval i and instantiating all the other interval variables in the action definition with new Skolem names. The temporal constraints between them are added and propagated through the interval network of the plan. The action assertion itself $Ac@j$ and all its post-conditions are added to the planning environment and all the pre-conditions are added to the goal set.

Collapsing If $p@i$ is in the goal set and $p@j$ is in the planning environment and the temporal constraint between i and j includes the equality relation, then we may remove $p@i$ from the goal set by collapsing i and j, that is, by asserting that $i = j$.

Notice that a goal assertion may be removed in more than one way: it may be that there is more than one interval with which it can collapse and there may be more than one action which has the desired post-condition. To clarify and simplify we introduce a definition.

Definition If $p@i$ is in the goal set and $p@j$ is in the planning environment and the temporal constraint between i and j includes the equality relation then we say that $p@j$ is a *potential causal explanation* (PCE) for $p@i$. A goal assertion with a PCE may be removed by collapsing, but a goal assertion without a PCE can only be removed by the introduction of an action.

So now we can state the planning algorithm.

1. For each goal assertion without a PCE introduce an action to remove it. If we simplify the problem by assuming that there is only one action which can bring about a given effect, then this step can be taken without the possibility of having to back-track later on.

2. When every goal assertion has a PCE attempt to collapse each goal assertion with its PCE. If this produces a consistent interval network then terminate successfully.

5.2 The Collapsing Problem

The problem, which was not clear in Allen and Koomen's original, is now staring us in the face: what do we do if step 2 above fails? What do we do if each goal assertion *individually* can collapse with its PCE but collapsing them *all at once* is inconsistent? A method of tackling this will be given later on.

For the moment an exhaustive search, though wasteful, is guaranteed to find a solution plan if one exists[1]. If it is inconsistent to collapse all the goal assertions with their PCE then at least one of them must be removed by introducing an action. So introduce an action for one of the goal assertions and then go back to the beginning of the planning algorithm. It may be necessary to back-track and choose another goal assertion so this search must be conducted breadth first. Clearly we are defining an intractable procedure, though in many typical planning problems step 2 of the planning algorithm may succeed immediately.

Example Let the initial planning problem be
(planning environment = $(\{p@i_0, q@i_1\}$, goal = $\{p@i_1, q@i_2\}$,
interval network = $\{i_0 \leq i_1 \leq i_2, i_0 < i_2\})$.

---

[1] A breadth first search will eventually find a correct plan if one exists, but this gives only semi-decidability. To be able to deal, in finite time, with an insoluble planning problem it would be necessary to include some kind of loop-checker, see [Hirsch, 1992].

Each goal has a potential causal explanation and each (individually) can be collapsed, but they cannot be collapsed together.

5.3 Causality

From "A causes B" we can always conclude $\Box(A \rightarrow B)$ where A and B are temporally qualified assertions and where $\Box\phi$ means that ϕ is true in any model. The converse is of course false. For example from $\Box(Action \rightarrow Precondition)$ we cannot conclude that the action causes the pre-condition, even though we know that in any model where the action occurs, the pre-condition must hold too. We cannot express notions of causality simply by making temporal assertions, something more is needed. Indeed, following Pelavin [Pelavin, 1988], the notions of pre-condition and post-condition convey a causal connection. We know in any model where the action occurs that both the pre- and post-conditions will also hold, but we know something more: the action *causes* the post-conditions but it does not cause the pre-conditions. So introducing an action can remove a post-condition from the goal set, as the action becomes the cause of that effect. It cannot remove a pre-condition, indeed the pre-conditions must be added to the causal gap. The action does not cause the pre-condition, it is *allowed* by it.

Allen and Koomen understood that a pre-condition and a post-condition cannot be distinguished by their temporal relation to an action, but drew no distinction between them in the action descriptors. Throughout this paper, however, the two notions have been separated explicitly in the action descriptors. Note that this generalises the traditional view of pre- and post-conditions as properties true immediately before and after an action [Fikes and Nilsson, 1971]. Here, in principle, the pre- and post-conditions can have any temporal relation to the action. However, the principle of causality requires that a pre-condition of an action should not refer to times strictly later than the action, and post-conditions should not refer to the past.[2]

---

[2] Puzzle: consider the action $Move(A, B)@i$. Suppose in the specification of this action we have $Clear(A)@j(i\{\text{during}\}j)$. Should this be a pre-condition or a post-condition? Pre-conditions should not refer to the future and post-conditions should not refer to the past.

Solution: a precondition for $Move(A, B)@i$ is that A is clear before it starts and while it continues. The pre-condition doesn't care what happens afterwards. So the precondition is $Clear(A)@j(i\{\text{ends, during}\}j)$.

5.4 Multi-agents

In the Pelavin set-up multi-agents are allowed, but the specification must include non-interference conditions between the actions. The Allen and Koomen framework allows you to specify some of these if it is known when an action is in progress. This can be expressed without adding anything new to the language, by the use of a special post-condition. Along with all the usual pre- and post-conditions for an action instance $Ac@i$, we have one extra post-condition: $Occurring(Ac)@i$. This tells us that at any sub-interval of i, Ac is in progress. Now we can specify various non-interference conditions like this: If Ac_1 can only occur when Ac_2 is in progress, then a precondition for $Ac_1@i$ is $(Occurring(Ac_2)@j$ where $i\{=, \text{during, finishes, starts}\}j$.

If Ac_1 and Ac_2 cannot occur together we could add a domain constraint which says given the assertions $Occurring(Ac_1)@i_1$ and $Occurring(Ac_2)@i_2$ we can deduce that i_1 and i_2 do not overlap (the relations between them are included in $\{<, >, \text{meets, met-by}\}$). This is just the same as the constraint given by Allen and Koomen that $Clear(B)$ and $On(A, B)$ cannot overlap.

More complex interactions can also be specified, but always a conjunction of binary non-interference conditions. Ternary or higher order non-interference conditions cannot be expressed so, for example, the problem of cooking three pans on two burners cannot be described in this set-up.

5.5 Forwards and Backwards

The Allen and Koomen planner is a backward planner. It works on the goal clauses, introduces actions to justify them and then works on the pre-conditions to the new actions. This continues till the pre-conditions to the actions can be collapsed with assertions in the planning environment. A simple modification can make the planner operate forwards and backwards simultaneously.

5.5.1 Reversibility

Given a planning problem (PE, G, IN) which extends to a solution plan $\Pi = (PE^*, \emptyset, IN^*)$, it is easy to find a solution $\Pi^\smile$ to the *reverse problem* $(G, PE, IN^\smile)$ where $IN^\smile$ is the same as IN except every relation is replaced by its converse (meets$^\smile$ = met by; before$^\smile$ = after; during$^\smile$ = during etc.). The solution $\Pi^\smile$ is like Π but the network is replaced by its

---

Similarly the post-condition says that A is clear immediately after the action. It doesn't care what happens before that, so we have $Clear(A)@k(i\{\text{meets, overlaps, starts, during}\}k)$.

Since we can deduce not $(j\{<, >\}k)$ and given that all intervals are maximal, we know $j = k$. Now we can eliminate disjuncts from the precondition and the post-condition to get $Clear(A)@j(i\{\text{during}\}j)$ as *both* the pre-condition and the post-condition.

