WORKING PAPER NO. 232

AN AXIOMATIC STUDY OF COMPUTATION IN HYPERTREES

by

Glenn Shafer

October 1991

An Axiomatic Study of Computation in Hypertrees[1]

Glenn Shafer[2]
Wednesday, October 16, 1991

Abstract. Theories of computation in acyclic hypergraphs (or hypertrees, as they are called here) have been developed for a variety of problems, from solving sparse linear equations to finding posterior probabilities in Bayesian expert systems. These theories are disparate in many ways, but they all exploit the tree-like structure of the hypertree. They all involve a step-by-step computation inward to the middle of the tree, and some involve a further step-by-step computation outward.

This paper undertakes to distill what these different theories have in common into general axioms relating a commutative semigroup to a lattice. These axioms can be varied to allow for differences in the uniqueness and flexibility with which various operations can be performed.

1. Introduction

Computation in acyclic hypergraphs (or hypertrees, as they are called here) has been studied for a variety of problems, including

—the solution of sparse linear equations (Rose 1970),
—dynamic programming (Bertelè and Brioschi 1972),
—the management of relational databases (Beeri, Fagin, Maier, and Yannakakis 1983; Maier 1983),
—constraint propagation (Dechter, Dechter, and Pearl 1990),
—belief computation in Bayesian expert systems (Lauritzen and Spiegelhalter 1988; Pearl 1988),
—belief computation in belief-function expert systems (Shafer, Shenoy, and Mellouli 1987; Dempster and Kong 1988),

and

—the solution of influence diagrams (Shenoy 1990; Shachter, Andersen, and Poh 1990, Ndilikilikesha 1991).

---

[1]This incomplete and tentative working paper builds on ideas developed by Prakash Shenoy and the author in a series of papers, including Shafer and Shenoy (1988, 1990) and Shenoy and Shafer (1990). The author has profited from conversations with Robert Cowell, Chandler Davis, Finn Jensen, Saunders Mac Lane, Steffen Lauritzen, Pierre Ndilikilikesha, and others. Further suggestions for improvement are welcome. The research has been partially supported by NSF grant IRI-8902444.

[2]School of Business, Summerfield Hall, University of Kansas, Lawrence, Kansas 66045, USA.

In each of these problems, one begins with several functions or other mathematical objects, each involving a different cluster of variables. The clusters form a hypergraph. When this hypergraph decomposes, the computational problem decomposes as well. In the most favorable case, when the hypergraph is a hypertree, the computational problem decomposes to the extent that it reduces to a sequence of computations within the clusters.

This paper undertakes to distill what these problems have in common into some general axioms. Because the purpose of the paper is to identify the bare essentials needed for hypertree computation to work, the setting for the axioms is abstract. The axioms are about domains (elements of a lattice), potentials (elements of a commutative semigroup), and relations among them. This level of abstraction is not fully needed, however, for the examples considered in the paper. In all these examples, the domains are sets of variables, with set inclusion as the lattice ordering. In the simplest examples, the potentials are functions on these variables, with either multiplication or addition as the semigroup operation.

In general, hypertree computation involves two phases, each with its own type of computation. The first phase moves inward in the hypertree, repeating a certain type of computation to successively simplify the problem. The second phase moves back out in the hypertree, repeating a different computation that uses the results of the first phase to solve the problem. In some problems, only the first phase is needed; but in other problems, both are needed.

In the inward phase, the emphasis is on some kind of reduction—some operation that simplifies a potential involving one domain to a potential involving a smaller domain. In general, reduction may involve some arbitrary choices, and the simpler potential on the smaller domain may not be unique. If it is unique, we call it a marginal. (In the probability problem, a marginal is obtained by summing out variables. Different operations are involved in other problems. We maximize out variables in dynamic programming, eliminate variables when solving linear equations, and so on.) All of our computational problems begin with potentials involving a number of relatively small domains. We want to combine these potentials using the semigroup operation and then reduce or marginalize the result, which may involve a large domain and hence be difficult to work with computationally, to one of the initial domains. It turns out that if the initial domains form a hypertree (sometimes additional conditions must also be satisfied), then this can be achieved by means of a sequence of combinations and reductions within domains as we move inward in the hypertree; the potential resulting from the combination over all the domains need not be dealt with explicitly or directly.

In the outward phase, the emphasis is on the solution of the problem that was successively reduced in the inward phase. We now suppose that along with each reduction we compute a partial solution, or extender, which relates the reduction to the potential being reduced. (An extender is a conditional probability function in the probability problem, a map showing where the maxima were attained in the dynamic programming problem, and a set of equations expressing the variables eliminated in terms of those that remain in the linear equations problem.) Formally, extenders are themselves potentials, but in many cases they can be represented and manipulated more easily than arbitrary potentials. In the outward phase, the extenders are combined in order to solve either the overall problem or a series of subproblems.

This paper gives several sets of axioms. This reflects the exploratory nature of the paper and also differences among problems in the uniqueness and flexibility with which computations can be performed. For each set of axioms, we derive the possibility of hypertree computation and give examples.

We consider only five computational examples: dynamic programming, factored probability distributions, sparse systems of linear equations, constraint propagation, and belief functions. Other examples are anticipated, however, by *some* of the material covered and by some aspects of the presentation. The use of arbitrary latttices anticipates examples where the objects with which we are dealing are not defined on sets of variables.

The reader will observe a tension between the desire for simplicity at the axiomatic level and the desire to stay as close as possible to computational reality. Both goals are important. Simple axioms can help us understand what is common to our different computational examples, but it is primarily the computational structure we want to understand, not the structure of a mathematical theory inspired by the computation.

Outline of the Paper. Chapter 2 reviews basic definitions for commutative semigroups, lattices, hypergraphs, and hypertrees.

Chapters 3 studies the inward phase of hypertree computation in the case where every potential has a marginal on every subdomain. This is an important special case, since it includes most of our examples.

Chapter 4 shows how an additional axiom, which is satisfied by most of our examples, allows us to simplify the axioms of Chapter 3. Unfortunately, the simplification is at the price of computational relevance.

Chapter 5 takes up the outward phase of computation in the case where every potential has a marginal on every subdomain, and in addition, extenders allow us to reconstruct potentials from their marginals. We call the extenders "continuers" in this case. Chapter 6 deals with extenders more generally, still assuming the existence of marginals.

Chapters 7 and 8 treat the case where reduction is always possible but may not be unique. Chapter 7 treats the inward phase of hypertree computation for this case, and Chapter 8 generalizes this treatment to account also for the outward phase.

Topics Not Discussed. This paper leaves aside many important issues and themes connected with the computational problems with which it deals. Most importantly, it does not discuss how to search for feasible hypertree covers or what to do when there are no feasible hypertree covers. These are both difficult questions.

There are only heuristics for searching for feasible hypertree covers (Bertelè and Brioschi 1972). There are iterative methods for some problems when there is no feasible hypertree cover, but there does not seem to be a theory at the level of generality at which we are working here.

We also do not discuss join trees. Roughly speaking, join trees provide an alternative representation for hypertrees. This representation makes the tree-like nature of hypertree computation much clearer, but it is not needed for the exposition here.

Finally, we do not discuss recursion or parallel computation. Many of the problems that fall under the theory presented here have been discussed in the context of recursion (because the same computation is required at each step inward or outward in the tree) and parallel computation (because computations in different branches of the tree can be carried out simultaneously). These are important aspects of the general theory as well, but there is no need to elaborate on them in this paper.

2. Mathematical Preliminaries

This chapter reviews elementary definitions and facts about several mathematical topics: commutative semigroups, lattices, graphs, hypergraphs, and hypertrees. Most of the definitions are standard. Some standard references are Petrich (1973) for semigroups, Birkhoff (1967) for lattices, and Berge (1973) for graphs and hypergraphs.

2.1. Commutative Semigroups

A semigroup is a pair $(\Phi,\otimes)$, where Φ is a set and $\otimes$ is an associative binary operation on Φ. When we say that $\otimes$ is a binary operation on Φ, we mean that $\otimes$ is a map that assigns to every ordered pair of elements of Φ another element of Φ. We write $\varphi_1\otimes\varphi_2$ for the element of Φ that $\otimes$ assigns to the ordered pair (φ_1,φ_2), and we call $\varphi_1\otimes\varphi_2$ the product of φ_1 and φ_2. When we say that the binary operation $\otimes$ is associative, we mean that

$$(\varphi_1\otimes\varphi_2)\otimes\varphi_3 = \varphi_1\otimes(\varphi_2\otimes\varphi_3)$$

for every triplet of $(\varphi_1,\varphi_2,\varphi_3)$ of elements of Φ. This implies that the result of combining any finite number of elements of Φ using the operation $\otimes$ does not depend on where the parentheses go. Thus we may write $\varphi_1\otimes\varphi_2\otimes\cdots\otimes\varphi_n$ for the product of $\varphi_1,\varphi_2,\ldots,\varphi_n$.

Though we must specify both the set Φ and the binary operation $\otimes$ when we specify a semigroup, we often mention only Φ explicitly, leaving $\otimes$ implicit. The next paragraph, for example, refers to a semigroup Φ.

A semigroup Φ is commutative if

$$\varphi_1\otimes\varphi_2 = \varphi_2\otimes\varphi_1$$

for every pair of (φ_1,φ_2) of elements of Φ. This implies that the result of combining any finite number of elements of Φ using the operation $\otimes$ does not depend on the order in which the elements are taken. Thus we may write $\otimes\{\varphi_i | i=1,\ldots,n\}$ for the product of $\varphi_1,\varphi_2,\ldots,\varphi_n$, and we may write $\otimes\{\varphi_h | h\varepsilon H\}$ for the product of $\{\varphi_h\}_{h\varepsilon H}$.

An element ι in a commutative semigroup Φ is an identity for Φ if $\varphi\otimes\iota = \varphi$ for every element φ of Φ. It is easy to see that identities are unique; there can be at most one identity in a given commutative semigroup. If a commutative semigroup does not have an identity, then we can simply adjoin one. This means that we replace $(\Phi,\otimes)$ by $(\Phi\cup\{\iota\},\otimes^*)$, where ι is not an element of Φ, and $\otimes^*$ is defined by setting $\iota\otimes^*\iota = \iota$, $\varphi\otimes^*\iota = \iota\otimes^*\varphi = \varphi$ for all $\varphi\varepsilon\Phi$, and $\varphi_1\otimes^*\varphi_2 = \varphi_1\otimes\varphi_2$ for all $\varphi_1\varepsilon\Phi$ and $\varphi_2\varepsilon\Phi$. It is easily verified that $\otimes^*$, defined is this way, is associative and commutative.

If Φ is a commutative semigroup, ϑ is a subset of Φ, and $\varphi_1\otimes\varphi_2$ is in ϑ whenever both φ_1 and φ_2 are, then we say that ϑ is a subsemigroup of Φ.

In this paper, we will call the elements of a commutative semigroup potentials.

2.2. Lattices

A partially ordered set is a pair $(\Im,\leq)$, where $\Im$ is a non-empty set and $\leq$ is a binary relation on $\Im$ that satisfies

$$x \leq x,$$

$$\text{if } x\leq y \text{ and } y\leq x, \text{ then } x=y, \qquad (2.1)$$

and

$$\text{if } x\leq y \text{ and } y\leq z, \text{ then } x\leq z.$$

An element x of a partially ordered set $\Im$ is called a zero for $\Im$ if $x\leq y$ for all y in $\Im$. It follows from (2.1) that if a zero exists, it is unique. If it does exist, we use the symbol $\varnothing$ to represent it. Similarly, an element x is called a unit if $y\leq x$ for all y in $\Im$. If a unit exists, it is unique, and we use the symbol Ξ to represent it.

Suppose x, y, and z are elements of a partially ordered set, $x\leq z$, and $y\leq z$. Suppose furthermore that $z\leq w$ for any other element w such that $x\leq w$ and $y\leq w$. In this case, we say that z is the least upper bound for x and y. Least upper bounds are unique when they exist. Greatest lower bounds are defined similarly, and they are also unique when they exist. The least upper bound of x and y is denoted by $x\vee y$, and the greatest lower bound is denoted by $x\wedge y$. Other names are also used: $x\vee y$ is sometimes called the join of x and y, and $x\wedge y$ is sometimes called the meet.

A partially ordered set in which least upper bounds and greatest lower bounds exist for all pairs of elements is called a lattice. In a lattice, both $\vee$ and $\wedge$ are associative and commutative. This means that for any finite subset H of the lattice, both $\vee H$ and $\wedge H$ are well defined. In fact, $\vee H$ is the least upper bound of the elements of H, and $\wedge H$ is the greatest lower bound of the elements of H.

In this paper, we call the elements of a lattice domains. If x and y are domains in a lattice, and $x\leq y$, then we say that x is a subdomain of y.

A lattice $\Im$ is modular if for all domains x, w, and y in $\Im$

$$x \le y \text{ implies } x \vee (w \wedge y) = (x \vee w) \wedge y. \tag{2.2}$$

The content of the equality in this condition is the requirement that $x \vee (w \wedge y) \ge (x \vee w) \wedge y$, for $x \vee (w \wedge y) \le (x \vee w) \wedge y$ will hold whether $\Im$ is modular or not. The following lemma gives another condition that is equivalent to (2.2).

Lemma 2.1. A lattice $(\Im, \le)$ is modular if and only if for all domains x,y, and z in $\Im$,

$$x \le z \le x \vee y \text{ implies } x \vee (y \wedge z) = z. \tag{2.3}$$

Proof: Suppose $(\Im, \le)$ is modular. Then by (2.2), $x \le z$ implies that $x \vee (y \wedge z) = (x \vee y) \wedge z$. But $z \le x \vee y$ implies that $(x \vee y) \wedge z = z$. So $x \le z$ and $z \le x \vee y$ together imply that $x \vee (y \wedge z) = z$.

Now suppose that (2.3) holds. Given domains x, w, and y such that $x \le y$, set $z = (x \vee w) \wedge y$. Then $x \le z \le x \vee w$, so by (2.3), $x \vee (w \wedge z) = z$, or $x \vee (w \wedge (x \vee w) \wedge y) = (x \vee w) \wedge y$. Since $w \wedge (x \vee w) = w$, this reduces to $x \vee (w \wedge y) = (x \vee w) \wedge y$.

Again, the content of the equality in (2.3) is the requirement that $x \vee (z \wedge y) \ge z$; the condition $x \vee (z \wedge y) \le z$ will hold whether $\Im$ is modular or not.

A lattice $\Im$ is distributive if for all domains x, w, and y in $\Im$

$$(x \vee w) \wedge y = (x \wedge y) \vee (w \wedge y),$$

or, equivalently,

$$(x \wedge w) \vee y = (x \vee y) \wedge (w \vee y).$$

Distributive lattices are modular, but there are modular lattices that are not distributive.

A subset $\Im_0$ of a lattice $\Im$ is called a sublattice of $\Im$ if $x \vee y$ and $x \wedge y$ are in $\Im_0$ whenever both x and y are in $\Im_0$. A sublattice of a modular lattice is modular, and a sublattice of a distributive lattice is distributive.

If x and y are domains in a lattice $\Im$, and

$$x \wedge y = \varnothing \quad \text{and} \quad x \vee y = \Xi,$$

then we say that y is a complement of x. In general, complements are not unique; a given domain in a lattice may have more than one complement. A lattice is said to be complemented if every domain has at least one complement.

A complemented distributive lattice with a unit and a zero is called a Boolean algebra. In a Boolean algebra, complements are unique.

The axioms in this paper begin with an arbitrary lattice $\Im$. They do not assume modularity, distributivity, or the existence of complements. All of the examples in the paper, however, involve a Boolean algebra consisting of all subsets of a finite set.

The Boolean algebra of all subsets of a finite set. If we take $\Im$ to consist of all the subsets of a set Ξ, and we write $x \le y$ whenever x is a subset of y, then $\Im$ a Boolean algebra. Its unit is Ξ, and its zero is the empty set. We have $x \vee y = x \cup y$ and $x \wedge y = x \cap y$, and the complement of x is the set difference Ξ-x.

More generally, any field of subsets is a Boolean algebra. Recall that a set $\Im$ of subsets of a set Ξ is a field if $A \cup B$, $A \cap B$, Ξ-A, and Ξ-B are in $\Im$ whenever A and B are in $\Im$. It is not necessary that $\Im$ include all the subsets of Ξ.

2.3. Graphs and Trees

In order to be comfortable with hypergraphs and hypertrees, we need to look from a couple different points of view at the more familiar ideas of graph and tree.

2.3.1. Graphs and Trees, Emphasizing Nodes

A graph is a pair (N,E), where N is a finite set and E is a collection of two-element subsets of N. The elements of N are called nodes and the elements of E are called edges. For our purposes, we can assume that every node is in at least one edge. A node that is in only one edge is called a leaf. The other node in that edge is called the bud for the leaf.

Given a subset M of N, set

$$E_M = \{e \varepsilon E | e \ldots M\}$$

The graph (M,E_M) is called the restriction of (N,E) to M.

Suppose $n_1,n_2,\ldots,n_k$ is an ordering of the nodes of a graph (N,E), and suppose that for i=2,...,k the node n_i is a leaf in the restriction of (N,E) to $\{n_1,n_2,\ldots,n_i\}$. In this case, we say that $n_1,n_2,\ldots,n_k$ is a tree construction ordering for the graph (N,E). A tree is a graph for which a tree construction ordering exists. In order to understand this definition, you should imagine literally constructing the tree by beginning with n_1, then adding n_2 and an edge between it and n_1, then adding n_3 and an edge between either it and either n_1 or n_2, etc. As you add a node, you always connect it to only one of the nodes already there; this is the content of the condition that it be a leaf in the restriction of the graph to the node set consisting of it and the nodes already there.

It is common to define a tree as a connected graph with no cycles, but it can easily be shown that the definition given here is equivalent (see, e.g., Golumbic 1980).

It is sometimes useful to pay attention to which nodes we connect to as we build a tree up following a tree construction ordering. Given a tree construction ordering $n_1,n_2,\ldots,n_k$ and an integer i between 2 and k, let b(i) be the integer between 1 and i-1 such that $h_{b(i)}$ is the bud for n_i in $\{n_1,n_2,\ldots,n_i\}$. We call the function b(i) the budding for the tree construction ordering.

The following lemma shows a more succinct way to introduce the ideas of tree construction ordering and budding.

> **Lemma 2.2.** A graph (N,E) is a tree if and only if there exists an ordering $n_1,n_2,\ldots,n_k$ of the elements of N, and there exist integers b(i), for $2 \leq i \leq k$, such that $1 \leq b(i) \leq i-1$ and $E=\{\{n_2,n_{b(2)}\},\ldots,\{n_k,n_{b(k)}\}\}$. Any such ordering is a tree construction ordering for the graph and the function b(i) is the budding for this tree construction ordering.

The proof is left to the reader.

2.3.2. *Graphs and Trees, Emphasizing Edges*

In order to pave the way for the generalization to hypergraphs and hypertrees, let us retrace, in a slightly different way, the road we have just travelled. This time, we emphasize edges.

To begin, we say simply that a graph is a finite set consisting of two-element sets. We call the two-element sets edges. We call the elements of the edges *nodes*. This is not different in substance from our first definition of graph. Since we are not considering graphs with isolated nodes, every node is in an edge, and so the nodes are specified as soon as the edges are specified.

Continuing this emphasis on edges, we now formulate edge ideas to replace the node ideas of leaf and bud.

Suppose t and b are distinct edges in a graph E. Let us say that t is a *twig*, and b is a branch for b, if exactly one of t's elements is in the rest of E, and this element is also in b. (This is the same as saying that one of t's elements is a leaf and the other, which is the leaf's bud, is in b. But we are now going to replace our talk about leaf and bud with talk about twig and branch.) Notice that a twig may have many branches (though a leaf has only one bud).

The next lemma gives an alternative way of defining twig and branch.

Lemma 2.3. If t and b are distinct edges in a graph E, then t is a twig and b is a branch for t if and only if

$$t \cap b \neq \varnothing \text{ and } t \cap (\cup(E-\{t\})) = t \cap b.$$

Proof: Since t and b are distinct, b must contain exactly one or none of t's elements. So $t \cap b \neq \varnothing$ is equivalent to the statement that it contains exactly one. And $t \cap (\cup(E-\{t\})) = t \cap b$ is equivalent to the statement that the other one is not in $E-\{t\}$.

How should we define the idea of tree now? In our first pass, we defined a tree construction ordering as an ordering of nodes, but now we want to define it as an ordering of edges. It is easy to make this change, because when we added a node, we always connected it to a node already there, and thus we were really adding an edge.

Formally, let us say that an ordering $e_1, e_2, \ldots, e_k$ of the elements of a graph E is a tree construction ordering in edges for E if for $i=2,\ldots,k$ the edge e_i is a twig in the graph $\{e_1, e_2, \ldots, e_i\}$. A *tree* is a graph for which a tree construction ordering in edges exists. Since every twig has at least one branch, we can choose, for i from 2 to k, an integer b(i) such that $1 \leq b(i) \leq i-1$ and $e_{b(i)}$ is a branch for e_i in $\{e_1, e_2, \ldots, e_i\}$. Let us call such a function b(i) a branching for E and $e_1, e_2, \ldots e_k$.

The number of nodes in a tree is one greater than the number of edges. The edge that we put down first in a ordering of edges contains two nodes, and the later edges each add one more node. This complicates the formal description of the correspondence between construction orderings in nodes and construction orderings in edges. Here is how it goes.

Lemma 2.4.

(i) Suppose $n_1,n_2,\ldots,n_k$ is a tree construction ordering in nodes, with budding b(i). Then $\{n_2,n_{b(2)}\},\ldots,\{n_k,n_{b(k)}\}$ is a tree construction ordering in edges (cf. Lemma 2.2). If we write $e_1,e_2,\ldots,e_{k-1}$ for this ordering, then $e_{b(i)1}$ is one branch for e_i.

(ii) Suppose $e_1,e_2,\ldots,e_k$ is a tree construction ordering in edges, and suppose b(i) is a branching for it. Suppose $e_i=\{p_i,q_i\}$ for i=1,..,k. Suppose further, for i=2,..,k, that p_i is the element of e_i that is not in any of the earlier e_j, while q_i is the element of e_i that is in $e_{b(i)}$ and perhaps in other e_j before e_i as well. Then both $p_1,q_1,p_2,\ldots,p_k$ and $q_1,p_1,p_2,\ldots,p_k$ are tree construction orderings in nodes. The bud for q_1 in the first of the ordering $p_1,q_1,p_2,\ldots,p_k$ is p_1. The bud for p_1 in the ordering $q_1,p_1,p_2,\ldots,p_k$ is q_1. The bud for a later p_i in either ordering is q_i.

Before leaving the topic of trees, we should contrast the roles played by the two conditions in Lemma 2.3. The condition $t\cap b\neq\varnothing$ implies that a twig is connected with the rest of the graph, and this, in turn implies that any graph built up by adding a sequence of twigs is connected. The condition $t\cap(\cup(E-\{t\})) = t\cap b$, on the other hand, implies that a twig is connected only through its branch, and this in turn implies that any graph built up by adding a sequence of twigs is acyclic. If we drop the condition $t\cap b\neq\varnothing$ from the definition of twig, then a graph we build up by adding a sequence of twigs will still be acyclic, but it might not be connected; instead of being a single tree, it might be a forest—a collection of unconnected trees.

2.4. Hypergraphs and Hypertrees

Now we generalize the preceding definitions of graph, twig, branch, and tree by dropping the requirement that the edges have only two elements. Now we allow the edges to be any finite sets. In fact, we go even further, and allow them to be domains in any lattice.

Formally, if H is finite subset of a lattice $\mathfrak{I}$, then we say that H is a hypergraph on $\mathfrak{I}$. We say that a domain t in a hypergraph H is a twig if there exists another domain b in H, distinct from t, such that $t\wedge(\vee(H-\{t\})) = t\wedge b$. We call any such b a branch for the twig t. A hypergraph is a hypertree if its domains can be ordered, say $h_1,h_2,\ldots,h_n$, so that h_i is a twig in $\{h_1,h_2,\ldots,h_i\}$, for i=2...,n. We call any such ordering a hypertree construction ordering for H. Given a hypertree construction ordering $h_1,h_2,\ldots,h_n$, we can choose, for i from 2 to n, an integer b(i) such that $1\leq b(i)\leq i-1$ and $h_{b(i)}$ is a branch for h_i in $\{h_1,h_2,\ldots,h_i\}$. We call a function b(i) satisfying this condition a branching for H and $h_1,h_2,\ldots h_n$.

In the sequel, we will write H^{-t} for the hypergraph H-{t}.

Notice that we have not required in the definition of twig that $t \wedge b \neq \emptyset$. This means that hypertrees are allowed to be disconnected. This is an argument against using the name "hypertree." Perhaps we should say "hyperforest" instead, or perhaps we should follow the lead of Beeri et al. (1983) and use the more awkward term "acyclic hypergraph." But since the prefix "hyper-" already suggests generalization, and since we have formulated the idea in the context of a general lattice that might not even have a zero, "hypertree" seems reasonable.

Suppose H is a hypergraph on a lattice $\mathfrak{I}$. A hypertree K on $\mathfrak{I}$ is called a hypertree cover for H on $\mathfrak{I}$ if for every domain h of H, there is a domain k(h) of K such that $h \leq k(h)$. In general, a hypergraph H has many hypertree covers. The simplest is the hypertree consisting of the single domain $\vee H$. The computational methods discussed in this paper work directly if we begin with a hypertree, but if we begin with a hypergraph H that is not a hypertree, we must replace H with a hypertree cover. Since the cost of the computation increases primarily with the size of the largest domain in the hypertree cover (and not so dramatically with the number of domains), $\vee H$ is usually not an attractive hypertree cover. We prefer instead a hypertree cover whose largest domains are as small as possible. Finding such a hypertree cover may itself be a difficult computational problem, however (Kong 1986).

3. Marginalization

In this chapter, we study a set of axioms that justifies the inward phase of hypertree computation in problems in which all potentials have unique reductions, which we therefore call marginals. If these axioms are satisfied, we are given a potential φ that factors on a hypertree, and it is feasible to find marginals in the domains of the hypertree, then we can move backwards in any construction ordering for the hypertree in order to find the marginal of φ on the root of the construction ordering.

Since any domain in a hypertree is the root of some construction ordering for the hypertree, the algorithm of this chapter allows us to find the marginal on any domain of the hypertree. By applying the algorithm repeatedly to different hypertree construction orderings, we can find marginals on all the domains of the hypertree. Shafer and Shenoy (1988) show how the duplication involved in such repeated application of the algorithm can be minimized, so that marginals on all the domains can be found with only two or three times as much work as marginals on a single domain. As we will see in Chapter 5, it is also possible, in some cases, to use the outward phase of hypertree computation to find marginals for the remaining domains once the inward phase has produced the marginal for one domain. This is often more efficient than the Shafer-Shenoy method.

The axioms that we use in this chapter are listed in Section 3.1, and the computational theory based on these axioms is developed in Section 3.3. Section 3.2 is concerned with vacuous extension. This concept is not needed for understanding computation in a hypertree, but it helps us understand how to extend the theory to the case where a potential factors on a hypergraph with a hypertree cover. Section 3.4

applies the computational theory to five examples: dynamic programming, probability, linear equations, constraint propagation, and belief functions. Section 3.5 comments on some ways of strengthening the axioms.

3.1. Axioms

Suppose $\Im$ is a lattice and Φ is a commutative semigroup. As in Chapter 2, we call the elements of $\Im$ domains and the elements of Φ potentials.

We label each potential φ in Φ with a domain in $\Im$, say $d(\varphi)$. When $d(\varphi) = x$, we say that x is the domain of φ and that φ is a potential on x. We make the following assumption about these labels:

Axiom A1. If φ_x and φ_y are potentials on x and y, respectively, then $d(\varphi_x \otimes \varphi_y) = x \vee y$.

This is the labelling axiom. It can be alternatively expressed by saying that the map d is a semigroup homomorophism from Φ to $\Im$, regarded as a semigroup with $\vee$ as its semigroup operation.

We set $\Phi_x = \{\varphi \varepsilon \Phi | d(\varphi) = x\}$. The labelling axiom implies that Φ_x itself is a commutative semigroup under the operation $\otimes$. We assume that this semigroup itself has an identity, say ι_x. And we adopt an axiom relating these identities:

Axiom A2. If $x \leq y$, then $\iota_x \otimes \iota_y = \iota_y$.

This is the identity axiom.

If $x \leq d(\varphi)$, then we associate with φ and x a potential with domain x, which we call the marginal of φ on x. We write $\varphi^{\downarrow x}$ for the marginal of φ on x. We adopt two axioms concerning marginalization:

Axiom M1. If φ_z is a potential on z, and $x \leq y \leq z$, then $(\varphi_z^{\downarrow y})^{\downarrow x} = \varphi_z^{\downarrow x}$.

Axiom M2. If φ_x and φ_y are potentials on x and y, respectively, then $(\varphi_x \otimes \varphi_y)^{\downarrow x} = \varphi_x \otimes \varphi_y^{\downarrow x \wedge y}$.

Axiom M1 is the transitivity axiom; it says that if φ_1 is the marginal of φ_2 and φ_2 is the marginal of φ_3, then φ_1 is the marginal of φ_3. Axiom M2 is the combination axiom; it says $(\varphi_x \otimes \varphi_y)^{\downarrow x}$ can be found by combining φ_x with another potential.

Axiom A1 is needed in the formulation of axioms A2 and M2. On the other hand, there is a certain independence between A2, which does not involve marginalization, and M1 and M2, which do not involve identities. Some of the consequences that we will derive from these axioms in this chapter involve only combination and marginalization and hence depend only on Axioms A1, M1, and M2. Others involve only combination and the existence of identities and hence depend only on Axioms A1 and A2.

The following lemma is not needed in this chapter, but it will be used in Chapter 5.

Lemma 3.1. If φ_x and φ_y are potentials on x and y, respectively, then $(\varphi_x \otimes \varphi_y)^{\downarrow x \wedge y} = \varphi_x^{\downarrow x \wedge y} \otimes \varphi_y^{\downarrow x \wedge y}$.

Proof: Using Axiom A2 once and Axiom M1 twice, we obtain $(\varphi_x \otimes \varphi_y)^{\downarrow x \wedge y} = ((\varphi_x \otimes \varphi_y)^{\downarrow x})^{\downarrow x \wedge y} = (\varphi_x \otimes \varphi_y^{\downarrow x \wedge y})^{\downarrow x \wedge y} = \varphi_x^{\downarrow x \wedge y} \otimes \varphi_y^{\downarrow x \wedge y}$.

3.2. Vacuous Extension

If φ_x is a potential on x, and $x \le y$, then we write $\varphi_x{\uparrow}^y$ for the product $\varphi_x \otimes \iota_y$, and we call $\varphi_x{\uparrow}^y$ the *vacuous extension* of φ_x to y.

Lemma 3.2. Suppose Axioms A1 and A2 hold. Then the following statements hold as well.

(i) If $x \le y$, then $\iota_x{\uparrow}^y = \iota_y$.

(ii) For any domains x and y, $\iota_x \otimes \iota_y = \iota_{x \vee y}$.

(iii) If φ_x is a potential on x, then $\varphi_x \otimes \iota_y = \varphi_x{\uparrow}^{x \vee y}$.

(iv) If φ_x is a potential on x, then $\varphi_x{\uparrow}^x = \varphi_x$.

(v) If φ_x is a potential on x, and $y \le x$, then $\varphi_x \otimes \iota_y = \varphi_x$.

(vi) If φ_x is a potential on x, and $x \le y \le z$, then $(\varphi_x{\uparrow}^y){\uparrow}^z = \varphi_x{\uparrow}^z$.

(vii) If φ_x and φ_y are potentials on x and y, respectively, $x \le z$, and $y \le z$, then $(\varphi_x \otimes \varphi_y){\uparrow}^z = \varphi_x{\uparrow}^z \otimes \varphi_y{\uparrow}^z$.

(viii) If φ_x and φ_y are potentials on x and y, respectively, then $\varphi_x \otimes \varphi_y = (\varphi_x{\uparrow}^{x \vee y}) \otimes (\varphi_y{\uparrow}^{x \vee y})$.

Proof:

(i) By definition, $\iota_x{\uparrow}^y = \iota_x \otimes \iota_y$. By the identity axiom, $\iota_x \otimes \iota_y = \iota_y$.

(ii) Since $x \le x \vee y$ and $y \le x \vee y$, we have both $\iota_{x \vee y} = \iota_{x \vee y} \otimes \iota_x$ and $\iota_{x \vee y} = \iota_{x \vee y} \otimes \iota_y$. Substituting one of these equations in the other, we obtain $\iota_{x \vee y} = \iota_{x \vee y} \otimes \iota_x \otimes \iota_y$. Since $\iota_{x \vee y}$ is the identity on $x \vee y$, this reduces to $\iota_{x \vee y} = \iota_x \otimes \iota_y$.

(iii) $\varphi_x \otimes \iota_y = (\varphi_x \otimes \iota_x) \otimes \iota_y = \varphi_x \otimes (\iota_x \otimes \iota_y) = \varphi_x \otimes \iota_{x \vee y} = \varphi_x{\uparrow}^{x \vee y}$.

(iv) By the definition of vacuous extension, $\varphi_x{\uparrow}^x = \varphi_x \otimes \iota_x$. Since ι_x is the identity on Φ_x, this is equal to φ_x.

(v) This follows from (iii) and (iv).

(vi) $(\varphi_x{\uparrow}^y){\uparrow}^z = (\varphi_x \otimes \iota_y){\uparrow}^z = \varphi_x \otimes \iota_y \otimes \iota_z = \varphi_x \otimes \iota_z = \varphi_x{\uparrow}^z$.

(vii) $\varphi_x{\uparrow}^z \otimes \varphi_y{\uparrow}^z = \varphi_x \otimes \iota_z \otimes \varphi_y \otimes \iota_z = \varphi_x \otimes \varphi_y \otimes \iota_z = (\varphi_x \otimes \varphi_y){\uparrow}^z$.

(viii) This follows from (vii) and (iv).

Lemma 3.3. Suppose Axioms A1, A2, and M2 hold. Suppose φ_x is a potential on x. Then $(\varphi_x{\uparrow}^{x \vee y}){\downarrow}^y = (\varphi_x{\downarrow}^{x \wedge y}){\uparrow}^y$.

Proof: $(\varphi_x{\uparrow}^{x \vee y}){\downarrow}^y = (\varphi_x \otimes \iota_y){\downarrow}^y = \varphi_x{\downarrow}^{x \wedge y} \otimes \iota_y = (\varphi_x{\downarrow}^{x \wedge y}){\uparrow}^y$.

3.3. Computational Theory

The computational theory developed here depends on Axioms A1-A2 and M1-M2 and also on certain further assumptions about the computational difficulty and feasibility of the operations $\otimes$, $\wedge$, $\vee$, d, and $\downarrow$. When all these assumptions are satisfied, the computational theory accomplishes two things. First, it tells us how to compute marginals of products on certain hypertrees. Second, it tells us how to use hypertree covers to compute marginals of products on certain hypergraphs that are not hypertrees.