In view of this, such an action should not be introduced to remove $Clear(A)@j$ from the goal set, even though it is an effect of the action. The action simultaneously removes it from the causal gap and puts it back again as a pre-condition.

converse (as above) and every assertion which refers to an action is replaced by its inverse. So

$$Move(a, from, to)^{\smile}@i = Move(a, to, from)@i$$

Other assertions, like $Clear(a)@j$ are left alone (self-inverse). Putting it more informally, if you see a film which starts in an initial state and progresses to a goal state, you can run the film backwards and move from the goal to the initial state. Now it is well known, in planning, that some problems are best tackled working forwards and some working backwards (it depends on the so-called 'fan out' - in this case the ratio of the number of pre-conditions to post-conditions). It is also known that there is some advantage to be gained by doing both at once. (As a rough guide if $f(n)$ is the complexity of the problem working only in one direction, then working forwards and backwards simultaneously can reduce the complexity to $\sqrt{f(n)}$). With the Allen and Koomen system, it should take only a small adjustment to make it work forwards and backwards.

Let the initial planning problem be (PE_0, G_0, IN_0). Modify step 1 of the planning algorithm thus:

1. (a) (As before) For each assertion $p@i$ in the goal set having no potential causal explanation, introduce an action to remove it. An assertion of the form $Action@j$ should be added to the planning environment along with its effects (including $p@i$). The pre-conditions are added to the goal set. The associated temporal relations should be added and propagated through the interval network. Let the output be (PE_1, G_1, IN_1)

   (b) Consider the planning problem $(G_1, PE_1, IN_1^{\smile})$ and perform step 1a to this problem. If the result of this step is $(G_2, PE_2, IN_2^{\smile})$ then invert this to (PE_2, G_2, IN_2) before going back to step 1a. These two steps can be performed in any order, it might be possible to look at the planning problem and heuristically decide whether it is easier to work forwards or backwards.

2. Collapse (as before).

Example: the Sussman Anomaly Let the initial planning environment be $\{C$ on $A@i_{ca}, Cl(C)@i_c, Cl(B)@i_b\}$ the goal set $\{Cl(A)@i_a, A$ on $B@i_{ab}, B$ on $C@i_{bc}\}$ and the interval network is $\{i_0 < i_g; i_{ca}, i_c, i_b \supseteq i_0; i_a, i_{ab}, i_{bc} \supseteq i_g\}$ (see below) as a set of assertions it looks like:

Let the only action be $Move(x, y)$ whose pre-conditions and post-conditions are given diagrammatically below (working forwards and backwards we only need one action).

Picking A on $B@i_{ab}$ and B on $C@i_{bc}$ as goal clauses, step 1a transforms the problem into:

$Cl(A)$ is the only assertion without a potential causal explanation, but we have no "Clear" action. Turning the problem round (step 1b) we introduce an action for the goal clause "C on A" and when this gets inverted we have:

Now every assertion has a potential causal explanation (dotted lines) and indeed step 2 succeeds to give a final plan:

which is the usual solution: move C off A; move B onto C; move A onto B.

5.6 Domain Variables

In the example above we have only one action $Mv(X, Y)$. This does not fully describe what is happening as it does not say where X was before it was moved. Something has been cleared but we can't tell what from the pre-conditions or post-conditions of this action. This can pose problems as the next example illustrates

$AonC@i_{ac}$ $AonB@i_{ab}$;
$Cl(A)@i_a; Cl(B)@i_b$ $Cl(C)@i_c$.

Neither A on B nor $Cl(C)$ has a potential causal explanation. Suppose first we introduce an action $Mv(A, B)$ to achieve A on B. Diagrammatically we have:

Notice that $Cl(C)$ does not have a potential causal explanation, even though in fact the move action will make C clear. This happens because the move action does not say where A was before it was moved. In fact A was on C and the action should make C clear, but this is not detected in this set-up. There is a similar problem with unstacking. This problem is avoided if we make two assumptions:

1. $Mv(X, Y)$ takes X off the table, so a pre-condition is Xon Table. (For unstacking we have to assume that we unstack X to the table)

2. The table is always clear (infinite table).

These two avoid any ambiguity and guarantee a solution, even though we may sometimes take two moves where one would suffice (Unstack A to the table, then put A on B instead of putting A directly on B).

However, if we have a finite table the problem becomes more interesting. We cannot assume that A was on the table before it was moved, there may not be enough room for that. We must use the more explicit notation $Mv(X, From, To)$. But here again, we run into a problem: in order to achieve a goal A on B we have many actions that bring it about - $Mv(A, C, B)$; $Mv(A, D, B)$, etc. Of course we could try each possibility in turn and back-track when necessary, but there may be a better solution which preserves the least commitment principle. This proposed solution uses domain variables and constraints.

Let us fix the objects in the world to be, say, $\Delta = \{A, B, C, D, T_1, T_2, T_3\}$ the last three indicating the three available table positions. If we have a goal (A on B)@i to achieve we can introduce the action $Mv(A, x, B)$ where x is a new *domain variable*, $x \in \Delta$. Domain constraints may also tell us that $x = A$ or B are impossible so we attach to x the set $D_x = \{C, D, T_1, T_2, T_3\}$. At this stage we do not choose any one of these values for x. Further calculations, later in the planning process may eliminate some of these values. Most of the original algorithm remains unaltered but we make one redefinition.

Definition Suppose that relation between intervals i and j includes equality. A *potential causal explanation* (PCE) for an assertion $p(x)@i$ is either

1. an assertion $p(c)@j$ for some constant c where $c \in D_x$ or

2. an assertion $p(y)@j$ where D_x intersects D_y

A PCE for p(c)@i is either

3. $p(c)@j$ or

4. $p(x)@j$ where $c \in D_x$.

When we attempt to collapse these assertions, case 3 does not involve domain variables and needs no modification. In cases 1 and 4 we try $x = c$ and in case 2 we put x and y equal to some element in the intersection of D_x and D_y and as before we put $i = j$. In the worst case the domains for the variables will be large and we will have to try all the possibilities. But in other cases the domains will become restricted and may even produce a unique solution.

Example We want to demonstrate the method on a limited table. A table with only two locations is too simple as there are only two possible moves in any situation (left → right or right → left). So we have three locations:- T_1, T_2 and T_3. Unfortunately, if this is to restrict our moves we will need at least four blocks : A, B, C and D. So the example is rather complicated, anyway here it is.

or diagrammatically:

B on C and C on D are selected as having no potential causal explanation (PCE) so we get

Now the A on B in the goal no longer has a PCE as the $Cl(B)$ rules out equality between the two A on B's. So we introduce $Mv(A, z, B)$

At this stage we have $D_x = \{T_2, T_3, A, D\}$; $D_y = \{T_2, T_3, A, B\}$ and $D_z = \{T_2, T_3, C, D\}$. We could pick $Cl(B)$ as a goal clause without a PCE but this will introduce an action

$Mv(s, B, t)$ which has two variables and will therefore be more difficult. Instead we use the reversibility property and work left to right. Now D on C and A on B become goal clauses without PCEs.

We have $D_v = \{T_1, T_3, A, B\}$ and $D_w = \{T_1, T_2, C, D\}$. Every assertion now has a PCE so we attempt to collapse. Consider first the variable x. The PCE for the assertion B on x is B on T_3. So we put $x = T_3$. Similarly, the PCE for the assertion C on y is C on T_2 so we put $y = T_2$. The PCE for A on z is A on w so this time we check that D_w intersects D_z (which it does) so we reset D_w and D_z to the intersection $\{T_2, C, D\}$. We could choose any one of the three, assign it to w and z and see where it got us, but it is preferable to delay such decisions as long as possible. So we will return to this later.