3.3.1. *Computational Assumptions*

The axioms themselves do not say anything about the relative difficulty of implementing the operations $\otimes$, $\wedge$, $\vee$, d, and $\downarrow$. The computational theory in which we are interested makes sense, however, only if we make certain further assumptions about their relative difficulty.

We assume that no computational difficulty is involved in finding the labels $d(\varphi)$. This is the point of calling them labels; we imagine that each potential is given to us with its label pasted on it. A function φ on a set of variables x, for example, comes to us labelled as such.

We also assume that there is little computational difficulty involved in the lattice operations. In most cases, these domains are simply sets of variables, and the lattice operations are simply union and intersection. Provided that the total number of variables is not inordinately large, finding the intersection of two sets of these variables, say, will not be a difficult problem.

The computational difficulties lie in implementing the marginalization map and also, sometimes, in implementing the semigroup operation. We assume that as we consider potentials with larger domains, finding their marginals becomes increasingly difficult. Finding their products may also become increasingly difficult. It may be difficult even to represent a potential on a large domain in any explicit way.

Let us call a domain x *feasible* if it is feasible to represent potentials on x adequately, combine them, and marginalize them. (This is not a precise definition, for what should count as an adequate representation can sometimes be debated, and the boundary between the feasible and the non-feasible is not sharp.) We assume that any subdomain of a feasible domain is also feasible.

We also assume that computational difficulties increase with the size of the domain so sharply that the number of domains with which we must deal is much less important to the feasibility of the computation than the size of the largest of these domains. Thus we call a hypertree feasible if each of its domains is feasible, regardless of the number of these domains. We call a hypergraph feasible if it has a feasible hypertree cover.

3.3.2. *Finding the Marginal of a Product on a Hypertree*

Suppose we are given a collection $\{\varphi_h\}_{h\in H}$ of potentials, where H is a hypertree on $\mathfrak{X}$, and φ_h is a potential on h. By the labelling axiom, its product, $\otimes\{\varphi_h | h\in H\}$, is a potential on $\vee H$. Suppose we want to find the marginal $(\otimes\{\varphi_h | h\in H\})^{\downarrow x}$. As we will now see, we can do this provided Axioms A1, M1, and M2 hold (Axiom A2 is not needed), our computational assumptions hold, the hypertree H is feasible, and x is a subdomain of some domain in H.

Let us begin with a something simpler—the computational significance of Axiom M2, the combination axiom. Suppose x and y are both feasible, but $x\vee y$ is not. The combination axiom says that we can compute $(\varphi_x\otimes\varphi_y)^{\downarrow x}$ in these circumstances even if we cannot even explicitly express the product $\varphi_x\otimes\varphi_y$. We first compute $\varphi_y^{\downarrow y\wedge x}$ (this only requires us to work in y), and then we combine it with φ_x (this only requires us to work in x).

Combining the transitivity and combination lemmas, we obtain the following lemma.

Lemma 3.4. Suppose $z \leq x$, and suppose φ_x and φ_y are potentials on x and y, respectively. Then $(\varphi_x \otimes \varphi_y)^{\downarrow z} = (\varphi_x \otimes \varphi_y^{\downarrow y \wedge x})^{\downarrow z}$.

Proof: $(\varphi_x \otimes \varphi_y^{\downarrow y \wedge x})^{\downarrow z} = ((\varphi_x \otimes \varphi_y)^{\downarrow x})^{\downarrow z}$ by combination

$= (\varphi_x \otimes \varphi_y)^{\downarrow z}$. by transitivity

Again, the computational significance is clear. If the domains x and y are feasible, and $z \leq x$, then we can compute $(\varphi_x \otimes \varphi_y)^{\downarrow z}$.

The next lemma concerns a product on a hypergraph with a twig. (Recall that H^{-t} is a compact way of writing H-{t}.)

Lemma 3.5. Suppose $\{\varphi_h\}_{h \varepsilon H}$ is a collection of potentials on the hypergraph H, t is a twig in H, and b is a branch for t. Then

(i) $(\otimes\{\varphi_h | h \varepsilon H\})^{\downarrow \vee H^{-t}} = (\otimes\{\varphi_h | h \varepsilon H^{-t}\}) \otimes \varphi_t^{\downarrow t \wedge b}$.

If we set $\varphi_b^{-t} = \varphi_b \otimes \varphi_t^{\downarrow t \wedge b}$ and $\varphi_h^{-t} = \varphi_h$ for all other h in H^{-t}, then

(ii) $(\otimes\{\varphi_h | h \varepsilon H\})^{\downarrow \vee H^{-t}} = \otimes\{\varphi_h^{-t} | h \varepsilon H^{-t}\}$, and

(iii) if $x \leq \vee H^{-t}$, then $(\otimes\{\varphi_h | h \varepsilon H\})^{\downarrow x} = (\otimes\{\varphi_h^{-t} | h \varepsilon H^{-t}\})^{\downarrow x}$.

Proof: By the labelling axiom, $\otimes\{\varphi_h^{-t} | h \varepsilon H^{-t}\}$ is a potential on $\vee H^{-t}$. So (i) follows directly from the combination axiom, with $\vee H^{-t}$ for x, $\otimes\{\varphi_h | h \varepsilon H^{-t}\}$ for φ_x, t for y, and φ_t for φ_y. To prove (ii), we merely rearrange the factors on the right-hand side of (i). Then we get (iii) by marginalizing both sides of (ii) to x and applying the transitivity axiom.

If the twig t and its branch b are feasible, then it is feasible to compute φ_b^{-t}, and statements (ii) and (iii) both have computational significance. Statement (ii) says that the task of marginalizing a product on H to a subdomain x of $\vee H^{-t}$ can be reduced to the smaller task of marginalizing a product on $\vee H^{-t}$ to x. Statement (iii) says that we can construct (i.e., compute the factors of) a product $\otimes\{\varphi_h^{-t} | h \varepsilon H^{-t}\}$ that is the marginal on $\vee H^{-t}$ of our original product $\otimes\{\varphi_h | h \varepsilon H\}$ on H. Of the two, statement (iii) is more immediately useful, but statement (ii) has the deeper significance, for it opens the way to the successive removal of twigs when H is a hypertree.

Suppose, indeed, that H is a feasible hypertree, and we want to compute $\varphi^{\downarrow x}$, where $\varphi = \otimes\{\varphi_h | h \varepsilon H\}$, and x is a subdomain of some domain in H, say h_1. Choose a hypertree construction ordering for H that begins with h_1, say $h_1, h_2, \ldots, h_n$, and choose a branching b(i) for this construction ordering. For i = 1,2,...,n, set

$$H^i = \{h_1, h_2, \ldots, h_i\}.$$

This is a sequence of hypertrees, each larger than the last; $H^1 = \{h_1\}$ and $H^n = H$. The domain h_i is a twig in H^i. So we can work backwards in this sequence, using the idea of Lemma 3.5 each time. At the step from H^i to H^{i-1}, we go from a collection $\{\varphi_h^i | h \varepsilon H^i\}$, say, which has $\varphi^{\downarrow \vee H^i}$ as its product, to a collection $\{\varphi_h^{i-1} | h \varepsilon H^{i-1}\}$, which has $\varphi^{\downarrow \vee H^{i-1}}$ its product. To go from $\{\varphi_h^i | h \varepsilon H^i\}$ to $\{\varphi_h^{i-1} | h \varepsilon H^{i-1}\}$, we omit h_i and change the potential on $h_{b(i)}$ from $\varphi_{h_{b(i)}}^i$ to

$$\varphi_{h_{b(i)}}{}^{i} \otimes (\varphi_{h_i}{}^{i})^{\downarrow h_{b(i)} \wedge h_i};$$

the other potentials are unchanged. The collection with which we begin, $\{\varphi_h{}^n | h \varepsilon H^n\}$, is simply $\{\varphi_h | h \varepsilon H\}$, and the collection with which we end, $\{\varphi_h{}^1 | h \varepsilon H^1\}$, consists of the single potential $\varphi^{\downarrow h_1}$. One more marginalization within h_1 reduces this to what we want, $\varphi^{\downarrow x}$.

Since the computation we have just described moves backwards in the construction ordering, from each domain h_i to its branch $h_{b(i)}$, we think of it as moving inward in the hypertree.

Notice that we have not appealed, so far, to the existence of identities, the idea of vacuous extension, or Axiom A2. We have only used Axioms A1, M1, and M2 and our computational assumptions.

3.3.3. Using Hypertree Covers

Now we replace the assumption that H is a feasible hypertree and x is a subdomain of some domain in x with the weaker assumption that $H \cup \{x\}$ is a feasible hypergraph—i.e., that $H \cup \{x\}$ has a feasible hypertree cover. We show that we can still compute $(\otimes\{\varphi_h | h \varepsilon H\})^{\downarrow x}$ in this case. Here we do need the existence of identities, the idea of vacuous extension, and Axiom A2.

Suppose K is a feasible hypertree cover for $H \cup \{x\}$. We may assume that $vK = v(H \cup \{x\})$. (Otherwise, we can get a feasible hypertree cover satisfying this condition by replacing each domain k in K by $k \wedge (v(H \cup \{x\}))$.) For each h in $H \cup \{x\}$, let k(h) be a domain in K such that $k(h) \leq h$. Construct a collection $\{\phi_k\}_{k \varepsilon K}$ on K from the collection $\{\varphi_h\}_{h \varepsilon H}$ on H by setting

$$\phi_k = \begin{cases} \otimes\{\varphi_h{}^{\uparrow k} | h \varepsilon H, k(h)=k\} & \text{if } \{\varphi_h | h \varepsilon H, k(h)=k\} \neq \varnothing \\ \iota_k & \text{if } \{\varphi_h | h \varepsilon H, k(h)=k\} = \varnothing . \end{cases}$$

Then

$$\begin{aligned} \otimes\{\phi_k | k \varepsilon K\} &= (\otimes\{\varphi_h{}^{\uparrow k(h)} | h \varepsilon H\}) \otimes (\otimes\{\iota_k | k \varepsilon K, k(h) \neq k \text{ for any h in H}\}) \\ &= (\otimes\{\varphi_h | h \varepsilon H\}) \otimes (\otimes\{\iota_{k(h)} | h \varepsilon H\}) \otimes (\otimes\{\iota_k | k \varepsilon K, k(h) \neq k \text{ for any h in H}\}) \\ &= \otimes\{\varphi_h | h \varepsilon H\}, \end{aligned}$$

by the definition of vacuous extension and statement (v) of Lemma 3.2. Hence

$$(\otimes\{\phi_k | k \varepsilon K\})^{\downarrow x} = (\otimes\{\varphi_h | h \varepsilon H\})^{\downarrow x}.$$

But we can compute $(\otimes\{\phi_k | k \varepsilon K\})^{\downarrow x}$, because it falls under the rubric of the preceding section; K is a feasible hypertree, and x is a subdomain of a domain of K.

3.3.4. Conclusion

We conclude that $(\otimes\{\varphi_h | h \varepsilon H\})^{\downarrow x}$ can be found whenever $H \cup \{x\}$ is a feasible hypergraph—i.e., is a feasible hypertree or has a feasible hypertree cover.

3.4. Examples

This section examines five problems in which our computational theory is instantiated. The problems vary in how potentials, combination, and marginalization are defined, and in what counts in practice as having computed a potential. In some cases, a potential is a function, and writing down a formula or an algorithm that allows us to compute its value at any configuration of its domain counts as having computed it. In other cases, a potential is a function, but we have to be in a position to sum it over arbitrary sets of configurations of its domain before we can say that we have computed it. In other cases, a potential is an equivalence class, and writing down one element of the equivalence class counts as having computed it.

In the problems discussed here, the lattice $\mathfrak{I}$ is the set of all subsets of a set of variables Ξ. Thus it has Ξ as its unit and the empty set as its zero. We write Θ_ξ for the frame (or state space or sample space) of the variable ξ. If we say that ξ is real-valued, we mean that Θ_ξ is the real line. In some other cases, we say only that Θ_ξ is finite. We call an element of Θ_ξ a configuration of ξ. If x is a domain, then we write Θ_x for the Cartesian product of the frames of the variables in x:

$$\Theta_x = \times_{\xi \varepsilon x} \Theta_\xi. \tag{3.1}$$

We call this the frame for x, and we call its elements configurations of x.

If x is a subdomain of y and c is a configuration for y, we write $c^{\downarrow x}$ for the configuration of x obtained by dropping the values of the variables in y and not in x. If, for example, $x = \{\xi_1,\xi_2\}$, $y = \{\xi_1,\xi_2,\xi_3,\xi_4\}$, and $c = (c_1,c_2,c_3,c_4)$, where $c_i \varepsilon \Theta_{\xi_i}$, then $c^{\downarrow x} = (c_1,c_2)$.

If x and y are disjoint domains, c_x is a configuration of x, and c_y is a configuration of y, then we write $(c_x.c_y)$ for the configuration of $x \vee y$ obtained by concatenating c_x and c_y. In other words, $(c_x.c_y)$ is the unique configuration of such that $(c_x.c_y)^{\downarrow x} = c_x$ and $(c_x.c_y)^{\downarrow y} = c_y$.

The problems discussed here are not the only ones that satisfy the axioms. Other problems of note that satisfy these axioms are discussed in Dawid (1991), Cowell and Dawid (1991), and Cowell (1991). Of particular interest is the problem of fast retraction, which involves the probability example, but which uses a marginalization that differs from the usual marginalization for probability but still satisfies the axioms.

3.4.1. Dynamic Programming

Suppose Ξ is a finite set of variables, each with a finite frame, and let $\mathfrak{I}$ be the lattice of all subsets of Ξ. Let us call any real-valued function on a subset x of Ξ a potential on x. A potential on $\varnothing$ is a constant. We combine potentials by adding them. In other words, $\varphi_1 \otimes \varphi_2$ is the potential with domain $d(\varphi_1) \vee d(\varphi_2)$ given by

$$(\varphi_1 \otimes \varphi_2)(c) = \varphi_1(c^{\downarrow d(\varphi_1)}) + \varphi_2(c^{\downarrow d(\varphi_2)})$$

for every configuration c in $d(\varphi_1) \vee d(\varphi_2)$. Thus Axiom A1 is satisfied by definition.

If $x \le d(\varphi)$, then $\varphi^{\downarrow x}$ is the potential on x obtained from φ by maximizing over the variables in $d(\varphi)$-x. In other words, for each configuration c of x,

$$\varphi^{\downarrow x}(c) = \max\{\varphi(c.c') \mid c' \text{ is a configuration of } d(\varphi)\text{-}x\}.$$

If $x = d(\varphi)$, so that $d(\varphi)-x = \varnothing$, then this reduces to $\varphi^{\downarrow x}(c) = \varphi(c)$. Axiom M1 holds because the order in which we maximize over variables does not matter.

Axiom M2 holds because addition of a constant distributes over maximization. Suppose, indeed, that φ_1 is a potential on x and φ_2 is a potential on y. Then for every configuration c of x,

$$\begin{aligned}(\varphi_1+\varphi_2)^{\downarrow x}(c) &= \max\{(\varphi_1+\varphi_2)(c.c') \mid c' \text{ is a configuration of } y\text{-}x\}\\ &= \max\{\varphi_1((c.c')^{\downarrow x}) + \varphi_2((c.c')^{\downarrow y}) \mid c' \text{ is a configuration of } y\text{-}x\}\\ &= \max\{\varphi_1(c^{\downarrow x}) + \varphi_2(c^{\downarrow y\wedge x}.c') \mid c' \text{ is a configuration of } y\text{-}x\}\\ &= \varphi_1(c^{\downarrow x}) + \max\{\varphi_2(c^{\downarrow y\wedge x}.c') \mid c' \text{ is a configuration of } y\text{-}x\}\\ &= \varphi_1(c^{\downarrow x}) + \varphi_2^{\downarrow y\wedge x}(c^{\downarrow y\wedge x})\\ &= (\varphi_1+\varphi_2^{\downarrow y\wedge x})(c).\end{aligned}$$

Finally, the commutative semigroup consisting of all real-valued functions on x, with addition as the semigroup operation, does indeed have an identity—the function on x that is identically equal to zero. Axiom A2 obviously holds.

Notice that the axioms would still hold if we substituted min for max, or if we required the potentials to be non-negative and substituted multiplication for addition. In the latter case, the identity on x would be the function identically equal to one.

Computational difficulties arise in this example when the frame Θ_x becomes too large. If the frame Θ_ξ for each variable ξ is large, and there are many variables in x, then Θ_x will be huge. So representing a potential φ on x may be impractical if it means writing down the real number that φ assigns to each configuration of x. And marginalizing φ to y may be impractical if it requires that for each configuration c of x we sort through all the configurations c' of y-x to see which $\varphi(c.c')$ is largest.

So feasibility is a matter of the size of Θ_x. We say that x is feasible if Θ_x is small enough for it to be feasible to list the elements of Θ_x and search through this list to perform the maximizations we require.

What we can accomplish using our computational theory in this example is to find the maximum of a function that is defined by adding functions on a feasible hypergraph. Indeed, if we begin with a collection $\{\varphi_h \mid h\varepsilon H\}$ of potentials on a feasible hypergraph H, and we set

$$\varphi = \Sigma\{\varphi_h \mid h\varepsilon H\},$$

then

$$\max\{\varphi(c) \mid c \text{ is a configuration of } \vee H\}$$

is the same as $(\otimes\{\varphi_h \mid h\varepsilon H\})^{\downarrow\emptyset}$. Our computational theory allows us to find $(\otimes\{\varphi_h \mid h\varepsilon H\})^{\downarrow\emptyset}$, because if H is a feasible hypertree, then $H\cup\{\emptyset\}$ is as well.

Usually, of course, we are interested not only in the maximum value of φ but also in a configuration c at which φ takes this maximum. As we will see in Chapter 6, we can find such a configuration using a second, outward sequence of computations.

3.4.2. *Marginals of Factored Probability Distributions*

This computational problem is very similar to our dynamic programming problem. It differs only in that we multiply instead of adding to combine potentials and we sum out instead of maximizing out to marginalize.

Again, Ξ is a finite set of variables, each with a finite frame, $\mathfrak{I}$ is the lattice of all subsets of Ξ, a potential on a non-empty subset x of Ξ is a real-valued function on x, and a potential on $\varnothing$ is a constant. But now the potential $\varphi_1 \otimes \varphi_2$ is the potential with domain $d(\varphi_1) \vee d(\varphi_2)$ given by

$$(\varphi_1 \otimes \varphi_2)(c) = \varphi_1(c^{\downarrow d(\varphi_1)}) \cdot \varphi_2(c^{\downarrow d(\varphi_2)})$$

for every configuration c in $d(\varphi_1) \vee d(\varphi_2)$. And if $x \leq d(\varphi)$, then $\varphi^{\downarrow x}$ is the potential on x given by

$$\varphi^{\downarrow x}(c) = \Sigma \{\varphi(c.c') \mid c' \text{ is a configuration of } d(\varphi)-x\}.$$

for each configuration c of x. By the same reasoning as in the dynamic programming problem, we find that Axioms A1, M1, and M2 hold. The identity is now the function on x that is identically equal to one, and hence Axiom A2 also holds. The computational difficulties are the same as in the dynamic programming problem, except that we are adding values instead of sorting through them to find the largest. We say that x is feasible if Θ_x is small enough for this to be feasible.

We can use our computational theory to find marginals of a product $(\Pi\{\varphi_h \mid h \varepsilon H\})^{\downarrow x}$ when $H \cup \{x\}$ is a feasible hypergraph. We can also find the constant $(\Pi\{\varphi_h \mid h \varepsilon H\})^{\downarrow \varnothing}$, which is the sum of the values of $\Pi\{\varphi_h \mid h \varepsilon H\}$ over all configurations of $\vee H$. This is most interesting in the case where the φ_h are all non-negative and their product is not identically zero, for then

$$P = ((\Pi\{\varphi_h \mid h \varepsilon H\})^{\downarrow \varnothing})^{-1} \cdot (\Pi\{\varphi_h \mid h \varepsilon H\}),$$

is a probability distribution on $\vee H$, and its marginal on x is given by

$$P^{\downarrow x} = ((\Pi\{\varphi_h \mid h \varepsilon H\})^{\downarrow \varnothing})^{-1} \cdot (\Pi\{\varphi_h \mid h \varepsilon H\})^{\downarrow x}. \tag{3.2}$$

Since our computational theory enables us to compute both marginals on the right-hand side of (3.2), it enables us to compute the marginal $P^{\downarrow x}$.

This example is more satisfying than the preceding one, because finding the marginal of a probability distribution is a reasonable goal in itself. There is not necessarily something more that has to be done, such as finding an optimal configuration.

3.4.3. *Sparse Systems of Linear Equations*

Suppose Ξ is a finite set of real variables, and $\mathfrak{I}$ is the lattice of all subsets of Ξ.

A system of linear equations is a pair (A,x), where x is a subset of Ξ, and A is a finite non-empty set of equations of the form

$$\sum_{\xi \varepsilon w} a_\xi \xi + a = 0,$$

where $w \leq x$ and a and the a_ξ are real numbers. (If $x = \varnothing$, then A can contain only equations of the form a = 0.) The set A may include redundant or contradictory equations. We say that (A,x) is a system on x, and we say that x is its label. If a configuration c of x satisfies all the equations in A, then we call c a solution of (A,x).

We say that two systems on x are equivalent if they have the same solutions. This is an equivalence relation. We call the equivalence classes potentials on x, and we call a system in a potential φ a representative of φ. If a configuration c of x is a solution of the representatives of a potential φ on x, we say that c is a solution of φ. In practice, computing a potential means writing down a representative of the potential.

The case where $x = \varnothing$ requires special comment. We divide the systems $(A,\varnothing)$ into two potentials—the potential ♣ consisting of those in which A contains false equations (such as $5 = 0$), and the potential ♠ consisting of those in which A contains the single equation $0 = 0$.

We combine potentials by taking unions of representatives. If φ_x and φ_y are potentials on x and y respectively, $(A,x)\varepsilon\varphi_x$, and $(B,y)\varepsilon\varphi_y$, then $\varphi_x\otimes\varphi_y$ is the potential that contains $(A\cup B, x\vee y)$. This potential is the same no matter which representatives we choose. Since $\cup$ and $\vee$ are commutative and associative, $\otimes$ is as well, and Axiom A1 holds by definition. The identity ι_x is the potential that contains the system $(\{0 = 0\},x)$, and Axiom A2 is obvious.

In order to marginalize a potential φ on y to a subdomain x of y, we choose a representative (A,y) from φ and successively eliminate from A any variables in y-x that are in equations in A. We do this in the usual way; if ξ is such a variable, we first remove it from any equations in A in which its coefficient is zero. If it remains in any equations in A, we choose one such equation, solve it for ξ, and substitute the resulting expression for ξ in any other equations that involve ξ. Each time we eliminate a variable, we obtain a system with at least one less equation. The system B we have when we have eliminated all the variables in y-x will depend on our choice of (A,y) from φ and on the choices we make when eliminating variables (the order in which we eliminate them and which equations we solve to eliminate them), but the potential to which (B,x) belongs will not depend on these choices. This potential is $\varphi^{\downarrow x}$. If x is non-empty, then the solutions of $\varphi^{\downarrow x}$ are the configurations of x of the form $c^{\downarrow x}$, where c is a solution of φ. Moreover, $\varphi^{\downarrow\varnothing}$ = ♠ if φ has solutions, and $\varphi^{\downarrow\varnothing}$ = ♣ if φ has no solutions.

Axioms M1 and M2 follow easily from this description. Axiom M1 is one aspect of the fact that the order in which we eliminate variables does not matter. Axiom M2 holds because when we eliminate variables, we do not change equations in which these variables do not appear. Indeed, suppose φ_x and φ_y are potentials on x and y, respectively, and consider representatives (A,x) of φ_x and (B,y) of φ_y. We obtain a representative of $(\varphi_1\otimes\varphi_2)^{\downarrow x}$ by eliminating the variables in y-x from all the equations in $A\cup B$, and we obtain a representative of $\varphi_1\otimes\varphi_2^{\downarrow y\wedge x}$ by eliminating these variables from the equations in B and adding the result to the equations in A, which comes down to the same thing.

Our computational theory is useful when a system is sparse—when it can be broken into subsystems that deal with relatively disjoint and relatively small clusters of variables. The total number of equations and the total number of variables may be too large to solve the system by eliminating the variables in an arbitrary order. But if the clusters of variables form a feasible hypergraph, then we are dealing with a product φ of potentials on this feasible hypergraph, and the computational theory will enable us to find a representative of $\varphi^{\downarrow h}$ for each cluster h, using only eliminations within the clusters. We may then be able to solve each $\varphi^{\downarrow h}$, thus obtaining a collection of partial solutions. At the very least, we can marginalize one of the $\varphi^{\downarrow h}$ further to $\emptyset$. Then we will know whether $\varphi^{\downarrow\emptyset}$ is ♠ or ♣, and hence whether φ has a solution.

Usually we will not be content to learn whether there is a solution, or to find a collection of partial solutions. We will want at least one complete solution. We can find a complete solution with very little more work than is required to find $\varphi^{\downarrow\emptyset}$, but this task belongs to the outward phase of the computation, which we will study in Chapters 6 and 8.

3.4.4. Subsets of Cartesian Products (Constraint Propagation)

A formal interpretation of our axioms can be obtained by taking potentials to be subsets of Cartesian products, with intersection of cylinder sets for combination and projection for marginalization. This is instructive in two respects. On the one hand, it gives us a way of seeing that our axioms are satisfied by constraint propagation, a practical computational problem in which subsets are defined by systems of constraints, and intersection and projection downward are implemented by manipulating these constraints. On the other hand, as will see next, it is a stepping stone towards understanding how our axioms are satisfied by belief functions.

To be concrete, let Ξ again be a finite set of variables, each with a finite frame, and let $\mathfrak{I}$ again be the lattice of all subsets of Ξ. A potential on x is a subset of Θ_x, labelled with x. We will leave the labelling implicit, except in the case of the empty set, where we will use a subscript. In other words, we distinguish between $\emptyset_x$, the empty subset of Θ_x, and $\emptyset_y$, the empty subset of Θ_y.

We define the semigroup operation by

$$\varphi_x \otimes \varphi_y = \{\, c \in \Theta_{x \vee y} \mid c^{\downarrow x} \in \varphi_x \text{ and } c^{\downarrow y} \in \varphi_y \,\}. \tag{3.3}$$

With this definition, the identity for Φ_x is Θ_x, and Axioms A1 and A2 are obvious. Notice that if φ_1 and φ_2 are potentials on the same domain, then

$$\varphi_1 \otimes \varphi_2 = \varphi_1 \cap \varphi_2. \tag{3.4}$$

We define the marginalization map by

$$\varphi_x^{\downarrow y} = \{\, c^{\downarrow y} \mid c \in \varphi_x \,\},$$

where $y \le x$. This is sometimes projection downward. Axiom M1 is obvious.

To verify Axiom M2, we note that

$$\begin{aligned}(\varphi_x \otimes \varphi_y)^{\downarrow x} &= \{\, c^{\downarrow x} \mid c \in \Theta_{x \vee y},\ c^{\downarrow x} \in \varphi_x, \text{ and } c^{\downarrow y} \in \varphi_y \,\} \\ &= \varphi_x \cap \{\, c^{\downarrow x} \mid c \in \Theta_{x \vee y} \text{ and } c^{\downarrow y} \in \varphi_y \,\}\end{aligned}$$

$= \varphi_x \cap \{ d\varepsilon\Theta_x \mid d{\downarrow}^{x\wedge y} = e{\downarrow}^{x\wedge y} \text{ for some } e\varepsilon\varphi_y \},$

while

$$\varphi_x \otimes \varphi_y{\downarrow}^{x\wedge y} = \{ c \mid c\varepsilon\varphi_x \text{ and } c{\downarrow}^{x\wedge y}\varepsilon\varphi_y{\downarrow}^{x\wedge y} \}$$
$$= \varphi_x \cap \{ d\varepsilon\Theta_x \mid d{\downarrow}^{x\wedge y}\varepsilon\varphi_y{\downarrow}^{x\wedge y} \}$$
$$= \varphi_x \cap \{ d\varepsilon\Theta_x \mid d{\downarrow}^{x\wedge y}= e{\downarrow}^{x\wedge y} \text{ for some } e\varepsilon\varphi_y \}.$$

The vacuous extension of a potential φ_x on x to a larger domain y is the cylinder set in Θ_y corresponding to φ_x. Indeed,

$$\varphi_x{\uparrow}^y = \varphi_x \otimes \iota_y = \{ c\varepsilon\Theta_{x\vee y} \mid c{\downarrow}^x\varepsilon\varphi_x \text{ and } c{\downarrow}^y\varepsilon\iota_y \}$$
$$= \{ c\varepsilon\Theta_{x\vee y} \mid c{\downarrow}^x\varepsilon\varphi_x \}.$$

By (3.4) and statement (vii) of Lemma 3.2,

$$\varphi_x \otimes \varphi_y = (\varphi_x{\uparrow}^{x\vee y}) \cap (\varphi_y{\uparrow}^{x\vee y}).$$

Thus φ_x and φ_y are combined vacuously extending and intersecting.

In the problem of constraint propagation, a subset φ_x of Θ_x is represented by a set of constraints on the variables in x—equations or inequalities relating these variables. With such a representation, combination is simply a matter of pooling two sets of constraints, but marginalization means manipulating the constraints so as to eliminate variables, and this may be computationally challenging if the domain is large. A feasible domain, in this example is one in which such manipulations are practical. Our computational theory is helpful when we pool a collection $\{\varphi_h\}_{h\varepsilon H}$, where φ_h is a set of constraints on the variables in h and H is a feasible hypergraph. We can find the implication of all the constraints for the variables in a particular domain h if we can marginalize within domains.

This problem of constraint propagation generalizes the problem of we studied in Section 3.4.3, where the constraints were all linear equations. Constraint propagation itself can be assimilated formally (though not necessarily computationally) to the problem of dynamic programming that we studied above, with multiplication instead of addition for combining potentials. A subset φ of a set Θ_x can be thought of as a non-negative function on Θ_x—the indicator function that assigns the value one to the elements of φ and the value zero to the other elements of Θ_x. Intersecting two sets corresponds to multiplying their indicator functions, and marginalizing a set to a subdomain corresponds to maximizing out the variables not in the subdomain.

3.4.5. Combining and Marginalizing Belief Functions

Mathematically, a belief function on a set of variables x can be thought of as a probability distribution for a random subset of Θ_x (Shafer 1990). A belief function on x is marginalized to a subdomain y of x by projecting the random subset down to y. A belief function on x is combined with a belief function on y by taking the two random subsets to be independent, vacuously extending both to x∨y, and intersecting them. It follows from this description, and from what we just learned about non-random subsets, that belief functions satisfy Axioms A1-A2 and M1-M2.

In computational practice, there are many ways of representing belief functions. First, there are many ways of representing subsets. A subset of Θ_x might be represented by a set of constraints, an explicit list of elements, or some mixture of the two. Second, there are many ways of representing a probability distribution over subsets.

Let us assume, for concreteness, that a subset is represented by an explicit list of elements. Then the implementation of our computational theory for belief functions might involve three different representations for a given belief function, which we may call the m-function, the b-function, and the q-function. Each of these functions assigns a non-negative number to each non-empty subset of Θ_x.

The m-function is often the initial representation for a belief function. In some applications, at least, a person assesses the strength of certain evidence by writing down an m-function. The m-function m_φ for a belief function φ on a set of variables x assigns each non-empty subset A of Θ_x the probability $m_\varphi(A)$ that the random subset is equal to A. Since the probabilities for all the subsets, including the empty set, must add to one, we must have

$$\sum_{A \ldots \Theta_x, A \neq \emptyset} m_\varphi(A) \leq 1. \tag{3.5}$$

The m-function is particularly useful for vacuous extension. Indeed, if $x \leq y$, then the m-function for the vacuous extension $\varphi^{\uparrow y}$ is given by

$$m_{\varphi^{\uparrow y}}(A) = \begin{cases} m_\varphi(B) & \text{if } A = B^{\uparrow y} \\ 0 & \text{if } A \neq B^{\uparrow y} \text{ for any subset B of } \Theta_x \end{cases}$$

for each subset A of Θ_y.

If m_φ is φ's m-function, then its b-function is given by

$$b_\varphi(A) = \sum_{B \ldots A, B \neq \emptyset} m_\varphi(B). \tag{3.6}$$

The number $b_\varphi(A)$ is the total probability that the random set is non-empty but is contained in A. This is of interest because it is interpreted (usually after normalization) as the degree of belief in A. The b-function is also useful for marginalization. Indeed, if $y \leq x$, then the b-function for the marginal $\varphi^{\downarrow y}$ is given by

$$b_{\varphi^{\downarrow y}}(A) = b_\varphi(A^{\uparrow x}) \tag{3.7}$$

for every subset A of Θ_y.

If m_φ is φ's m-function, then its q-function is given by

$$q_\varphi(A) = \sum_{B \supseteq A} m_\varphi(B).$$

We can combine belief functions simply by multiplying their q-functions:

$$q_{\varphi_x \otimes \varphi_y}(A) = q_{\varphi_x}(A^{\downarrow x}) \cdot q_{\varphi_y}(A^{\downarrow y})$$

for every subset A of $\Theta_{x \vee y}$.

Each of the three functions, m_φ, b_φ, and q_φ, completely specifies the belief function φ. If we start with m_φ, we can obtain b_φ and q_φ by (3.6) and (3.7). If we start with b_φ, we can obtain m_φ and q_φ by

$$m_\varphi(A) = \sum_{B...A, B\neq\emptyset}(-1)^{|A-B|}b_\varphi(B)$$

and

$$q_\varphi(A) = \sum_{B...A, B\neq\emptyset}(-1)^{|A-B|}b_\varphi(B\cup(\Theta_x\text{-}A)).$$

If we start with q_φ, we can obtain m_φ and b_φ by

$$m_\varphi(A) = \sum_{B\supseteq A}(-1)^{|B-A|}q_\varphi(B)$$

and

$$b_\varphi(A) = \sum_{B\cap A\neq\emptyset}(-1)^{|B|+1}q_\varphi(B).$$

Since hypertree computation involves alternate marginalization and combination, a straightforward implementation would use a number of these formulas. Suppose, indeed, that we want to marginalize a product on a hypertree H. For each domain h in H, there might be several belief functions, some on h and some on subdomains of h, that are included in the product (cf. Section 3.3.3). Suppose these belief functions are initially expressed as m-functions. Then to eliminate a twig t with branch b, we first vacuously extend the m-functions on subdomains of t to m-functions on t. We then transform all the m-functions on t into q-functions and multiply them together. Then we transform the result into a b-function and marginalize it to $b\wedge t$. Then we transform it back into an m-function on $b\wedge t$. This produces a new collection of m-functions, all on subdomains of domains in H-{t}. We can continue eliminating twigs in this way until only one domain h is left in the hypergraph. Then we can go through the process one more time, and stop with a b-function on h or on some subdomain of h that interests us.