The PCE for D on T_1 is D on v so we set $v = T_1$. We now have the following picture

Notice that we have enforced the ordering $Mv(C, T_2, D); Mv(B, T_3, C); Mv(A, z, B)$. We know that $Mv(D, C, T_1)$ is before $Mv(C, T_2, D)$ and that $Mv(A, B, w)$ is before $Mv(A, z, B)$, but we do not know the exact order of the other actions. We now have a further restriction on D_z (currently $\{T_2, C, D\}$). The assertion $Cl(z)$ must be *during* the assertion C on D (so $z = D$ is not possible) and also *during* the assertion B on C (so $z \neq C$). So we can put $w = z = T_2$ to get which in fact is the correct solution: $Mv(D, C, T_1)$; $Mv(C, T_2, D)$; $Mv(A, B, T_2)$; $Mv(B, T_3, C)$ and then $Mv(A, T_2, B)$. Notice that this was deduced without any backtracking.

5.7 Solving the Collapsing Problem

Suppose we have a planning problem and every goal assertion has a PCE. We wish to collapse a (large) number of edges, but we find it inconsistent to do so. We know that an action must be introduced for at least one interval in the causal gap but we don't know which one. Instead of trying each one in turn and backtracking we would like to either decompose the problem using the notion of independence (section 4.5) or to focus our search in some other way. We use a special version of a *dependency* (see chapter 4) which involves only the equality relation.

5.7.1 Folding

Let N be an interval network and $U \subseteq N \times N$ be a set of edges which we wish to collapse. Suppose this is inconsistent with N.

Definition Define a *fold* as a dependency involving only the equality relation. A fold can be represented simply as a set of edges since there is no ambiguity about the relation.

Examples

In the diagram there is only one fold - $\{(i,j),(j,k)\}$.

In this diagram there are four folds: $\{a,b,c\},\{a,v\},\{u,c\},\{u,v\}$.

5.7.2 Using folds for collapsing

Idea: suppose we wish to collapse 100 edges but edge47 and edge83 cannot collapse simultaneously. We know that an action must be introduced somewhere and a fold can help focus our attention. Our fold is {edge47, edge83} and this tells us that an action must be introduced for one of these two intervals. We may still have to backtrack, but only once instead of 100 times. In this way the principle of least commitment is preserved as far as possible.

Modified planning algorithm We have a planning problem (G, PE, IN) where G is a set of goal clauses, PE is the planning environment and IN is the interval network. A goal clause $p@i$ has a potential causal explanation (PCE) if there is an assertion $p@j$ in the planning environment with the same property (p) such that the set of relations in the interval network between i and j includes equality. We will denote the *causal explanation* by the edge (i, j).

1. Continue to introduce actions for goal clauses without PCEs until every goal clause has a PCE.
2. Let $U = \{\text{causal explanations } (i_g, j_g)\}_{g \in G}$. If there are no folds then U is consistent so collapse all the arcs and terminate successfully. Otherwise, find a fold in U (a tractable algorithm which finds this is described later).
3. Back-tracking point. Choose one arc from the fold and introduce an action for it. Go back to step 1

Note : if the fold is of size one (only one arc) then equality is not consistent on that arc. This is not a genuine causal explanation and an action must be introduced. If the propagation algorithm were complete it would have been deleted.

Question : Is it possible that the size of the smallest fold can exceed two? (Answer: probably yes, though examples of networks with only folds of size three or more have not been discovered by the author).

5.7.3 Algorithm to calculate folds

We have an interval network N and a set of arcs
$U \subseteq N \times N$. Let $U = \{(i_k, j_k) : k = 1 \ldots n\}$.

1. Base case. If N is already inconsistent then $\emptyset$ is a fold.
   Else,
2. Propagate $i_1 = j_1$ in N to get N^*.
3. Recursively, take N^* and find a fold D^* in N^* contained in $U \setminus \{(i_1, j_1)\}$.
4. (a) If collapsing D^* is inconsistent with N then D^* is a fold in N.
   (b) If collapsing D^* is consistent with N then $\{(i_1, j_1)\} \cup D^*$ is a fold in N.

Note :

- this algorithm has linear complexity, over the complexity of propagation. If we use a polynomial but incomplete propagation algorithm (like the Allen propagation algorithm) we will still produce an inconsistent set of collapsing arcs, but it may not be minimal. This may mean that we have slightly more back-tracking than needed. It is still very efficient at focusing the search space to a small set of choices.

- to *minimise* back-tracking we require not just *any* fold (minimal inconsistent collapsing set of arcs) but a fold of smallest size (global minimum). Global minima are always harder to find than local minima, but if we assume that the size of a smallest fold is bound by a relatively small integer B, say, our search space will be $O(n^B)$. Simply try to collapse all arcs singly, then try all pairs, then all triples and so on. We must find an inconsistent set by the time we check all sets of size B (of which there are ${}^nC_B = O(n^B)$). The first set we find to be inconsistent must be a global minimum.

Is it reasonable to assume that the size of the smallest fold is small? In the Allen and Koomen planner we search for goal clauses with no potential causal explanation and introduce actions for them. In our terminology, we find a fold of size one. Having dealt with all these they assume that it is consistent to collapse all the goal clauses. In other words, the size of the smallest fold is assumed to be one. Now, we know this isn't always the case. But it does seem fair to relax their assumption but keep the size of the smallest fold quite small (perhaps two or three).

So, in the planning algorithm (above) at step 2 we should look first for a fold of size 1, then 2 and so on till we find an inconsistent set of collapsing arcs which will be not only a fold, but a global minimum.

We shall see in the following section that the collapsing problem is not peculiar to the Allen and Koomen planner: it is a particular instance of the persistence problem and variants of it exist in many other temporal reasoner. The Allen and Koomen planner, as originally described, did not address this problem and was thus an incomplete planning system. In this section we have tackled the collapsing problem without any avoidable loss of efficiency by focusing the search using a *fold.*

Chapter 6

A Comparison of Allen and Koomen's Planner, TMM and the Event Calculus

6.1 Introduction

The Allen interval algebra and Dean and McDermott's TMM have both been used extensively in planning. There appear to be some similarities as well as a number of differences between the two approaches. Both methods have encountered serious problems which must be overcome if they are to develop into useful artificial planning architectures. The purpose of this section is to compare the two approaches and to synthesise some of the best features of both. Allen's system is described in section 1.1.1 and [Allen, 1983] and the reference for TMM is [Dean and McDermott, 1987]. At the most general level it can be seen that Allen's system, being interval based, is more expressive than the point based TMM. On the other hand this increased expressive power also renders the consistency-checking problem intractable for intervals. TMM is also able to represent metric information and has a polynomial time, complete algorithm for checking consistency. However, it is unable to express certain interval relations which commonly arise - there is a trade off between expressive power and tractability. An investigation of TMM's clipping and stretching is made and analogues are found in the interval algebra.

6.2 TMM

TMM is a point based system built on the relation algebra $\mathcal{M}$ and it represents quantitative, metric information. This architecture is based on a point based time map which is described in [Dechter *et al.*, 1991]. Temporal reasoning is performed by the Floyd-Walshall algorithm which is sound and complete for STCPs (simple temporal constraint problems, time maps or, in the language of this thesis, $\mathcal{M}$-networks with a single interval on each edge).

$$D(t, s, [a, b])$$

asserts that $s - t$ lies in the interval $[a, b]$ and

$$M(t, s, [a, b])$$

represents the *minimised* temporal relation (tightest bound) between s and t, calculated by the algorithm. The difficulty arises when you try to deal with intervals (or in the terminology of [Dechter *et al.*, 1991] *tokens*). Let us represent a token by a triple (p, st, end) where p is an atomic proposition (the token type), st and end are time points representing the start and finish of the token. These may not be known exactly, but there will be constraints on them in the usual way for STCPs. The token need not represent a maximal interval for p but given that p is *point-based* (defined later) and that the flow of time is Dedekind complete it is always possible to choose a maximal interval if we want to. We will say more on maximal intervals later. Let's not worry too much about closed or open intervals.