Condition (3.5) is not needed for our computational theory. In the standard version of belief-function theory, however, we use a scale from zero to one to measure belief or support for a subset, with zero meaning no evidence that the truth is in the set, and one meaning certainty that it is. If we begin with certainty that the truth is in Θ_x, then we want

$$b_\varphi(\Theta_x) = \sum_{A...\Theta_x, A\neq\emptyset}m_\varphi(A) = 1$$

for the b-functions that we interpret as degrees of belief—i.e., the inputs and the final output. Making the inputs satisfy this condition is not enough to ensure that the output does. It ensures only that the output satisfies $b_\varphi(\Theta_x) < 1$. This is because the intersection of non-empty sets can be empty. The standard theory therefore includes a final step in which the output b-function b_φ is normalized—i.e. multiplied by $[b_\varphi(\Theta_x)]^{-1}$. This will be possible unless b_φ is identically zero, which will happen only when the random sets being combined have empty intersection with probability one. The interpretation in this case is that combination is impossible because the belief functions being combined express contradictory certainties.

3.5. Other Comments

Here we comment on two ways the axioms of Section 3.1 can be strengthened.

3.5.1. Marginalizing a Potential to Its Own Domain

In all the examples we studied in Section 3.4, it does not change a potential to marginalize it to its own domain: if φ_x is a potential on x, then $\varphi_x{\downarrow}^x = \varphi_x$. This property is not implied by Axioms A1-A2 and M1-M2, however. It is easy to construct examples of these axioms in which it is violated.

Indeed, given any instance of the axioms, we can construct another instance for which $\varphi_x{\downarrow}^x = \varphi_x$ does not always hold. We simply double the number of potentials by distinguishing two versions of each potential, one marked and one unmarked. Both behave the same, except that the marginal of every potential is marked, and when any potential in a product is marked, the product is marked. Then the axioms will be satisfied, with the unmarked version of ι_x as the identity. And if φ_x is an unmarked potential on x, then $\varphi_x{\downarrow}^x$ is not equal to φ_x.

The idea of marking potentials whenever any computation is performed on them might be useful in some applications, so this example is not entirely artificial.

3.5.2. Are There Non-Modular Examples?

The axioms we have studied in this chapter can be satisfied with any lattice $\mathfrak{I}$. We simply take the semigroup Φ to be $\mathfrak{I}$ itself, with $\vee$ as the semigroup operation. Then we define d and $\downarrow$ by $d(x) = x$ and $y{\downarrow}^x = y \wedge x$. The semigroup Φ_x then has only one element, x itself, which is therefore the identity ι_x. And Axioms A1-A2 and M1-M2 are all satisfied.

The lattices in our computational examples, on the other hand, are very well-behaved. In fact, they are all Boolean algebras. This raises the question of how general the axioms are in a practical sense. Are there interesting computational examples in which the lattice is non-modular, say?

The following lemma, which depends on modularity, will be useful in Chapter 5.

Lemma 3.6. If $\mathfrak{I}$ is modular, $x < z < x \vee y$, and φ_x and φ_y are potentials on x and y, respectively, then $(\varphi_x \otimes \varphi_y){\downarrow}^z = \varphi_x \otimes \varphi_y{\downarrow}^{y \wedge z}$.

Proof: By Axiom M2,

$$(\varphi_x \otimes \varphi_y){\downarrow}^z = (\varphi_x \otimes \iota_z \otimes \varphi_y){\downarrow}^z = (\varphi_x \otimes \iota_z) \otimes \varphi_y{\downarrow}^{y \wedge z} = (\varphi_x \otimes \varphi_y{\downarrow}^{y \wedge z}) \otimes \iota_z.$$

By Lemma 2.1, $x \vee (y \wedge z)$, which is the domain of $(\varphi_x \otimes \varphi_y{\downarrow}^{y \wedge z})$, is equal to z.

So $(\varphi_x \otimes \varphi_y{\downarrow}^{y \wedge z}) \otimes \iota_z = \varphi_x \otimes \varphi_y{\downarrow}^{y \wedge z}$.

If the lattice $\mathfrak{I}$ is not modular, then the conclusion of Lemma 3.6 will not hold. According to Lemma 2.1, there will be x, y, and z such that $x < z < x \vee y$ and yet $x \vee (y \wedge z) \neq z$, and in this case, the two potentials equated by the lemma will have different domains.

4. Strongly Vacuous Extension

In most of the examples that we studied in Chapter 3, vacuous extension can be reversed by marginalization; whenever we vacuously extend a potential up to a larger domain and then marginalize back down again, we get the potential with which we began. In this case, we say that vacuous extension is strongly vacuous.

When vacuous extension is strongly vacuous, the mathematical theory of the preceding chapter simplifies, and we can reformulate the axioms in a several different ways. In particular, we can eliminate the distinction between a potential and its vacuous extension altogether. We will explore these possibilities in this chapter.

The purpose of this chapter is mathematical insight. It breaks no new ground computationally, and it can be omitted by readers who prefer to move on immediately to the further computational theory in the succeeding chapters.

In Section 4.1, we introduce the stability axiom, which we call Axiom M3. In Section 4.2, we show the equivalence of Axioms A1-A2 and M1-M3 to a set of axioms that use vacuous extension as a primitive. In Section 4.3, we reaxiomatize again, using the idea of projection between arbitrary domains as a primitive. In Section 4.4, we show how the axioms for projection simplify when we identify potentials that differ only by vacuous projection.

4.1. The Stability Axiom

Vacuous extension is strongly vacuous if and only if the marginal of an identity is always an identity.

Lemma 4.1. Suppose Axioms A1-A2 and M1-M2 hold, and suppose $x \le y$. Then $(\varphi^{\uparrow y})^{\downarrow x} = \varphi$ for every potential φ on x if and only if $\iota_y^{\downarrow x} = \iota_x$.

Proof: If $x \le y$ and φ is a potential on x, then $(\varphi^{\uparrow y})^{\downarrow x} = (\varphi \otimes \iota_y)^{\downarrow x} = \varphi \otimes \iota_y^{\downarrow x}$. Hence $(\varphi^{\uparrow y})^{\downarrow x} = \varphi$ for every potential φ on x if and only if $\varphi \otimes \iota_y^{\downarrow x} = \varphi$ for every potential φ on x. But the latter condition is equivalent to saying that $\iota_y^{\downarrow x}$ is the identity on Φ_x.

Since Axioms A1-A2 and M1-M2 use identities rather than vacuous extension as a primitive, it is natural to adopt the condition $\iota_y^{\downarrow x} = \iota_x$ as an additional axiom if we want to stipulate that vacuous extension is strongly vacuous. So we formulate the following axiom:

Axiom M3. If $x \le y$, $\iota_y^{\downarrow x} = \iota_x$.

We call this the stability axiom.

By setting y equal to x in Lemma 4.1, we see that if vacuous extension is strongly vacuous (i.e., the stability axiom holds), then $\varphi_x^{\downarrow x} = \varphi_x$ for every potential φ_x on x. As we saw in Section 3.5, this property does not follow from Axioms A1-A2 and M1-M2 alone.

The stability axiom holds in all but one of the examples we studied in Section 3.4. The exception is the probability example, in 3.4.2. In that example, $\iota_y{\downarrow}^x = k \cdot \iota_x$, where k is the number of elements in the frame Θ_{y-x}. We can force this example into a form that satisfies Axiom M3, but at only at a cost to the computational relevance of the theory.

One way to make the probability example satisfy Axiom M3 is to drop any distinction between potentials that differ by a multiplicative constant. In other words, we say that a potential on x is an equivalence class of real-valued functions on x, where two functions are equivalent if one is equal to a non-zero constant times the other. The equivalence classes inherit the semigroup operation, the labelling, and the marginalization map, and Axioms A1-A2 and M1-M3 are all satisfied. This loses computational relevance because it talks about equivalence classes, while we are actually working with concrete functions, and because it puts the computation of the constant $(\Pi\{\varphi_h | h \varepsilon H\}){\downarrow}^{\varnothing}$, which is important in most applications, outside the theory.

Another approach is to consider only real-valued functions that take positive values and sum to one, and to replace marginalization with conditional expectation (in other words, $\varphi{\downarrow}^x(c)$ is the average of $\varphi{\downarrow}^x(d)$ over all d for which $d{\downarrow}^x = c$). This takes us even farther away from actual computational problems. In practice we want to work with functions that are not yet normalized to sum to one, and we will avoid conditional expectation in favor of marginalization when possible, because conditional expectation requires divisions, which are computationally costly.

This example is reason enough to omit Axiom M3 from our general computational theory. But it would be interesting to know if there are other computational examples in which Axiom M3 does not hold, and it would also be interesting to see examples that diverge from the axiom more sharply, in the sense that it cannot be satisfied by taking equivalence classes that retain significant aspects of the computational problem.

On the other hand, it is worthwhile to study Axiom M3 further, because a theory that includes this axiom is the natural framework for a number of ideas and properties that occur in many examples.

4.2. Axioms for Strongly Vacuous Extension

When vacuous extension is strongly vacuous, we can use it in place of identities as an axiomatic primitive.

Here is how it goes. We begin with the operations d, $\otimes$, $\uparrow$, and $\downarrow$ as our primitives. We assume that

d maps Φ to $\mathfrak{I}$,

$\otimes$ is a commutative binary operation on Φ,

if $x \leq y$, and φ_x is a potentiat on x, then $\varphi_x{\uparrow}^y$ is a potential on y, and

if $x \leq y$, and φ_y is a potentiat on y, then $\varphi_y{\downarrow}^x$ is a potential on x.

And we adopt the following axioms.

Axiom V1. If φ_x and φ_y are potentials on x and y, respectively, then $d(\varphi_x \otimes \varphi_y) = x \vee y$.
Axiom V2. If φ_x is a potential on x, then $\varphi_x{\uparrow}^x = \varphi_x$.
Axiom V3. If φ_x is a potential on x, and $x \le y \le z$, then $(\varphi_x{\uparrow}^y){\uparrow}^z = \varphi_x{\uparrow}^z$.
Axiom V4. If φ_x and φ_y are potentials on x and y, respectively, $x \le z$, and $y \le z$, then $(\varphi_x \otimes \varphi_y){\uparrow}^z = \varphi_x{\uparrow}^z \otimes \varphi_y{\uparrow}^z$.
Axiom V5. If φ_z is a potential on z, and $x \le y \le z$, then $(\varphi_z{\downarrow}^y){\downarrow}^x = \varphi_z{\downarrow}^x$.
Axiom V6. If φ_x and φ_y are potentials on x and y, respectively, then $(\varphi_x \otimes \varphi_y){\downarrow}^x = \varphi_x \otimes \varphi_y{\downarrow}^{y \wedge x}$.
Axiom V7. If φ_x is a potential on x, then $(\varphi_x{\uparrow}^{x \vee y}){\downarrow}^y = (\varphi_x{\downarrow}^{x \wedge y}){\uparrow}^y$.
Axiom V8. If φ_x is a potential on x, and $x \le y$, then $(\varphi_x{\uparrow}^y){\downarrow}^x = \varphi_x$.

We have already verified that Axioms V1-V8 are implied by Axioms A1-A2 and M1-M3:

Axiom V1 is Axiom A1.
Axiom V2 is statement (iv) of Lemma 3.2.
Axiom V3 is statement (vi) of Lemma 3.2.
Axiom V4 is statement (vii) of Lemma 3.2.
Axiom V5 is the same as Axiom M1.
Axiom V6 is the same as Axiom M2.
Axiom V7 is Lemma 3.3.
Axiom V8 is part of Lemma 4.1.

It is also true, that Axioms V1-V8 imply Axioms A1-A2, and M1-M3. More precisely, if Axioms V1-V8 hold, then we can adjoin identities to the semigroups Φ_x and extend the operations d, $\otimes$, $\uparrow$, and $\downarrow$ to include these identities in such a way that Axioms A1-A2 and M1-M3 hold.

Suppose, indeed, that Axioms V1-V8 hold. Then it follows from Axiom V1 that Φ_x, the set of all potentials with domain x, is a semigroup. Hence we can adjoin an identity ι_x to it. We can then extend $\otimes$ to an operation on $\Phi \cup \{\iota_x | x \varepsilon \mathfrak{I}\}$ by setting $\iota_x \otimes \iota_y$ equal to $\iota_{x \vee y}$ and by setting $\varphi_x \otimes \iota_y$ and $\iota_y \otimes \varphi_x$ equal to $\varphi_x{\uparrow}^{x \vee y}$. Axiom A2 is satisfied by definition. The extended operation $\otimes$ is obviously commutative and satisfies Axiom A1; using Axioms V1-V4, we can show it is associative as well.

Next, we extend the definitions of $\uparrow$ and $\downarrow$ to $\Phi \cup \{\iota_x | x \varepsilon \mathfrak{I}\}$ by setting $\iota_x{\uparrow}^y = \iota_y$ and $\iota_y{\downarrow}^x = \iota_x$ whenever $x \le y$. Axiom M1 follows immediately from these definitions and Axiom V5. To derive Axiom M2, we use Axiom V6, and we note that

$$\begin{aligned}(\iota_x \otimes \varphi_y){\downarrow}^x &= (\varphi_y{\uparrow}^{x \vee y}){\downarrow}^x \\ &= (\varphi_y{\downarrow}^{y \wedge x}){\uparrow}^x && \text{by Axiom V7} \\ &= \iota_x \otimes \varphi_y{\downarrow}^{y \wedge x}\end{aligned}$$

and

$$\begin{aligned}(\varphi_x \otimes \iota_y){\downarrow}^x &= (\varphi_x{\uparrow}^{x \vee y}){\downarrow}^x \\ &= \varphi_x && \text{by Axiom V8}\end{aligned}$$

$= \varphi_x{\uparrow}^x$ by Axiom V2

$= \varphi_x \otimes \iota_{y \wedge x}$

$= \varphi_x \otimes \iota_y{\downarrow}^{y \wedge x}.$

Note the essential role of Axiom V8 in the last derivation.

Axioms V1-V7 follow from Axioms A1-A2 and M1-M2 alone, without Axiom M3. Going the other way, however, it does not appear possible to adjoin identities and derive Axioms A1-A2 and M1-M2 without using Axiom V8. So although Axioms V1-V8 succeed in reaxiomatizing the theory of Axioms A1-A2 and M1-M3 in terms of vacuous projection, it does not appear that the theory of Axioms A1-A2 and M1-M2 alone can be similarly reaxiomatized.

4.3. Projection

Recall Lemma 3.3 and Axiom V7:

$$(\varphi{\uparrow}^{x \vee y}){\downarrow}^y = (\varphi{\downarrow}^{x \wedge y}){\uparrow}^y \qquad (4.1)$$

whenever φ is a potential on x. This equation suggests that the map that takes φ into the potential $(\varphi{\uparrow}^{x \vee y}){\downarrow}^y$ might also be used as a primitive. In fact, doing so considerably simplifies Axioms V1-V8.

Given a potential φ on x, let us write $\varphi^{\rightarrow y}$ for $(\varphi{\uparrow}^{x \vee y}){\downarrow}^y$, and let us call $\varphi^{\rightarrow y}$ the *projection* of φ to y. (Notice that $\varphi^{\rightarrow y}$ is defined for any potential φ and any domain y.) It can be shown that if Axioms V1-V8 hold for the operations d, $\otimes$, $\uparrow$, and $\downarrow$, then the following axioms hold for the operations d, $\otimes$, and $\rightarrow$:

Axiom P1. If φ_x and φ_y are potentials on x and y, respectively, then $d(\varphi_x \otimes \varphi_y) = x \vee y$.

Axiom P2. If φ_x is a potential on x, then $\varphi_x^{\rightarrow x} = \varphi_x$.

Axiom P3. $(\varphi^{\rightarrow x})^{\rightarrow y} = (\varphi^{\rightarrow x \wedge y})^{\rightarrow y} = (\varphi^{\rightarrow x \vee y})^{\rightarrow y}$.

Axiom P4. If φ_x and φ_y are potentials on x and y, respectively, then $(\varphi_x \otimes \varphi_y)^{\rightarrow x} = \varphi_x \otimes \varphi_y{}^{\rightarrow x}$.

Conversely, if Axioms P1-P4 hold for d, $\otimes$, and $\rightarrow$, then Axioms V1-V8 hold for d, $\otimes$, $\uparrow$, and $\downarrow$. (If we start with $\rightarrow$, then we define $\uparrow$ and $\downarrow$ as special cases of $\rightarrow$; when $x \le y$, we write $\varphi_x{\uparrow}^y$ for $\varphi_x^{\rightarrow y}$, and when $y \le x$, we write $\varphi_x{\downarrow}^y$ for $\varphi_x^{\rightarrow y}$.) Proofs are left to the reader.

Though (4.1) holds even if vacuous extension is not strongly vacuous, it does not appear that Axioms P1-P4 can be adapted in any simple way to make them equivalent to Axioms A1-A2 and M1-M2 alone. Apparently the theory of Axioms A1-A2 and M1-M2 cannot be axiomatized in terms of projection.

4.4. Projection Without Labels

Now we formulate axioms that dispense with labels and hence with the distinction between a potential and its vacuous extension.

4.4.1. *Axioms*

Here is the way it goes. We begin with a lattice $\mathfrak{I}$ and a commutative semigroup $(\vartheta,\oplus)$. We associate with every potential ϕ in ϑ and every domain x in $\mathfrak{I}$ a potential $\phi^{\Rightarrow x}$ that we call the abstract projection of ϕ on x. If $\phi^{\Rightarrow x} = \phi$, we say that x *supports* ϕ, or that ϕ is measurable with respect to x. We adopt two axioms.

Axiom P*1. $(\phi^{\Rightarrow x})^{\Rightarrow y} = \phi^{\Rightarrow x \wedge y}$.

Axiom P*2. If x supports ϕ_1, then $(\phi_1 \oplus \phi_2)^{\Rightarrow x} = \phi_1 \oplus (\phi_2^{\Rightarrow x})$.

We assume, without loss of generality, that the semigroup contains an identity. If it did not, then we could adjoin an identity ι and set $\iota^{\Rightarrow x} = \iota$ for all x, and Axioms P*1 and P*2 would still hold.

The following lemma lists some consequences of Axioms P*1-P*2.

Lemma 4.2.

(i) x supports $\phi^{\Rightarrow x}$.
(ii) If x supports ϕ, then x supports $\phi^{\Rightarrow y}$.
(iii) If x and y both support ϕ, then $x \wedge y$ supports ϕ.
(iv) If x supports ϕ, then $x \wedge y$ supports $\phi^{\Rightarrow y}$.
(v) If x supports ϕ, then $\phi^{\Rightarrow y} = \phi^{\Rightarrow x \wedge y}$.
(vi) If x supports ϕ and $x \le y$, then y supports ϕ.
(vii) If x supports both ϕ_1 and ϕ_2, then x supports $\phi_1 \oplus \phi_2$.
(viii) If x supports ϕ_1 and y supports ϕ_2, then $x \vee y$ supports $\phi_1 \oplus \phi_2$.

Proof:

(i) Substituting x for y in Axiom P*1, we obtain $(\phi^{\Rightarrow x})^{\Rightarrow x} = \phi^{\Rightarrow x}$. So x supports $\phi^{\Rightarrow x}$.

(ii) It follows from Axiom P*1 that $(\phi^{\Rightarrow y})^{\Rightarrow x} = (\phi^{\Rightarrow x})^{\Rightarrow y}$. If x supports ϕ, then we can substitute ϕ for $\phi^{\Rightarrow x}$ on the right-hand side of this equation, obtaining $(\phi^{\Rightarrow y})^{\Rightarrow x} = \phi^{\Rightarrow y}$. This means that x supports $\phi^{\Rightarrow y}$.

(iii) If x and y both support ϕ, then by substituting ϕ first for $\phi^{\Rightarrow x}$ and then for $\phi^{\Rightarrow y}$ in Axiom P*1, we find that $\phi = \phi^{\Rightarrow x \wedge y}$, which means that $x \wedge y$ supports ϕ.

(iv) This follows from (i), (ii), and (iii).

(v) This follows from Axiom P*1 and the definition of support.

(vi) If $x \le y$, then $x \wedge y = x$. So Axiom P*1 says that $(\phi^{\Rightarrow x})^{\Rightarrow y} = \phi^{\Rightarrow x}$. If x supports ϕ, then this reduces to $\phi^{\Rightarrow y} = \phi$, which means that y supports ϕ.

(vii) If x supports ϕ_1, then Axiom P*2 says that $(\phi_1 \oplus \phi_2)^{\Rightarrow x} = \phi_1 \oplus (\phi_2^{\Rightarrow x})$. If x also supports ϕ_2, then we can substitute ϕ_2 for $\phi_2^{\Rightarrow x}$ in this equation, obtaining $(\phi_1 \oplus \phi_2)^{\Rightarrow x} = \phi_1 \oplus \phi_2$, which means that x supports $\phi_1 \otimes \phi_2$.

(viii) If x supports ϕ_1 and y supports ϕ_2, then, by (vi), $x \vee y$ supports both ϕ_1 and ϕ_2. Hence, by (vii), $x \vee y$ supports $\phi_1 \oplus \phi_2$.

4.4.2. *Translation to Labelled Potentials*

Axioms P*1-P*2 are essentially equivalent to Axioms A1-A2 and M1-M3, to Axioms V1-V8, and to Axioms P1-P4. Since we are most interested in Axioms A1-A2 and M1-

M3, which are closest to our computational theory, let us spell out formally the equivalence of P*1-P*2 with A1-M3. First, let us say how to go from P*1-P*2 to A1-M3.

Lemma 4.3. Suppose Axioms P*1-P*2 are satisfied by the lattice $\mathfrak{I}$, the commutative semigroup $(\vartheta,\oplus)$ and the abstract projection $\Rightarrow$. Define $(\Phi,\otimes)$ by

$$\Phi = \{(\phi,x) \mid \phi\varepsilon\vartheta,\ x\varepsilon\mathfrak{I}, \text{ and } x \text{ supports } \phi\}$$

and

$$(\phi_1,x)\otimes(\phi_2,y) = (\phi_1\oplus\phi_2, x\vee y),$$

define a map d from Φ to $\mathfrak{I}$ by

$$d((\phi,x)) = x,$$

and define a map $\downarrow$ by

$$(\phi,x)^{\downarrow y} = (\phi^{\Rightarrow y},y)$$

whenever $y \leq x$. Then $(\Phi,\otimes)$ is a commutative semigroup, and Axioms A1-A2 and M1-M3 are satisfied by $\mathfrak{I}$, $(\Phi,\otimes)$, d, and $\downarrow$.

Then let us say how to go from Axioms A1-A2 and M1-M3 to Axioms P*1-P*2.

Lemma 4.4. Suppose Axioms A1-A2 and M1-M3 are satisfied by the lattice $\mathfrak{I}$, the commutative semigroup $(\Phi,\otimes)$, the labelling d, and the marginalization $\downarrow$. Define a relation $\approx$ on the set Φ by

$$\varphi_1 \approx \varphi_2 \text{ if and only if } \varphi_1\otimes\iota_{d(\varphi_2)} = \varphi_2\otimes\iota_{d(\varphi_1)}.$$

Then $\approx$ is an equivalence relation, and any two identities ι_x and ι_y are in the same equivalence class. Let $[\varphi]$ denote the equivalence class containing φ, and let ϑ denote the set of all the equivalence classes.

If $\varphi_1 \approx \varphi_2$ and $\varphi_3 \approx \varphi_4$, then $\varphi_1\otimes\varphi_3 \approx \varphi_2\otimes\varphi_4$. Thus a binary operation $\oplus$ on ϑ can be defined by $[\varphi_1]\oplus[\varphi_2] = [\varphi_1\otimes\varphi_2]$. The pair $(\vartheta,\oplus)$ is a commutative semigroup, with $[\iota_x]$ as its identity.

If $\varphi_1 \approx \varphi_2$, and x is a subdomain of both their labels, then $\varphi_1^{\downarrow x} \approx \varphi_2^{\downarrow x}$. So we can define a map $\Rightarrow$ by $[\varphi]^{\Rightarrow x} = (\varphi\otimes\iota_x)^{\downarrow x}$.

Axioms P*1-P*2 are satisfied by $\mathfrak{I}$, $(\vartheta,\oplus)$ and $\Rightarrow$.

4.4.3. Least Support

If the lattice $\mathfrak{I}$ is finite, then every potential φ has a least support. (This follows from statement (iii) of Lemma 4.2.) We could call this the label for φ, but it would not necessarily follow the labelling axiom, for the least support of a product may be smaller than the least support of its factors. A variable can influence two real-valued functions, for example, without influencing the product, in which case it is be in the least support of both functions but not the least support of the product. (Example: Suppose f and g are functions on the variable ξ, which takes the values 0 and 1. Suppose f(0)=2, f(1)=4, g(0)=6, and g(1)=3. Then d(f)=d(g)=$\{\xi\}$, but d(f·g)=$\varnothing$, because f·g is a constant; it is always equal to 12.)

4.4.4. *Semi-Computational Theory*

By relying on the idea of support rather than on the specification of labels, we can derive from Axioms P*1-P*2 a theory that is similar to the computational theory of Section 3.3. The following lemma illustrates the point.

Lemma 4.5. Suppose H is a hypergraph. For each domain h in H, suppose ϕ_h is a potential supported by h. Set

$$\phi = \oplus\{\phi_h | h \varepsilon H\}.$$

Suppose t is a twig in H and b is a branch for t. Then

$$\phi^{\Rightarrow \vee H^{-t}} = \oplus\{\phi_h^{-t} | h \varepsilon H^{-t}\},$$

where $\phi_b^{-t} = \phi_b \oplus (\phi_t^{\Rightarrow b \wedge t})$ and $\phi_h^{-t} = \phi_h$ for all other h in H^{-t}.

Proof. Apply Axiom P*2 with $\vee H^{-t}$ for x, $\oplus\{\phi_h | h \varepsilon H^{-t}\}$ for ϕ_1, t for y, and ϕ_t for ϕ_2. This yields

$$\phi^{\Rightarrow \vee H^{-t}} = (\oplus\{\phi_h | h \varepsilon H^{-t}\}) \oplus \phi_t^{\Rightarrow \vee H^{-t}}.$$

Since ϕ_t is supported by t, $\phi_t = \phi_t^{\Rightarrow t}$, and hence, by statement (v) of Lemma 4.2, $\phi_t^{\Rightarrow \vee H^{-t}} = \phi_t^{\Rightarrow (\vee H^{-t}) \wedge t}$. Since b is a branch for t in H, $(\vee H^{-t}) \wedge t = b \wedge t$. So

$$\phi^{\Rightarrow \vee H^{-t}} = (\oplus\{\phi_h | h \varepsilon H^{-t}\}) \oplus (\phi_t^{\Rightarrow b \wedge t}).$$

By the commutativity and associativity of $\oplus$, we can shift ϕ_b within the right-hand side of this equation to obtain

$$\phi^{\Rightarrow \vee H^{-t}} = (\oplus\{\phi_h | h \varepsilon (H^{-t} - \{b\})\}) \oplus (\phi_b \oplus \phi_t^{\Rightarrow b \wedge t}).$$

This lemma allows us to eliminate twigs step-by-step in a hypertree. Since ϑ contains an identity, we can also use a hypertree cover as in Section 3.3.3.

This theory we obtain in this way is less interesting than the computational theory based directly on Axioms A1-A2 and M1-M3, because in practice, computation does require labels in some form. And if we use labels but think in terms of support, we are complicating our story. We may even be creating new computational problems, for it is not always easy to find the least support for a potential.

5. Continuation and Normalization

This chapter deals with an important special case of the outward phase of hypertree computation—the case where the axioms of Chapter 3 are satisfied and, in addition, continuers exist. In this case, after moving inward in the hypertree to compute the marginal on the root, we can move back out to compute marginals on the other domains in the hypertree.

We call ψ a continuer of φ from x to y if combining ψ with φ's marginal on x produces φ's marginal on y:

$$\varphi^{\downarrow x} \otimes \psi = \varphi^{\downarrow y}. \tag{5.1}$$

As we will see, continuers exist in all the examples of Chapter 3. Finding a continuer in these examples is sometimes a little harder than finding a product. Intuitively, it requires a division; solving (5.1) for ψ requires that we divide each side by $\varphi^{\downarrow x}$.

We will assume that the lattice $\mathfrak{I}$ contains a zero $\varnothing$, and we call a continuer from $\varnothing$ to y a normalization on y. In our examples, a normalization is nearly the same as a marginal. (In the case of a probability distribution P, it is exactly the same; we have $P^{\downarrow\varnothing} = 1$, and hence $P^{\downarrow\varnothing}\otimes\psi = P^{\downarrow y}$ means $\psi = P^{\downarrow y}$.)

Section 5.1 of this chapter studies the definition of continuer and explores the implications of the existence of continuers. Section 5.2 shows how continuers can be used in the outward phase of hypertree computation. Section 5.3 applies this computational theory to the examples of Chapter 3.

5.1. Properties and Implications of Continuation

We begin with the framework of Chapter 3. In other words, we adopt Axioms A1-A2 and M1-M2.

Here is the definition of continuer again. Suppose φ is a potential on z, and $x \le y \le z$. Suppose

$$\varphi^{\downarrow x}\otimes\psi = \varphi^{\downarrow y}. \tag{5.1}$$

Then we say that ψ continues φ from x to y, or that ψ is a continuer of φ from x to y.

We have encountered continuation before. If $x \le y$, φ is a potential on x, and Axiom M3 holds, then ι_y is a continuer of $\varphi^{\uparrow y}$ from x to y.

The following lemma lists some consequences of the definition of continuer:

Lemma 5.1.

(i) If ψ continues φ from x to y, then $x \vee d(\psi) = y$.

(ii) If $w \le x \le y \le z$, and φ is a potential on z, then ψ continues φ from w to x if and only if it continues $\varphi^{\downarrow y}$ from w to x.

(iii) Suppose the lattice $\mathfrak{I}$ is modular. Then if $w \le x \le y \le z$, φ is a potential on z, and ψ continues φ from w to y, then $\psi^{\downarrow x \wedge d(\psi)}$ continues φ from w to x.

(iv) If $w \le x \le y \le z$, φ is a potential on z, ψ_1 continues φ from w to x, and ψ_2 continues φ from x to y, then $\psi_1\otimes\psi_2$ continues φ from w to y.

(v) If φ_x and φ_y are potentials on x and y, respectively, and ψ continues φ_y from $x \wedge y$ to y, then ψ also continues $\varphi_x\otimes\varphi_y$ from x to $x \vee y$.

(vi) If φ_x and φ_y are potentials on x and y, respectively, and ψ continues φ_y from $x \wedge y$ to y, then ψ also continues $\varphi_x\otimes\varphi_y$ from $x \wedge y$ to y.

Proof.

(i) This follows from (5.1) and Axiom A1.

(ii) Since $\varphi^{\downarrow w} = (\varphi^{\downarrow y})^{\downarrow w}$ and $\varphi^{\downarrow x} = (\varphi^{\downarrow y})^{\downarrow x}$, the relation $\varphi^{\downarrow x} = \varphi^{\downarrow w}\otimes\psi$ can also be written $(\varphi^{\downarrow y})^{\downarrow x} = (\varphi^{\downarrow y})^{\downarrow w}\otimes\psi$.

(iii) By Lemma 3.6, when we marginalize both sides of $\varphi^{\downarrow y} = \varphi^{\downarrow w}\otimes\psi$ to x, we obtain $\varphi^{\downarrow x} = (\varphi^{\downarrow w}\otimes\psi)^{\downarrow x} = \varphi^{\downarrow w}\otimes\psi^{\downarrow x \wedge d(\psi)}$.

(iv) We have $\varphi^{\downarrow w}\otimes\psi_1\otimes\psi_2 = \varphi^{\downarrow x}\otimes\psi_2 = \varphi^{\downarrow y}$.

(v) We have $(\varphi_x\otimes\varphi_y)^{\downarrow x}\otimes\psi = \varphi_x\otimes\varphi_y{}^{\downarrow x \wedge y}\otimes\psi = \varphi_x\otimes\varphi_y$.

(vi) Using Lemma 3.1, we find that $(\varphi_x \otimes \varphi_y)^{\downarrow x \wedge y} \otimes \psi = \varphi_x^{\downarrow x \wedge y} \otimes \varphi_y^{\downarrow x \wedge y} \otimes \psi$ $= \varphi_x^{\downarrow x \wedge y} \otimes \varphi_y = (\varphi_x \otimes \varphi_y)^{\downarrow y}$.

[NOTE: We could add to statement (iii) the comment that any continuer of ψ from $x \wedge d(\psi)$ to $d(\psi)$ is also a continuer of φ from x to y. (Proof: If ψ^* continues ψ from $x \wedge d(\psi)$ to $d(\psi)$, then we can substitute the equation $\psi = \psi^{\downarrow x \wedge d(\psi)} \otimes \psi^*$ in the equation $\varphi^{\downarrow w} \otimes \psi = \varphi^{\downarrow x}$ to obtain $\varphi^{\downarrow w} \otimes \psi^{\downarrow x \wedge d(\psi)} \otimes \psi^* = \varphi^{\downarrow y}$, or $\varphi^{\downarrow x} \otimes \psi^* = \varphi^{\downarrow y}$.) Is this of any use?]

We can only use the results in Lemma 5.1 if continuers exist. So in addition to Axioms A1-A2 and M1-M2, we now adopt the following axiom:

Axiom T. If φ is a potential on z, and $x \le y \le z$, then there is at least one continuer of φ from x to y.

As we will see in Section 5.3, Axiom T is satisfied by all the examples we studied in Chapter 3. It would be interesting to know whether there are computationally interesting examples of Axioms A1-A2 and M1-M2 in which continuers do not always exist.

As the following lemma indicates, the existence of continuers can help us even if we do not compute them.

Lemma 5.2. The following statements follow from Axioms A1-A2, M1-M2, and T.

(i) Suppose $x \le y$, φ_1 and φ_2 are potentials on x, φ_y is a potential on y, and

$$\varphi_1 \otimes \varphi_y^{\downarrow x} = \varphi_2 \otimes \varphi_y^{\downarrow x}.$$

Then

$$\varphi_1 \otimes \varphi_y = \varphi_2 \otimes \varphi_y.$$

(ii) Suppose φ_x is a potential on x, and φ_y is a potential on y. Then there exists at least one potential χ on $x \wedge y$ such that

$$\chi \otimes \varphi_y^{\downarrow x \wedge y} = (\varphi_x \otimes \varphi_y)^{\downarrow x \wedge y}. \quad (5.2)$$

Moreover, if χ is any potential satisfying (5.2), then

$$\chi \otimes \varphi_y = (\varphi_x \otimes \varphi_y)^{\downarrow y}.$$

Proof. (i) Let ψ be a continuer of φ_y from x. Then

$$\varphi_1 \otimes \varphi_y = \varphi_1 \otimes \varphi_y^{\downarrow x} \otimes \psi = \varphi_2 \otimes \varphi_y^{\downarrow x} \otimes \psi = \varphi_2 \otimes \varphi_y.$$

(ii) By Lemma 3.1, $\varphi_x^{\downarrow x \wedge y}$ satisfies (5.2). If χ is another potential satisfying (5.2), then by statement (i), together with Axiom M2,

$$\chi \otimes \varphi_y = \varphi_x^{\downarrow x \wedge y} \otimes \varphi_y = (\varphi_x \otimes \varphi_y)^{\downarrow y}.$$

Continuers are not, in general, unique. One aspect of their non-uniqueness is the possibility of continuers to y whose domain is smaller than y. If ψ continues φ from x to y, and $d(\psi)$ is smaller than y, then the vacuous extension $\psi^{\uparrow y}$ is a different continuer of φ from x to y. It is not bad to have a continuer to y whose domain is smaller than y. This happens whenever a potential factors non-trivially on a hypergraph, and the point of this chapter is to take advantage of it. As we will see in Section 5.3, however, the non-uniqueness of continuation often goes beyond this. There are often several different potentials on y that continue φ from x to y.