6.3 Clipping

Clipping is introduced in TMM to deal with potential contradictions, for example the tokens $T_1 = (p, s_1, e_1)$ and $T_2 = (q, s_2, e_2)$ may contradict if p and q are *contradictory token types* and if the intervals $[s_1, e_1], [s_2, e_2]$ possibly overlap. Because the relation between points is not known exactly, the relation between the two intervals is either

1. definitely do not overlap
2. possibly overlap or
3. definitely overlap.

The latter two cases lead to *potential* or *definite* contradictions between the tokens. In the original TMM there was only one procedure to detect such a potential contradiction. If it is known definitely that T_1 starts before T_2 i.e. $D(s_1, s_2, (0, \infty))$ and that T_1 might continue into T_2 ($M(e_1, s_2, [a, b])$ $a < 0$) then there is a potential contradiction which can be resolved by *clipping* : assert that T_1 is entirely before T_2 by asserting $D(e_1, s_2, [0, \infty))$.

But potential contradictions may arise in other ways too (see [Materne and Hertzberg, 1991]). For example, suppose instead that T_1 definitely ends after T_2. In this case T_1 should be clipped from the left by saying T_1 starts after T_2.

So, given two tokens T_1 and T_2 with contradictory types:

- if s_1 is definitely before s_2 then clip by asserting e_1 before s_2 (similarly if s_2 is definitely before s_1 clip by asserting e_2 before s_1)

- if e_1 is definitely after e_2 then clip by asserting s_1 after e_2 (similarly if e_2 is definitely after e_1 then clip by asserting s_2 after e_2).

Of course if we have definite overlap then clipping will produce an inconsistency -there is simply no way to resolve this type of contradiction.

There is one case left though: what if it is not known which token starts first and it is not known which ends first either?

Example : suppose that between 0 hrs. and 10 hrs p occurs for at least one hour and q occurs for at least two hours. $T_1 = (p, s_1, e_1)$ $T_2 = (q, s_2, e_2)$. $D(0, t, [0, 10])$ for $t = s_1, s_2, e_1, e_2$. T_1 lasts at least one hour so $D(s_1, e_1, [1, 10])$ and T_2 lasts at least two hours so $D(s_2, e_2, [2, 10])$. It is not known which starts first or which ends first. We are confronted with the problem: you can't express the disjointedness of two intervals using only constraints on points. A disjunction is needed:- either T_1 is before T_2 - $D(e_1, s_2, [0, \infty))$ or T_2 is before T_1 - $D(e_2, s_1, [0, \infty))$. But this disjunction is not expressible in TMM. It is unsatisfactory to leave this choice to the user (as in [Materne and Hertzberg, 1991]) - it must be dealt with by the temporal reasonner even if this means a loss of tractability.

We are trying to express that the intervals $i_1 = [s_1, e_1]$ and $i_2 = [s_2, e_2]$ are disjoint. In the Allen algebra where intervals are primitive this can be asserted by saying $i_1\{<, m, m^{\smile}, >\}i_2$. The disjunction $i_1\{<, m\}i_2$ can be expressed in a point based system as $D(e_1, s_2, [0, \infty))$ and similarly the disjunction $i_1\{m^{\smile}, >\}i_2$ can be asserted as $D(s_1, e_2, (-\infty, 0])$. In certain situations there might be additional information which rules out one or other of these cases. For example it might be known that s_1 is definitely before s_2, in which case we can rule out $i_1\{m^{\smile}, >\}i_2$ and assert the other disjunct, so $D(e_1, s_2, [0, \infty))$. That is what TMM calls *clipping*.

Alternatively, it might be certain that e_1 is after e_2, in which case $i_1\{m, <\}i_2$ can be ruled out and $D(s_1, e_2, (-\infty, 0])$ can be asserted. However, there are two additional points to be made. Firstly, as mentioned above there is another case where it is not known which way round the two intervals should go. Clipping in this case cannot be expressed in TMM. Secondly, comparison with the Allen algebra yields a more accurate test for clipping. It is a test which will cover both cases above and include additional situations too. In order to rule out $i_1\{m^{\smile}, >\}i_2$ for contradictory token types, it is sufficient to know that s_1 is definitely before e_2. This subsumes the case where we know that s_1 is before s_2 which is too strict a requirement.

Clipping Rule:

Let $T_1 = (p, s_1, e_1)$ and $T_2 = (q, s_2, e_2)$ be tokens with contradictory token types p and q. If $M(s_1, e_2, (a, b))$ with $a \geq 0$ (s_1 definitely before e_2) then clip by asserting $D(e_1, s_2, (0, \infty))$

$$((s1 < e2) \rightarrow (T1 \text{ entirely before } T2) \rightarrow (e1 < s2)).$$

Similarly if it is known that e_1 is after s_2 then rule out $i_1\{m, <\}i_2$ and assert that s_1

is after e_2. This is covered by the same rule but with the two tokens the other way round. Note that an application of the rule may be thought of either as clipping T_1 on the right (T_1 must stop before T_2 starts) or as clipping T_2 on the left (T_2 mustn't start until T_1 stops). In reality neither view is quite correct. Consider the diagram below:

Clearly this conflict could be resolved by terminating T_1 at time s and putting no new restriction on T_2 (clipping T_1 on the right) or by terminating T_1 at time t and starting T_2 also at t, but neither case can be deduced *a priori* from the situation. All that is known is that the end of T_1 must be before the beginning of T_2 and this is all that the rule asserts.

The problematic case is where it is possible for the two tokens to be arranged in either order. If $M(s_1, e_2, [a, b])$ and $a \leq 0$ (can't rule out $i_1\{>, mi\}i_2$) and $M(s_2, e_1, [c, d])$ with $c \leq 0$ (can't rule out $i_2\{>, mi\}i_1$) then it is not possible to constrain the time points in any way so as to be sure that no inconsistency arises (p and q true together) but which allows all consistent values to remain. In order to rule out all inconsistent values but leave all consistent values in, it is necessary to specify a choice of two cases: either T_1 is before T_2 or T_2 is before T_1. This extends the language of TMM and means a loss of tractability but it is unavoidable. This case is referred to as the disjunctive clipping rule and produces a choice of two simple clippings.

Disjunctive Clipping Rule:
Let $T_1 = (p, s_1, e_1)$ and $T_2 = (q, s_2, e_2)$ be tokens with contradictory token types p and q and let $M(s_1, e_2, (a, b)), M(s_2, e_1, (c, d))$ be the constraints on the end-points.
If $a \geq 0$ then assert $D(e_1, s_2, [0, \infty))$ (*ordinary clipping*)
If $c \geq 0$ then assert $D(e_2, s_1[0, \infty))$ (" ")
Otherwise, $(a, c < 0)$ assert $D(e_1, s_2, [0, \infty)) \vee D(e_2, s_1[0, \infty))$

6.4 Maximality of Intervals

There are two distinct uses of the word 'maximal' in the literature which should not be confused.

Definitions

- Let h : {propositions} $\rightarrow$ {intervals} be an assignment of the atomic propositions to intervals. We call h *point-based* if for all propositions p and intervals i

$$i \in h(p) \text{ iff } \forall t \in i \ [t, t] \in h(p)$$

(p is true at i if and only if p is true at all the points in i). This is rather stronger than the usual definition of downward closure (see [Richards *et al.*, 1989]).

- Given an interval assignment h and a property p the interval i is said to be *maximal for* p if p is true at i but p is not true at any interval which properly contains i.

  This is what Allen and Koomen have in mind when they choose maximal intervals for all assertions.

PROPOSITION 35 *Let $(T, <)$ be a Dedekind complete flow of time and let h be a point-based assignment. If p is true at some interval i then there is an interval j containing i such that j is a maximal interval for p.*

The other use of the word maximal comes about when there is no fixed assignment h, only constraints on an assignment.

Definition An interval i is a *maximal extension* of p if it is *consistent* for p to be true at i but it is not consistent for p to be true at any extension of i.

The persistence principal in the event calculus [Kowalski and Sergot, 1986] uses maximal extensions for propositions.

The difference between the two definitions can be illustrated by an example.

Example

An exact assignment for p is not given but it is known that $[s, t] \subseteq h(p)$ (p is true at $[s, t]$) and it is possible that p remains true up to u. The maximal extension of p is $[s, u]$. The maximal interval for p is not known exactly, it will depend on the assignment h, but it will be of the form $[s, v]$ where v is between t and u and v depends on the assignment h. It is implicit that there is a maximal interval for p which starts at s and ends sometime between t and u.

In the Allen and Koomen planner, maximal intervals are chosen to simplify the interval network. If i and j are maximal intervals for the same property p it can be deduced that the only possible relations between i and j are $\{<, =, >\}$. Two such intervals are either disjoint or identical. The clipping rule (above) can be used to achieve the same result. In the example above let T be the token (p, s, e) (s is the start and e is the end) where $t \leq e \leq u$. Define two new tokens $S = (\neg p, s, s)$ and $E = (\neg p, e, e)$[1][2] The clipping rule

---

[1] Strictly, only atomic propositions are allowed as token types. An alternative is to define a token $S = (q, s, s)$ and assert that q and p are contradictory properties.

[2] This assumes that p is true on the open interval (s, e). Closed intervals can be dealt with by defining $S = (\neg p, w, s)$ where $w < s$. The problems of dealing with closed, open or semi-open intervals is not dealt with here, but can be handled by this type of trick.

will enforce the maximality of the interval (s, e) for p, as we now show. Suppose there is another token $T^* = (p, s^*, e^*)$ and as before define $S^* = (\neg p, s^*, s^*), E^* = (\neg p, e^*, e^*)$. Suppose also that the ordering of s and s^* is not known and neither is the ordering of s and e^*, e and s^* nor e and e^*. So initially any of Allen's thirteen relations could hold between the intervals (s, e) and (s^*, e^*). The clipping rule will rule out all but three of these possibilities. Since S^* and T have contradictory properties (p and $\neg p$) and might overlap (potential contradiction) the disjunctive clipping rule should enforce that either $s^* \leq s$ or $s^* \geq e$. Similarly $e^* \leq s$ or $e^* \geq e$ can be deduced. Since there is a potential contradiction between S and T^* and between E and T^* it follows that $s \leq s^*$ or $s > e^*$ and $e \leq s^*$ or $e \geq e^*$.
Case 1: $e^* \leq$ s. Here $(s^*, e^*) < (s, e)$.
Case 2: $s^* \geq e.(s^*, e^*) > (s, e)$.
Case 3: $s^* \leq s, e^* \geq e, s \leq s^*$ and $e \geq e^*$. So $s^* = s$ and $e^* = e$. $(s^*, e^*) = (s, e)$.

6.5 Collapsing, stretching, persistence

Consider now the problem of database query. It is given that p holds from s to t (see diagram) and the query "is p true at (v, u)?" is asked.

It is consistent that p is true over (v, u) but it is not necessarily true. If the answer given is 'yes' then it may be necessary to change it later. This is a non-monotonic step and is referred to in TMM as *stretching*. In the Allen and Koomen planner the exact analogue is called *collapsing*. For example, suppose that "p true at i" is in the planning environment and there is a goal (or query) - "is p true at j?". Let the relations between i and j be $\{<, =, >\}$. It is possible to satisfy the goal (or give a positive answer to the query) by collapsing i and j and asserting $i = j$. Again, although this is a consistent possibility, it is not necessarily true, it may be necessary to change this answer later. In TMM the stretching rule given by Dean and McDermott is the following: let the token (p, s, t) be in the database and suppose the query (p, v, u) is asked. If it is known that s is definitely before v and it is consistent for t to be as late as u then the query may be answered positively provided it is asserted that t is greater than or equal to u. As with clipping, a comparison with Allen's system will indicate a generalisation of this rule. Note that the intervals corresponding to the tokens are not necessarily maximal for p.

Let $p@i$ be in the database and let the query "is $p@j$ true?" be asked. The modified stretching rule should be: if it is consistent that j is contained in i ($j\{eq, st, end, dur\}i$) then you may answer the query positively by stretching and asserting that i is indeed contained in j.

Modified Stretching Rule

Let $T = (p, s, t)$ be in the database and let the query (p, v, u)? be asked. If it is consistent that

$$s \leq v \text{ and } t \geq u(\text{ or } M(s, v, [a, b]), b \geq 0 \text{ and } M(t, u, [c, d]), c \leq 0)$$

then you can stretch by asserting

$$s \leq v \text{ and } t \geq u(\text{ or } D(s, v, [0, \infty))) \text{ and } (D(t, u, (-\infty, 0]))$$

Notes

1. This generalises the notion of stretching in [Dean and McDermott, 1987] because here we allow stretching to the left as well as to the right.
2. Under the additional assumption that both intervals are maximal then stretching becomes identical to collapsing. For, the only consistent way that one maximal interval for p can be contained in another is if they are made identical.
3. The modified stretching rule does not introduce any disjunction (as for clipping). There is no additional cost for achieving a more general rule.

6.6 The Allen and Koomen Planner in TMM

The Allen and Koomen planner handles the search space well because of its adherence to the principle of least commitment. The non-monotonic step of collapsing can be delayed as long as possible until the only goals left are those with potential causal explanations (PCEs) or those for which collapsing is a possibility. There is a remaining problem of what to do at this stage if every goal clause has a PCE but it is not consistent to collapse all the goal clauses simultaneously. See section 5.7 for a more detailed discussion of this problem. A more serious problem for this planner is that it is based on the Allen interval algebra which is intractable. The correspondence between some aspects of the interval algebra and TMM (above) indicates that it might be fruitful to translate Allen and Koomen's method into TMM. It would of course remain intractable because it is sometimes necessary to introduce disjunctions (see Clipping, above). On the other hand it is arguable that in TMM the cause of the intractability would be transparent (the disjunctions) and we would also gain TMM's ability to express metric relations between points.

The Planning Problem All assertions will be represented by tokens (p, s, e). To be more accurate the notation should express whether the intervals are closed, open or semi-open, but these details are omitted here. A planning problem will be a triple (PE, G, TM) where PE and G are sets of assertions (tokens) representing the planning environment and goal set respectively. If it is intended that the intervals are maximal intervals then additional tokens must be added as described in the section on maximality (see section 6.4). TM is the time map and consists of metric constraints between all pairs of points.

The Algorithm Assertions in $g \in G$ are removed until there are none left. This can be done in two ways.

Stretching $g = (p, s, e)$ has a PCE if there is some assertion $(p, s^*, e^*) \in PE$ and it is possible that $[s, e] \subseteq [s^*, e^*]$. These goal clauses should be left till last.

Actions If g has no PCE then an action must be introduced to remove it. g can then be removed from G but the preconditions for the action must be added to G. If there is only one action which has p as the property of one of its post-conditions, then this step can be taken without the possibility of having to back-track.

The action introduction step is continued until every goal has a potential causal explanation.