In order to use statement (iii) of Lemma 5.1, we need $\mathfrak{I}$ to be modular. For the sake of simplicity in our computational theory, we also want $\mathfrak{I}$ to contain a zero. So we adopt one more axiom:

Axiom A3. The lattice $\mathfrak{I}$ is modular and contains a zero $\varnothing$.

This axiom, too, is satisfied by the computational examples we considered in Chapter 3.

If ψ is a continuer of φ from $\varnothing$ to y, where $\varnothing$ is the zero of $\mathfrak{I}$, then we call ψ a normalization of φ on y. Notice that if ψ is a normalization of φ on y, and $x \leq y$, then $\psi^{\downarrow x}$ is a normalization of φ on x.

We say that a continuer ψ of φ from x to y is proper if $\psi^{\downarrow x \wedge d(\psi)} = \iota_{x \wedge d(\psi)}$. As we shall see, proper continuers exist in all the computational examples of Chapter 3. In some cases, it is may be worthwhile to use proper continuers; in other cases, the extra effort needed to find them may not be justified.

5.2. Computational Theory

In Chapter 3, we learned how to find the marginal on the root of a hypertree construction ordering of a potential for which we have a factorization on the hypertree. We do this by marginalizing and combining within domains as we move inward in the hypertree—i.e., backward from twig to branch in the hypertree construction ordering. In this section, we learn how to move back out in a hypertree—forward from branch to twig in the sequence—using continuers to find marginals for the other domains in the hypertree.

We actually give three different theories for the outward computation.

In the first two theories, we find continuers as we marginalize moving inward. Each time we marginalize a potential on a twig to the meet with the twig's branch, we also find and store a continuer of the potential from the meet back to the twig. After we reach the root, we move back out along the construction ordering, marginalizing to meets and combining with the continuers we have stored in order to find the marginals for successive twigs. In the first theory, we start back out as soon as we have found the marginal for the root. In the second theory, we first use this marginal to find a normalization for the root, and the potentials we find for the other domains as we move outward are also normalizations rather than marginals.

In the third theory, we simplify the computation. Instead of combining and storing continuers to the twigs, we merely store all the potentials that we compute as we move inward to find the marginal of the root. This includes a potential on each domain and its marginal on the meet with the branch of that domain. On our way back out, we solve equations that are analogous to (5.1) but which involve potentials on the meets rather than on the larger domains.

The third theory is equivalent to a method explained by Jensen et al. (1990) for the probability example. It is more efficient than the first two theories, since it does its work on smaller domains. The first two theories are theoretically interesting, however, because they help us understand the third theory, and because the second theory works for the more general axioms we consider in the next chapter.

5.2.1. *Using Continuers to Find Marginals*

Finding continuers, like finding products and marginals, often becomes more difficult as the domain is enlarged. So we must modify the definition of feasible domain. In Chapter 3, we called a domain feasible when it is feasible to combine and marginalize potentials on the domain. Now we require that it also be feasible to find continuers of potentials on the domain. We will still say that a hypertree is feasible when all its domains are feasible.

The following lemma provides the key to finding marginals in a feasible hypertree.

Lemma 5.3. Suppose φ is a product on a hypergraph H:

$$\varphi = \otimes\{\varphi_h | h \varepsilon H\}.$$

Suppose that t is a twig in H, and b is a branch for t. And suppose that ψ continues φ_t from $b \wedge t$ to t. Then

(i) ψ continues φ from $b \wedge t$ to t, and

(ii) ψ continues φ from vH^{-t} to vH.

Proof. Statements (i) and (ii) are special cases of statements (vi) and (v), respectively, of Lemma 5.1.

Suppose H is a feasible hypertree, and recall Chapter 3's recipe for finding the marginal of φ on a domain h_1 of H. First we find a construction ordering $h_1,h_2,\ldots,h_n$, and a branching b(i) for this ordering. Then we marginalize from H^n to H^{n-1} to H^{n-2} and so on, where H^i is the hypertree $\{h_1,h_2,\ldots,h_i\}$. At the step from H^i to H^{i-1}, we go from a collection $\{\varphi_h^{\,i} | h \varepsilon H^i\}$, which has $\varphi^{\downarrow vH^i}$ as its product, to a collection $\{\varphi_h^{\,i-1} | h \varepsilon H^{i-1}\}$, which has $\varphi^{\downarrow vH^{i-1}}$ its product. To go from $\{\varphi_h^{\,i} | h \varepsilon H^i\}$ to $\{\varphi_h^{\,i-1} | h \varepsilon H^{i-1}\}$, we omit h_i and change the potential on $h_{b(i)}$ from $\varphi_{h_{b(i)}}^{\,i}$ to $\varphi_{h_{b(i)}}^{\,i} \otimes (\varphi_{h_i}^{\,i})^{\downarrow h_{b(i)} \wedge h_i}$; the other potentials are unchanged. The collection with which we begin, $\{\varphi_h^{\,n} | h \varepsilon H^n\}$, is simply $\{\varphi_h | h \varepsilon H\}$, and the collection with which we end, $\{\varphi_h^{\,1} | h \varepsilon H^1\}$, consists of the single potential $\varphi^{\downarrow h_1}$.

Now suppose that at each step we also find a continuer. As we are work within h_i to marginalize $\varphi_{h_i}^{\,i}$ to $h_{b(i)} \wedge h_i$, we also work within h_i to find a continuer of $\varphi_{h_i}^{\,i}$ from $h_{b(i)} \wedge h_i$ to h_i. We write ψ_i for this continuer.

Lemma 5.4. The potential ψ_i continues φ from $h_{b(i)} \wedge h_i$ to h_i and also from vH^{i-1} to vH^i, for i=2,...,n.

Proof. By Lemma 5.3, ψ_i continues $\varphi^{\downarrow vH^i}$ from $h_{b(i)} \wedge h_i$ to h_i and from vH^{i-1} to vH^i. By statement (ii) of Lemma 5.1, it also continues φ from $h_{b(i)} \wedge h_i$ to h_i and from vH^{i-1} to vH^i.

Notice that $d(\psi_i) \leq h_i$.

At the last step, we find $\varphi^{\downarrow h_1}$, together with ψ_2, which continues φ from $h_1 \wedge h_2$ to h_2. According to the second part of Lemma 5.4, we have the following:

$\varphi^{\downarrow h_1}$, the marginal of φ on h_1,

ψ_2, which continues φ from h_1 to $h_1 v h_2$,

ψ_3, which continues φ from $h_1 \vee h_2$ to $h_1 \vee h_2 \vee h_3$,

and so on, up to

ψ_n, which continues φ from $h_1 \vee h_2 \vee \cdots \vee h_{n-1}$ to $h_1 \vee h_2 \vee \cdots \vee h_n$.

By the definition of continuer, together with statement (iv) of Lemma 5.1, these potentials constitute a new factorization of φ.

Now let us describe what we have using the first part of Lemma 5.4:

$\varphi^{\downarrow h_1}$, the marginal of φ on h_1,

ψ_2, which continues φ from $h_1 \wedge h_2$ to h_2,

ψ_3, which continues φ from $h_{b(3)} \wedge h_3$ to h_3,

and so on, up to

ψ_n, which continues φ from $h_{b(n)} \wedge h_n$ to h_n.

This description makes it obvious how to use these potentials to find marginals on all the h_i.

First we marginalize $\varphi^{\downarrow h_1}$ to $h_1 \wedge h_2$ (an operation within h_1). This gives the marginal of φ on $h_1 \wedge h_2$, which we can combine with ψ_2 (an operation within h_2) to obtain the marginal of φ on h_2. Next we find the marginal on h_3. We do this by taking the marginal on h_1 or h_2, whichever is the branch of h_3, and marginalizing it to $h_{b(3)} \wedge h_3$ (an operation within $h_{b(3)}$) and then combining the result with ψ_3 (an operation within h_3). Continuing in this way back out the hypertree construction ordering, we successively obtain marginals on each h_i, working in each case only in the domains h_i and $h_{b(i)}$. The computation at each step is the same as the computation we performed at each step during the inward phase of the computation: marginalization to the meet with a neighbor and then combination with the potential on that neighbor, except that now we move from $h_{b(i)}$ to h_i instead of from h_i to $h_{b(i)}$.

5.2.2. Using Continuers to Find Normalizations

Suppose that instead of stopping the inward phase of the computation when we have found $\varphi^{\downarrow h_1}$, we continue the process one more step, finding $\varphi^{\downarrow \varnothing}$ and a potential ψ_1 that continues φ from $\varnothing$ to h_1.

We now have

ψ_1, which continues φ from $\varnothing$ to h_1,

ψ_2, which continues φ from h_1 to $h_1 \vee h_2$,

ψ_3, which continues φ from $h_1 \vee h_2$ to $h_1 \vee h_2 \vee h_3$,

and so on, up to

ψ_n, which continues φ from $h_1 \vee h_2 \vee \cdots \vee h_{n-1}$ to $h_1 \vee h_2 \vee \cdots \vee h_n$.

By statement (iv) of Lemma 5.1, $\psi_1 \otimes \psi_2 \otimes \cdots \otimes \psi_n$ continues φ from $\varnothing$ to $h_1 \vee h_2 \vee \cdots \vee h_n$—i.e., a normalization of φ on $h_1 \vee h_2 \vee \cdots \vee h_n$.

Now we shift our attention to the fact that

ψ_1 continues φ from $\varnothing$ to h_1, (5.3)

ψ_2 continues φ from $h_1 \wedge h_2$ to h_2, (5.4)

ψ_3 continues φ from $h_{b(3)} \wedge h_3$ to h_3, (5.5)

and so on, up to

ψ_n continues φ from $h_{b(n)} \wedge h_n$ to h_n.

We can use these continuers to find normalizations of φ on all the h_i.

We begin with ψ_1. Its domain is h_1. By (5.3), together with statement (iii) of Lemma 5.1, its marginal $\psi_1^{\downarrow h_1 \wedge h_2}$ continues φ from $\varnothing$ to $h_1 \wedge h_2$. We can find $\psi_1^{\downarrow h_1 \wedge h_2}$ working within h_1, and then we can combine it with ψ_2 working within h_2; this yields, by (5.4) and statement (iv) of Lemma 5.1, a continuer of φ from $\varnothing$ to h_2.

Now we have normalizations on h_1 and h_2 (i.e., continuers of φ from $\varnothing$ to h_1 and from $\varnothing$ to h_2). The next step is to find a normalization on h_3 (i.e., a continuer of φ from $\varnothing$ to h_3). We do this by taking the normalization on h_1 or h_2, whichever is the branch of h_3, and marginalizing it to $h_{b(3)} \wedge h_3 \wedge d(\psi_3)$. By statement (iii) of Lemma 5.1, this yields a continuer of φ from $\varnothing$ to $h_{b(3)} \wedge h_3$. By (5.5) and statement (iv) of Lemma 5.1, we can combine this continuer with ψ_3 to obtain a continuer of φ from $\varnothing$ to h_3.

Continuing in this way back out the tree, we successively obtain normalizations on each h_i, working in each case only in the domains h_i and $h_{b(i)}$.

Lemma 5.5. If the continuers $\psi_1, \psi_2, \ldots, \psi_n$ are proper, then the normalizers on the h_i found by the preceding algorithm are marginals of the normalizer $\psi_1 \otimes \psi_2 \otimes \cdots \otimes \psi_n$.

Proof. Let us write $i_1, i_2, \ldots, i_k$ for the sequence $i, b(i), b((i)), \ldots, 1$. Since the $\psi_1, \psi_2, \ldots, \psi_n$ are proper, $\psi_{i_1} \otimes \psi_{i_2} \otimes \cdots \otimes \psi_{i_k}$ is the marginal of $\psi_1 \otimes \psi_2 \otimes \cdots \otimes \psi_n$ on $h_{i_1} \vee h_{i_2} \vee \cdots \vee h_{i_k}$. To see that this is true, notice that we can marginalize to $h_{i_1} \vee h_{i_2} \vee \cdots \vee h_{i_k}$ by successively eliminating twigs from the hypertree construction sequence. We begin with a twig j other than i, and we find that

$$\begin{aligned}(\psi_1 \otimes \psi_2 \otimes \cdots \otimes \psi_n)^{\downarrow h_1 \vee h_2 \ldots \vee h_{j-1} \vee h_{j+1} \ldots \vee h_n} &= (\psi_1 \otimes \psi_2 \otimes \cdots \otimes \psi_{j-1} \otimes \psi_{j+1} \otimes \cdots \otimes \psi_n) \otimes \psi_j^{\downarrow h_j \wedge h_{b(j)} \wedge d(\psi_j)} \\ &= (\psi_1 \otimes \psi_2 \otimes \cdots \otimes \psi_{j-1} \otimes \psi_{j+1} \otimes \cdots \otimes \psi_n) \otimes \iota_{h_j \wedge h_{b(j)} \wedge d(\psi_j)} \\ &= \psi_1 \otimes \psi_2 \otimes \cdots \otimes \psi_{j-1} \otimes \psi_{j+1} \otimes \cdots \otimes \psi_n.\end{aligned}$$

We can continue eliminating twigs until only $h_{i_1}, h_{i_2}, \ldots, h_{i_k}$ are left.

To complete the proof, note that the algorithm amounts to marginalizing $\psi_{i_1} \otimes \psi_{i_2} \otimes \cdots \otimes \psi_{i_k}$ by eliminating first h_{i_k} (or h_1), then $h_{i_{k1}}$, and so on, until only h_i remains.

[NOTE: The exposition of this proof needs to be improved, but this seems to require some groundwork on join trees or directed acyclic graphs. I am also troubled because I think the conclusion of the lemma is true even if the continuers are not normal. How can this be proven? If it is true, does it generalize to the next chapter?]

5.2.3. Using Division in Separators to Find Marginals

The meets $h_{b(i)} \wedge h_i$, for $i=2,\ldots,n$, are called the separators for the hypertree construction ordering. In our third computational theory, we marginalize and combine, as usual, on the h_i. But we do not compute continuers on the h_i. Instead, we solve equations of the form

$$\chi \otimes \varphi_1 = \varphi_2 \tag{5.6}$$

on the separators. Instead of assuming that it is feasible to find continuers on these domains, we assume only that if φ_1 and φ_2 are potentials on a separator, and χ satisfying (5.6) exist, then it is feasible to find one.

The following lemma brings us closer to seeing how solving equations of the form (5.6) can help us find marginals on the domains of the hypertree.

Lemma 5.6. Suppose φ is a product on a hypergraph H:

$$\varphi = \otimes\{\varphi_h | h \varepsilon H\}.$$

Suppose that t is a twig in H, and b is a branch for t. And suppose that ψ continues φ_t from $b \wedge t$ to t. Then there exists at least one potential χ on $b \wedge t$ such that

$$\chi \otimes \varphi_t^{\downarrow b \wedge t} = \varphi^{\downarrow b \wedge t}. \tag{5.7}$$

Moreover, if χ is any potential satisfying this equation, then

$$\chi \otimes \varphi_t = \varphi^{\downarrow t}. \tag{5.8}$$

Proof. This is a special case of statement (ii) of Lemma 5.2.

In a typical step of the outward computation described in Section 5.2.1 above, we find the marginal $\varphi^{\downarrow b}$, marginalize it to the separator $b \wedge t$, and combine the result with a continuer to find $\varphi^{\downarrow t}$. Lemma 5.6 shows us that the continuer is not needed. If we save $\varphi_t^{\downarrow b \wedge t}$, which we had computed during the inward phase of the computation, we can solve (5.7) for χ and then use (5.8) to find $\varphi^{\downarrow t}$.

So here is what we do. As we move inward, from $\{\varphi_h^i | h \varepsilon H^i\}$ to $\{\varphi_h^{i-1} | h \varepsilon H^{i-1}\}$, we save both $\varphi_{h_i}^i$ and its marginal $(\varphi_{h_i}^i)^{\downarrow h_{b(i)} \wedge h_i}$. Thus at the end of the inward phase, we have

$\varphi^{\downarrow h_1}$, the marginal of φ on h_1,

$\varphi_{h_2}^2$ on h_2 and $(\varphi_{h_2}^2)^{\downarrow h_1 \wedge h_2}$ on the separator $h_1 \wedge h_2$,

$\varphi_{h_3}^3$ on h_3 and $(\varphi_{h_3}^3)^{\downarrow h_{b(3)} \wedge h_3}$ on the separator $h_{b(3)} \wedge h_3$,

and so on, up to

$\varphi_{h_n}^n$ on h_n and $(\varphi_{h_n}^n)^{\downarrow h_{b(n)} \wedge h_n}$ on the separator $h_{b(n)} \wedge h_n$,

On the first step back out, we marginalize $\varphi^{\downarrow h_1}$ to $h_1 \wedge h_2$, then solve the equation

$$\chi \otimes (\varphi_{h_2}^2)^{\downarrow h_1 \wedge h_2} = \varphi^{\downarrow h_1 \wedge h_2}$$

for χ, and then compute

$$\varphi^{\downarrow h_2} = \chi \otimes \varphi_{h_2}^2.$$

On the next step, we marginalize $\varphi^{\downarrow h_1}$ or $\varphi^{\downarrow h_2}$, depending on whether h_1 or h_2 is the branch for h_3, to $h_{b(3)} \wedge h_3$, and then we solve

$$\chi \otimes (\varphi_{h_3}^3)^{\downarrow h_{b(3)} \wedge h_3} = \varphi^{\downarrow h_{b(3)} \wedge h_3}$$

for χ, and then we compute

$$\varphi^{\downarrow h_3} = \chi \otimes \varphi_{h_3}^3.$$

And so on, until we have found $\varphi^{\downarrow h_i}$ for all the h_i.