The Collapsing Problem At this point every goal has a PCE so for all the goal assertions (simultaneously) try to stretch the PCE so that it covers the goal. If this can be done consistently, then quit - the problem is solved. However, it is known for intervals that even when each goal clause has a potential causal explanation, it may be inconsistent to collapse them all at the same time. It turns out that this problem carries over to the point-based version too.

6.7 Planning in the Event Calculus

The event calculus is another point-based system where all properties are deduced from what is known at local events (see [Kowalski and Sergot, 1986]). There are no interval tokens here though a property p may be inferred to hold over an interval (e_1, e_2) using the *persistence axiom* (which is based on negation as failure) if

- e_1 starts p
- e_2 terminates p and
- not clipped(e_1, p, e_2).

Clipped(e_1, p, e_2) holds if there is an event e which terminates p and the time of e is between the times of e_1 and e_2. The event calculus is implemented in prolog, but in order to assert that two intervals are disjoint it is necessary to use non-horn clauses. Intractability has crept in already.

A version of the Allen and Koomen planner, translated into the language of the event calculus could be constructed as follows. Let each event consist of the event name, the properties initiated by the event and the properties terminated by the event. An action consists of a group of events together with temporal constraints on the times of occurrence of the events. This definition allows actions to be non-instantaneous. For example the action "$move(A, from, to)$" could be defined as a group of four linearly ordered events $\{e_1 < e_2 < e_3 < e_4\}$ where

- e_1 initiates held(A)
- e_1 terminates clear(A)
- e_2 initiates moving(A)
- e_2 terminates on(A, from)
- etc.

A planning problem consists of a set of events and temporal relations on their times, together with a list of available actions. A planning problem is solved if there is a linear arrangement of the events satisfying all the temporal constraints such that for each event e if e terminates p then there is an event e^* occurring before e such that e^* initiates p and p is not clipped between e^* and e.

Examples

1.

   This planning problem consists of two events :e, f. Event e initiates $on(A, B)$ and $clear(A)$. Event f terminates $on(A, B)$ and $clear(A)$. Event e is constrained to be before event f. This problem is solved already since both properties terminated by event f are initiated by event e.

2.

> Here again we have two events: e and f. Event e initiates $clear(A)$ and $clear(B)$. Event f terminates $on(A,B)$ and $clear(A)$. e is constrained to be before f. This problem is not yet solved since f terminates $on(A,B)$ and there is no event which initiates that property. In order to solve this problem an action must be introduced which initiates $on(A,B)$ and the event which initiates $on(A,B)$ must be constrained to be before f.

In line with previous comments on the Allen and Koomen planner, a property p terminated by an event e has a *potential causal explanation* if there is an event e^* which initiates p such that it is *consistent* that e^* occurs before e and it is consistent that no other property which terminates p occurs between e^* and e. Collapsing corresponds to the assertion that e^* does indeed occur before e and any event which terminates p lies either before e^* or after e.

Analogous with the Allen and Koomen planner, a planning algorithm can thus be devised which introduces new actions for each property terminated by an event which possesses no potential causal explanation. When all such properties have potential causal explanations then an attempt to linearise all the events so that each property collapses is made. If this succeeds, the plan is solved. If not, the 'collapsing problem' has to be dealt with. This can be done with by the method described in section 5.7

6.8 Conclusion

The differences between the Allen interval algebra, TMM and the event calculus are not as big as all that. TMM and the event calculus have the advantage of being point-based and therefore tractable. But, in order to be useful they must express the disjointedness of intervals which requires the introduction of disjuncts (or non-Horn clauses), thus becoming intractable. The Allen and Koomen planning method can be translated into TMM and the event calculus but the collapsing problem is inherited. Nevertheless, TMM has two remaining advantages: its ability to express metric constraints on time points and a search mechanism based on truth maintenance (not described here).

Chapter 7

Final Conclusions

In the end the final score was 15 (theorems) to 12 (problems) so the theorems won. However this gives no grounds for complacency as some of the problems are very hard and the solution of a problem often raises at least two new ones. The day when the study of relation algebra is fully transparent is still some way off.

It turned out that handling binary relations algebraically was more problematic than for unary relations. The first contribution of this thesis was to apply the methods of amalgamation theory and forcing to that problem. The study of finite substructures (networks) clarified and deepened our understanding of the representations of relation algebras.

Then, given a representation, we showed how to construct a new relation algebra which could handle binary relations on the intervals. The move from points to intervals involved a loss of tractability.

Perhaps the reason for the difficulty with relation algebra is the enormous power that binary relations possess. Tarski showed that they could be used as a vehicle for set-theory and hence all of mathematics. This does mean that we are not dealing with an artificial subject here, but one which arises directly from a range of scientific problems and has many applications to them.

We illustrated this here by applying the theoretical work to temporal reasoning in artificial planning. Formidable difficulties occur in this very practical subject which have not been overcome by ad-hoc techniques restricted to the immediate problem field. In this thesis a considerable advantage was gained by translating from the application to the more abstract language of relation algebra. One application of the work on interval algebras was the construction of a relation algebra which represented metric as well as interval relations. The so-called collapsing problem from the Allen and Koomen planner was solved - a problem closely related to the persistence problem in temporal reasoning.

A number of important questions have been raised in the text and remain unanswered: typically these are phrased in an abstract way but have direct consequences for the application areas.

- Is an atomic relation algebra with a single TAP-class ω-categorical?
- Under what conditions can a relation algebra be extended to a complex group? Is it

sufficient for the networks of the algebra to form an amalgamation class? A positive answer here would translate an important subclass of relation algebras to a relatively well understood subject.

- For non-atomic networks, when is it the case that transitive closure guarantees consistency? A necessary condition is that for every subalgebra $\mathcal{B}$ of $\mathcal{A}$ there is a representation of $\mathcal{A}$ which is fully universal over $\mathcal{B}$. Is this condition sufficient? This is directly applicable to temporal reasoning and a positive result would identify nearly all of the tractable relation algebras.
- Characterise the representable relation algebras for which the network satisfaction problem is tractable.
- Apply the game-theoretic approach of chapter I to the characterisation of representable cylindric algebras.

We look forward to further research in this fascinating area and hope to answer some, at least, of these questions.

Part IV

Appendices

Appendix A

Proof that the fixpoint algorithm terminates on TCPs

A TCP is defined in [Dechter *et al.*, 1991] as an $\mathcal{M}'$-network where the metric constraint on an edge is a finite union of intervals, possibly with irrational endpoints. e.g.

Let us gloss over the problem that algorithms do not handle real numbers as these are not all finitely representable and answer an open problem raised in [Dechter *et al.*, 1991]

THEOREM 36 *The fixpoint algorithm (page 66) always terminates finitely.*

PROOF:

One iteration of the algorithm takes some triangle (a, b, c) from the network N and replaces $N(a, c)$ by $N(a, c) \ \cap \ (N(a, b) * N(b, c))$. This composition is calculated by addition of intervals. Let $S(N)$ be the (finite) set of all endpoints of intervals mentioned in N. If a number occurs more than once in $S(N)$ then label each instance separately in order to distinguish them.

Claim: at each stage the relation between nodes n and m is either $\emptyset$ (inconsistent) or equal to a finite union of intervals and each endpoint is a sum of distinct[1] elements from $S(N)$. This claim can be proved by induction on the

---

[1] Why must the elements in the sum be distinct? Because if the same element occurred twice it would correspond to a constraint on the edge (a, c) created by a looping path. However, either a loop produces

number of iterations of the algorithm. Now there are only a finite number of possible sums that can be produced this way and therefore only a finite number of possible intervals that can occur on an edge at any stage of the algorithm. The relations on each edge are never increasing so each edge can be placed in the queue a finite number of times and therefore the algorithm must terminate. □

---

an inconsistency (so the algorithm terminates) or an equally tight constraint is produced from the path with the loop deleted.

Glossary

Age	$Age(X)$ is the class of finite, atomic networks which can be isomorphically embedded in the representation X
Allen Propagation Algorithm	The fixpoint algorithm that calculates the transitive closure of an interval network.
Amalgamation class	A class of finite, atomic $\mathcal{A}$-networks closed under isomorphism, downward closed and having the amalgamation property and joint embedding property.
Amalgamation property	A class $\mathcal{K}$ of $\mathcal{A}$-networks has the amalgamation property (AP) if for any $L, N_1, N_2 \in \mathcal{K}$ and embeddings $f_i : L \to N_i$ $(i = 1, 2)$ there is a network $M \in \mathcal{K}$ and embeddings $g_i : N_i \to M$ such that $f_1 g_1 = f_2 g_2$.
Atom	A minimal non-zero element of a relation algebra.
Atomic network	A network where each edge is labelled by an atom.
Atomic relation algebra	A relation algebra where every non-zero element contains an atom.
Complete representation	A representation is complete if it preserves arbitrary unions. For an atomic relation algebra this is equivalent to saying that every pair of points in the representation is in the range of some atom.
Chain condition	A class of networks has the d-chain condition if for every pair of networks $N \subseteq M$ from the class there is a chain $N = N_0 \subset N_1 \subset \ldots \subset N_k$ such that $M \subseteq N_k$ and for each i N_{i+1} has the d-support property in N_i.
Complete network	A (not necessarily atomic) network is complete if for each edge and each atom contained in the relation on that edge there is a model of the network which realises that atomic relation.
Complete relation algebra	A relation algebra (or Boolean algebra) is complete if arbitrary intersections and unions exist.
Complex Group	A relation algebra where the atoms form a group
Dependency	A minimal inconsistent labelling.
Extension property	A (not necessarily atomic) finite network N has the extension property if for any subnetwork M of N every model of M extends to a model of N.
Fold	A dependency with each edge labelled with equality.

Fraïssé limit	The unique countable, homogeneous and universal representation of a relation algebra whose atomic networks form an amalgamation class.
Fully Universal Relation Algebra	$\mathcal{A}$ is fully universal if $Net\mathcal{A})$ is a TAP-class with JEP.
Fully Universal Representation	A representation which embeds every transitively closed atomic network.
Homogeneous	A representation is homogeneous if any local isomorphism extends to a full automorphism of the representation.
Independence	Two sets of edges E and F from a network N are independent if every pair of labellings of E and F consistent with N has a joint extension to a labelling of N.
Integral	A relation algebra is integral if $a * b = 0$ implies $a = 0$ or $b = 0$.
Joint embedding property	A class $\mathcal{K}$ of $\mathcal{A}$-networks has the joint embedding property (JEP) if any two networks from $\mathcal{K}$ embed in a third also belonging to $\mathcal{K}$.
Labelling	A labelling of a network is a tightening of the network to an atomic network.
Model	A model of a network N is an isomorphism from N into some representation of the relation algebra.
Network	An $\mathcal{A}$-network is a transitively closed, directed graph with each edge labelled by an element of $\mathcal{A}$. In chapters 1 to 3 only atomic networks are considered.
Non-singular	An element a of a relation algebra is non-singular if $a * a^{\smile} = Id$.
Potential causal explanation	A goal assertion $p@i$ has a potential causal explanation (PCE) if there is an assertion in the planning environment $p@j$ and it is consistent that $i = j$.
Proper relation algebra	A set of binary relations on a domain D closed under the Boolean operations, composition, converse and containing the equality relation.
RRA	The class of all relation algebras isomorphic to a proper relation algebra.
Reference Intervals	Given an interval network N a subset $S \subseteq N$ is a set of reference intervals if every interval in the network is *contained* in some reference interval from S.

Regular representation	The representation of a non-singular relation algebra with domain equal to the set of atoms and an element a of the relation algebra interpreted as the set of pairs (α, β) where $\alpha * a \geq \beta$.
Relation algebra	A tuple $(A, \vee, -, 0, 1, *, {}^{\smile}, Id)$ which satisfies the Tarski axioms (p 9).
Representation	A representation of a relation algebra $\mathcal{A}$ is a domain D and a map I from $\mathcal{A}$ to $\mathcal{P}(D \times D)$ such that the image forms a proper relation algebra and I preserves all the operations.
STCP	Simple Temporal Constraint Problem. An $\mathcal{M}$-network with a single interval on each edge.
Normal Representation	A complete, fully universal, homogeneous representation.
Complete Representation	A representation which preserves arbitrary unions (and intersections) wherever they are defined.
Support property	A network $M \supset N$ has d-support in N if there is a sub-network $L \subseteq N$ of size d or less such that the type of M over L determines the type of M over N.
Tightening	A tightening of a network N is a network M with the same set of nodes such that for all $a, b \in N$ we have $M(a, b) \leq N(a, b)$.
Triangle Addition Property	A class $\mathcal{K}$ of atomic $\mathcal{A}$-networks has the triangle addition property (TAP) if it is closed under isomorphism, downward closed and for each $N \in \mathcal{K}$ and each consistent triangle T such that there is an edge (s, t) of T and an edge (n, m) of N labelled by the same atom r, it is possible to embed N and T in a network M say in such a way that (s, t) and (n, m) embed to the same edge of M and $M \in \mathcal{K}$.
TAP-class	A class having the triangle addition property.
Transitively closed	A network N is transitively closed if for all nodes $l, m, n \in N$, $N(l, m) * N(m, n) \geq N(l, m)$.
Unit	An atomic relation contained in the identity of a relation algebra. "Unit" is also used to describe the top element of a boolean algebra.

Universal	A representation is universal over $\mathcal{K}$ if every network from $\mathcal{K}$ embeds in the representation. A representation X is universal if it is universal over the maximal TAP-class $\tau(\mathcal{A})$ or, equivalently, if every network from every representation embeds into X.

Notation

$\mathcal{A}^2_r$	The interval algebra generated by the relation algebra $\mathcal{A}$ and the atom r.
$B^{<\omega}$	The set of finite sequences of elements of B.
$\mathcal{C}$	The containment algebra.
$F(a)$	The principal filter generated by a.
$\mathcal{I}$	The Allen interval algebra.
$L(\mathcal{A})$	The first-order language with one binary predicate symbol for each element of $\mathcal{A}$.
$\mathcal{M}$	The metric system of Dechter, Meiri and Pearl. The elementary relations are of the form $[p,q]$ where p and q are rationals, which asserts that the distance between two points lies inside the interval $[p,q]$.
$\mathcal{M}'$	Same as $\mathcal{M}$ except irrational endpoints to the intervals are permitted.
$\mathcal{M}^2$	The metric interval algebra (p 64).
$Net(\mathcal{A})$	The class of all finite atomic $\mathcal{A}$-networks.
$N(n,m)$	The relation in the network N which holds between the nodes n and m.
$N^\dagger$	The $\mathcal{A}$-structure formed from the network N defined by $N^\dagger(a) = \{(m,n) \in N \times N : N(m,n) \leq a\}$.
$\mathcal{P}$	The point algebra with three atoms: $Id, <$ and $>$, with composition and converse as for dense linear order.
$Th(\mathcal{A})$	The first-order theory that defines representations of $\mathcal{A}$.
$\tau(\mathcal{A})$	The maximal TAP-class of $\mathcal{A}$.
X^{-1}	X is a representation, d, e are points in the representation, $X^{-1}(d,e) = \{a \in \mathcal{A} : (d,e) \in X(a)\}$. For simple relation algebras, $X^{-1}(d,e)$ forms an ultrafilter.
$\overleftarrow{X}(x,y)$	The unique atomic relation which holds between the points x and y in the complete representation X.

$\overleftarrow{X}(S)$	X is a complete representation, $S \subseteq X$. $\overleftarrow{X}(S)$ is the atomic network with S the set of nodes and for each pair $s, t \in S$ the edge (s, t) is labelled by the unique atomic relation which holds on (s, t) in X i.e. $\overleftarrow{X}(S)(s, t) = \overleftarrow{X}(s, t)$.
X_r^2	The set of r-intervals from a representation X i.e. the set of pairs (x, y) such that $\overleftarrow{X}(x, y) = r$.

Bibliography

[Allen and Koomen, 1983] J F Allen and J Koomen. Planning using a temporal world model. *IJCAI*, 8:711–714, 1983.

[Allen, 1983] J F Allen. Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26(11):832–843, November 1983.

[Allen, 1984] J F Allen. Towards a general theory of action and time. *Artificial Intelligence*, 23(2):123–154, 1984.

[Andréka *et al.*, 1991] H Andréka, D Monk, and I Németi. *Algebraic Logic*. North-Holland, 1991.

[Andréka, 1989] H Andréka. On the representation problem of distributive semilattice-ordered semigroups, 1989. preprint.

[Boole, 1948] G Boole. *The mathematical analysis of logic, being an essay towards a calculus of deductive reasoning*. Basil Blackwell, 1948.

[De Morgan, 1964] A De Morgan. On the syllogism, no. iv, and on the logic of relations. *Transactions of the Cambridge Philosophical Society*, 10:331–358, 1964.

[Dean and McDermott, 1987] T Dean and D McDermott. Temporal database management. *AI*, 32:1–55, 1987.

[Dechter *et al.*, 1991] R Dechter, I Meiri, and J Pearl. Temporal constraint networks. *Artificial Intelligence*, 49, 1991.

[Fikes and Nilsson, 1971] R Fikes and N Nilsson. Strips: a new approach to the application of theorem proving to problem solving. *Artificial Intelligence*, 2:189–208, 1971.

[Henkin *et al.*, 1971] L Henkin, J Monk, and A Tarski. *Cylindric Algebras Part I*. North-Holland, 1971.

[Henkin *et al.*, 1985] L Henkin, J Monk, and A Tarski. *Cylindric Algebras Part II*. North-Holland, 1985.

[Hirsch, 1992] R Hirsch. The problem of persistence in planning. Technical report, Imperial College Planning Project, 1992.

[Hirsch, 1994] R Hirsch. From points to intervals. *Journal of Applied Non-Classical Logics*, 1994.

[Hodges, 1985] W Hodges. *Building Models by Games.* Number 2 in London Mathematical Society Student Texts. Cambridge University Press, 1985.

[Jónsson and Tarski, 1948] B Jónsson and A Tarski. Representation problems for relation algebras. *Bulletin of the American Mathematical Society*, 54, 1948.

[Jónsson and Tarski, 1951] B Jónsson and A Tarski. Boolean algebras with operators. *American Journal of Mathematics*, 73:891–939, 1951.

[Jónsson and Tarski, 1952] B Jónsson and A Tarski. Boolean algebras with operators. *American Journal of Mathematics*, 74:127 – 162, 1952.

[Kautz and Ladkin, 1991] H Kautz and P Ladkin. Integrating metric and qualitative temporal reasoning. In *Proceedings of the ninth national conference on artificial intelligence, Angeim, California*, pages 241–246, 1991.

[Koomen, 1989] J Koomen. Localising temporal constraint propagation. *Knowledge Representation*, 1989.

[Kowalski and Sergot, 1986] R Kowalski and M Sergot. A logic-based calculus of events. *New Generation Computing (Japan)*, 4, 1986.

[Ladkin and Maddux, 1987a] P Ladkin and R Maddux. The algebra of convex time intervals. Technical report, The Kestrel Institute, Palo Alto, California, 1987.

[Ladkin and Maddux, 1987b] P Ladkin and R Maddux. The algebra of convex time intervals. Technical Report KES.U.87.2, The Kestrel Institute, 1987.

[Ladkin and Maddux, 1993] P Ladkin and R Maddux. On binary constraint problems. *Journal of the American Mathematical Society*, 1993.

[Ladkin, 1989] P Ladkin. Metric constraint satisfaction with intervals. Technical Report TR-89-083, ICSI, Berkeley, 1989.

[Ligozat, 1990] G Ligozat. Weak representation of interval algebras. In *Proceedings of eighth national conference on artificial intelligence, Mass.*, pages 715–720, 1990.

[Lyndon, 1950] R Lyndon. The representation of relational algebras. *Annals of Mathematics*, 51(3):707–729, 1950.

[Lyndon, 1956] R Lyndon. The representation of relation algebras, ii. *Annals of Mathematics*, 63(2):294–307, 1956.

[Lyndon, 1961] R Lyndon. Relation algebras and projective geometries. *Michigan Mathematics Journal*, 8:207–210, 1961.

[Maddux, 1989] R D Maddux. Non-finite axiomatizability results for cylindric and relation algebras. *Journal of Symbolic Logic*, 54(3):951–974, 1989.

[Maddux, 1991] R Maddux. The origin of relation algebras in the development and axiomatization of the calculus of relations. *Studia Logica*, 3/4, 1991.

[Materne and Hertzberg, 1991] S Materne and J Hertzberg. MTMM - correcting and extending time map management. In *Proceedings of the European Workshop on Planning*, March 1991.

[McKenzie, 1970] R McKenzie. Representation of integral relation algebras. *Michigan Mathematics Journal*, 17:279–287, 1970.

[Meiri, 1991] I Meiri. Combining qualitative and quantitative constraints in temporal reasoning. Technical Report R-160 UCLA, UCLA, December 1991.

[Monk, 1964] D Monk. On representable relation algebras. *Michigan Mathematics Journal*, 11:207–210, 1964.

[Németi, 1991] I Németi. Algebraisations of quantifier logics, an introductory overview. *Studia Logica*, 50(3/4):485 - 570, 1991.

[Pelavin, 1988] R Pelavin. *A formal approach to planning with concurrent actions and external events*. PhD thesis, University of Rochester, 1988.

[Richards *et al.*, 1989] B Richards, I Bethke, J Van der Does, and J Oberlander. *Temporal Representation and Inference*. Academic Press, 1989.

[Stone, 1936] M Stone. The theory of representations for boolean algebras. *Transactions of the American Mathematical Society*, 40:37–111, 1936.

[Tarski, 1941] A Tarski. On the calculus of relations. *Journal of Symbolic Logic*, 6:73–89, 1941.

[Tarski, 1954 and 1955] A Tarski. Contributions on the theory of models. In *Proc. Koninkl. Nedel. Akad. van Wetensch*, volume 57 and 58 of *A*, pages 572–588 and 56–64 resp., 1954 and 1955.

[van Beek and Cohen, 1990] P. van Beek and R. Cohen. Exact and approximate reasoning about temporal relations. *Computational intelligence*, 6:132–144, 1990.

[Villain and Kautz, 1986] M Villain and H Kautz. Constraint propagation algorithms for temporal reasoning. *Knowledge Representation*, pages 377–382, 1986.

[Villain *et al.*, 1989] M Villain, H Kautz, and P van Beek. Constraint propagation algorithms for temporal reasoning — a revised report. *Readings in Qualitative Reasoning about Physical Systems*, pages 373–381, 1989.