5.3. Examples

In this section, we review each of the examples we studied in Section 3.4, to verify that Axiom T holds. This means checking that the equation

$$\varphi^{\downarrow x} \otimes \psi = \varphi^{\downarrow y}. \qquad (5.1)$$

has a solution whenever φ is a potential on z, and $x \leq y \leq z$. We also discuss how to solve this equation and the equation

$$\chi \otimes \varphi_t^{\downarrow b \wedge t} = \varphi^{\downarrow b \wedge t}. \qquad (5.7)$$

in each example. And we discuss the interpretation of continuers and normalizations in each example.

As we will see, equations (5.1) and (5.7) can be solved easily except in the example of belief functions. In this example, we must enlarge the semigroup Φ in order to solve these equations.

It should not be thought that continuers exist in all cases of marginalization. An interesting example where continuers do not always exist is provided by the fast retraction marginalization studied by Cowell and Dawid (1991) and Dawid (1991). When continuers do not exist, marginals for all domains in the hypertree can be efficiently computed using the simultaneous propagation scheme of Shafer and Shenoy (1988).

5.3.1. Dynamic Programming

In this example, Ξ is a finite set of variables, each with a finite frame, $\mathfrak{I}$ is the lattice of all subsets of Ξ, and a potential on a subset x of Ξ is a real-valued function on x. We combine potentials by adding them: $\varphi_1 \otimes \varphi_2$ is the potential with domain $d(\varphi_1) \vee d(\varphi_2)$ given by

$$(\varphi_1 \otimes \varphi_2)(c) = \varphi_1(c^{\downarrow d(\varphi_1)}) + \varphi_2(c^{\downarrow d(\varphi_2)})$$

for every configuration c of $d(\varphi_1) \vee d(\varphi_2)$. We marginalize by maximizing out variables:

$$\varphi^{\downarrow x}(c) = \max\{\varphi(c.c') \mid c' \text{ is a configuration of } d(\varphi)-x\}.$$

for every configuration c of x.

In this example, (5.1) and (5.7) have unique solutions, and finding them is simply a matter of subtraction:

$$\psi(c) = \varphi^{\downarrow y}(c) - \varphi^{\downarrow x}(c^{\downarrow x})$$

for every configuration c of y, and

$$\chi(c) = \varphi^{\downarrow b \wedge t}(c) - \varphi_t^{\downarrow b \wedge t}(c)$$

for every configuration c of $b \wedge t$. Notice that $\psi^{\downarrow x}$ is identically equal to zero. Thus the continuer ψ is proper.

In general, $\psi(c)$ is always negative or zero. It tells how far $\varphi^{\downarrow y}(c)$ is from the maximum value that can be achieved by holding the coordinates in $c^{\downarrow x}$ constant while varying the coordinates in $c^{\downarrow y-x}$. The value of a normalization tells how far the particular configuration is from attaining the maximum of the potential.

5.3.2. *Factored Probability Distributions*

Again, Ξ is a finite set of variables, each with a finite frame, $\Im$ is the lattice of all subsets of Ξ, and a potential on a subset x of Ξ is a real-valued function on x. But in this example we combine by multiplying: $\varphi_1 \otimes \varphi_2$ is the potential with domain $d(\varphi_1) \vee d(\varphi_2)$ given by

$$(\varphi_1 \otimes \varphi_2)(c) = \varphi_1(c^{\downarrow d(\varphi_1)}) \cdot \varphi_2(c^{\downarrow d(\varphi_2)})$$

for every configuration c of $d(\varphi_1) \vee d(\varphi_2)$, and we marginalize by summing out:

$$\varphi^{\downarrow x}(c) = \Sigma \{\varphi(c.c') \mid c' \text{ is a configuration of } d(\varphi)\text{-}x\}.$$

for each configuration c of x.

Here solving (5.1) and (5.7) is a matter of division. in the case of (5.1), we want to find ψ such that

$$\varphi^{\downarrow x}(c^{\downarrow x}) \cdot \psi(c) = \varphi^{\downarrow y}(c) \qquad (5.9)$$

for every configuration c of y. This is easily done. If $\varphi^{\downarrow x}(c^{\downarrow x}) = 0$, then $\varphi^{\downarrow y}(c) = 0$, so any value for $\psi(c)$ will satisfy (5.9). If $\varphi^{\downarrow x}(c^{\downarrow x}) > 0$, on the other hand, then we must have

$$\psi(c) = \frac{\varphi^{\downarrow y}(c)}{\varphi^{\downarrow x}(c^{\downarrow x})}. \qquad (5.10)$$

So we use (5.10) to define $\psi(c)$ for c such that $\varphi^{\downarrow x}(c^{\downarrow x}) > 0.$, and we define $\psi(c)$ arbitrarily for c such that $\varphi^{\downarrow x}(c^{\downarrow x}) = 0$. We handle (5.7) similarly; we set

$$\chi(c) = \frac{\varphi^{\downarrow b \wedge t}(c)}{\varphi_t^{\downarrow b \wedge t}(c)}.$$

for every configuration c of $b \wedge t$ such that the denominator is non-zero. It follows from Lemma 5.6 that if the denominator is zero, then the numerator is also, so that we can define $\chi(c)$ arbitrarily.

In this example, continuers give conditional probabilities. More precisely, if the denominator of (5.10) is not zero, then the ratio is the conditional probability of c given $c^{\downarrow x}$, according to the probability distribution that is proportional to ψ.

In the case where φ does not take any zero values, continuers and normalizations are unique. All the values of a continuer are conditional probabilities, and a normalization on y is simply the marginal on y of the probability distribution that is proportional to φ. When φ does have some zero values, continuers and normalizations may not be completely unique, because some of their values may be arbitrary. In the extreme case, where φ is identically zero, the continuers and normalizations are completely arbitrary.

In this example, a continuer ψ from x to y is proper if

$$\Sigma \{\varphi(c_x.c_{y-x}) \mid c_{y-x} \text{ is a configuration of } y\text{-}x\} = 1 \qquad (5.11)$$

for each configuration c_x of x. When $\varphi^{\downarrow x}(c_x) > 0$, this holds, because then the $\varphi(c_x.c_{y-x})$ are the conditional probabilities, according to the probability distribution proportional to φ, of c_{y-x} given c_x If $\varphi^{\downarrow x}(c_x) = 0$, then we can define arbitrarily $\varphi(c_x.c_{y-x})$, and hence we can make (5.11) hold if we wish. If we are interested only in finding marginals for, however, there will be no reason to go to the trouble of doing so.

5.3.3. *Linear Equations*

In this example, (5.1) and (5.7) are both trivial to solve in a formal sense; $\varphi^{\downarrow y}$ is a solution of (5.1), and $\varphi^{\downarrow b \wedge t}$ is a solution of (5.7).

This means that the ideas of Section 5.2 simplify, and the distinctions among the three computational approaches more or less disappear. The simplest way to describe the computation is to say that we save only potentials on the domains (not on the separators) as we move inward. After we have found the marginal on the root h_1, we marginalize it to $h_1 \wedge h_2$, and we combine this with the potential we have on h_2 to obtain the marginal for h_2, and so forth.

5.3.4. *Subsets of Cartesian Products*

Our comments on the preceding example apply to this one as well.

5.3.5. *Belief Functions*

In the theory of belief functions, combination can be implemented by multiplying q-functions.

$$q_{\varphi_x \otimes \varphi_y}(A) = q_{\varphi_x}(A^{\downarrow x}) \cdot q_{\varphi_y}(A^{\downarrow y})$$

This suggests that we can define continuers for commonality functions just as in the preceding example. In other words, we solve (5.1) by means of a q-function q_ψ on y with

$$q_\psi(A) = \frac{q_{\varphi^{\downarrow y}}(A)}{q_{\varphi^{\downarrow x}}(A^{\downarrow x})} \tag{5.12}$$

for each subset A of Θ_x such that $q_{\varphi^{\downarrow x}}(A^{\downarrow x}) > 0$.

As it turns out, we can use (5.12) to define a non-negative function q_ψ that satisfies

$$q_{\varphi^{\downarrow x}}(A^{\downarrow x}) \cdot q_\psi(A) = q_{\varphi^{\downarrow y}}(A). \tag{5.13}$$

If $q_{\varphi^{\downarrow x}}(A^{\downarrow x}) = 0$, then $q_{\varphi^{\downarrow y}}(A) = 0$, so (5.13) will hold no matter how we define q_ψ for A such that $q_{\varphi^{\downarrow x}}(A^{\downarrow x}) = 0$. However, q_ψ may fail to be a q-function.

We can legitimize (5.12), however, by enlarging the commutative semigroup of belief functions to a larger commutative semigroup. The belief functions on y form a semigroup Φ_y, which consists, in its q-function representation, of all set functions q on Θ_y for which the numbers

$$m(A) = \sum_{B \supseteq A} (-1)^{|B-A|} q(B)$$

are non-negative and add to one. Every element of Φ_y is itself non-negative, but not every non-negative set function q on Θ_y is in Φ_y. If we let Φ_y^* be the commutative semigroup consisting of all non-negative set functions on Θ_y, still with multiplication as the semigroup operation, then Φ_y will be a subsemigroup of Φ_y^*, and (5.12) will define a legitimate element of Φ_y^*. If we use the same definition of marginalization for Φ_y^* as we used for Φ_y, then we have a more general theory of belief functions, in which

Axioms A1-A7 all hold. We can implement the algorithm of Section 5.2 in this more general theory.

A normalization of φ on y, in this general theory, will be a set function q_ψ satisfying

$$q_{\varphi}{\downarrow}\varnothing \cdot q_{\psi}(A) = q_{\varphi}{\downarrow}y(A). \tag{5.14}$$

Since $q_{\varphi}{\downarrow}\varnothing$ is a constant, we see that here, as in the probability example, normalizations are essentially the same as marginals. If $q_{\varphi}{\downarrow}\varnothing=0$, then q_φ is identically zero. Whenever is not identically zero, we can rewrite (5.14) as

$$q_{\psi}(A) = \frac{q_{\varphi}{\downarrow}y(A)}{q_{\varphi}{\downarrow}\varnothing}.$$

Thus the normalization φ on y is the proper belief function proportional to the marginal of φ on y.

If the belief function φ is in the subsemigroup Φ, then its marginals and normalizations will be as well, and hence the final results of the algorithms described in Section 5.2 will be in Φ, even if some or all of the factors or some or all of the continuers used in the computation are not.

6. Extension and Solution

Normalizations sometimes provide more information than we need. In the case of dynamic programming, for example, we may not want to know how much each configuration of y falls short of being optimal. We only want to know which configurations of y are optimal. In this chapter, we generalize the computational theory of the preceding chapter so that it will apply to cases where we seek less information.

6.1. Extension Relations

We adopt Axioms A1-A3 and M1-M2. We also assume the existence of an extension relation.

We call a set E of ordered triplest (x,ψ,φ) an extension relation if Axioms A1-A3 and M1-M2 hold and E satisfies the following axioms:

Axiom E1. If $(x,\psi,\varphi)\varepsilon E$, then φ and ψ are in Φ, x is in $\Im$, and $x \vee d(\psi) = d(\varphi)$.

Axiom E2. If $w \le x \le d(\varphi)$, $(w,\psi_1,\varphi^{\downarrow x})\varepsilon E$, and $(x,\psi_2,\varphi)\varepsilon E$, then $(w,\psi_1\otimes\psi_2,\varphi)\varepsilon E$.

Axiom E3. If φ_x and φ_y are potentials on x and y, respectively, and $(x\wedge y,\psi,\varphi_y)\varepsilon E$, then $(x,\psi,\varphi_x\otimes\varphi_y)\varepsilon E$.

Axiom E4. If $x \le d(\varphi)$, then there is at least one potential ψ such that $(x,\psi,\varphi)\varepsilon E$.

Axiom E5. If $w \le x \le d(\varphi)$ and $(w,\psi,\varphi)\varepsilon E$, then $(w,\psi^{\downarrow x\wedge d(\psi)},\varphi^{\downarrow x})\varepsilon E$.

The most important axioms here are E2 and E3. Axiom E2 is the combination axiom for extenders. Axiom E3 is the lifting axiom for extenders.

[NOTE: Should I strengthen Axiom E5 by requiring that any extender ψ^* of ψ from $x \wedge d(\psi)$ to $d(\psi)$ should also be an extender of φ from x to $d(\varphi)$ and should satisfy $\psi = \psi^* \otimes \psi^{\downarrow x \wedge d(\psi)}$? I think these things are true in the examples.]

We say that ψ extends φ from x to y (or that ψ is an extender of φ *from* x to y) whenever $x \leq y \leq d(\varphi)$ and $(x,\psi,\varphi^{\downarrow y}) \varepsilon E$. (If $d(\varphi) = y$, this condition reduces to $(x,\psi,\varphi) \varepsilon E$.)

Continuers provide one example of an extension relation. More precisely:

Lemma 6.1. If Axiom T holds, and we set

$$E = \{(x,\psi,\varphi) \mid \psi \text{ continues } \varphi \text{ from } x \text{ to } d(\varphi)\},$$

then E is an extension relation.

On the other hand, extension relations have all the properties that we used in the last chapter's computational theory for continuers. Indeed, we can prove the following lemma, which differs from Lemma 5.1 only in that "continues" has been replaced by "extends."

Lemma 6.2. Suppose we are given an extension relation. Then the following statements hold.

(i) If ψ extends φ from x to y, then $x \vee d(\psi) = y$.

(ii) If $w \leq x \leq y \leq z$, φ is a potential on z, and ψ extends φ from w to x, then ψ extends $\varphi^{\downarrow y}$ from w to x.

(iii) If $w \leq x \leq y \leq z$, φ is a potential on z, and ψ extends φ from w to y, then $\psi^{\downarrow x \wedge d(\psi)}$ extends φ from w to x.

(iv) If $w \leq x \leq y \leq z$, φ is a potential on z, ψ_1 extends φ from w to x, and ψ_2 extends φ from x to y, then $\psi_1 \otimes \psi_2$ extends φ from w to y.

(v) If φ_x and φ_y are potentials on x and y, respectively, and ψ extends φ_y from $x \wedge y$ to y, then ψ also extends $\varphi_x \otimes \varphi_y$ from x to $x \vee y$.

(vi) If φ_x and φ_y are potentials on x and y, respectively, and ψ extends φ_y from $x \wedge y$ to y, then ψ also extends $\varphi_x \otimes \varphi_y$ from $x \wedge y$ to y.

Proof: Statements (i) through (v) follow from the definition of "extends" together with Axioms E1, M1, E5, E2, and E3, respectively.

To prove (vi), note that by (v), ψ extends $\varphi_x{}^{\downarrow x \wedge y} \otimes \varphi_y$ from $x \wedge y$ to y. Since $\varphi_x{}^{\downarrow x \wedge y} \otimes \varphi_y = (\varphi_x \otimes \varphi_y)^{\downarrow y}$, this means that ψ extends $(\varphi_x \otimes \varphi_y)^{\downarrow y}$ and hence $\varphi_x \otimes \varphi_y$ from $x \wedge y$ to y.

If ψ is an extender of φ from $\varnothing$ to y, where $\varnothing$ is the zero of $\mathfrak{I}$, then we call ψ a solution of φ on y.

6.2. Computational Theory

The computational theory for extenders follows from Lemma 6.2 exactly as the computational theory for extenders followed from Lemma 5.1. We begin, of course, by modifying appropriately our concept of a feasible domain. A domain is now feasible if it is feasible to combine and marginalize potentials and to find extenders within the domain.

Here is the analogue of Lemma 5.3.

Lemma 6.3. Suppose φ is a product on a hypergraph H:

$$\varphi = \otimes\{\varphi_h | h \varepsilon H\}.$$

Suppose that t is a twig in H, and b is a branch for t. And suppose that ψ extends φ_t from $b \wedge t$ to t. Then

(i) ψ extends φ from $b \wedge t$ to t, and

(ii) ψ extends φ from vH^{-t} to vH.

Proof. Statements (i) and (ii) are special cases of statements (vi) and (v) of Lemma 6.2, respectively.

Suppose H is a feasible hypertree, and we find $\varphi^{\downarrow h_1}$ by the step-by-step method described in Chapter 3. And at each step, we also find an extender. As we work within h_i to find the marginal of $\varphi_{h_i}{}^i$ on $h_{b(i)} \wedge h_i$, we also find an extender of $\varphi_{h_i}{}^i$ from $h_{b(i)} \wedge h_i$ back to h_i. Let us write ψ_i for this extender.

Lemma 6.4. The potential ψ_i extends φ from $h_{b(i)} \wedge h_i$ to h_i and also from vH^{i-1} to vH^i, for i=2,...,n.

Proof. By Lemma 6.3, ψ_i extends $\varphi^{\downarrow vH^i}$ from $h_{b(i)} \wedge h_i$ to h_i and from vH^{i-1} to vH^i. So by statement (ii) of Lemma 6.2, it extends φ from $h_{b(i)} \wedge h_i$ to h_i and from vH^{i-1} to vH^i.

At the last step, we find $\varphi^{\downarrow h_1}$, together with ψ_2, which extends φ from $h_1 \wedge h_2$ to h_2. Then we continue one more step, finding $\varphi^{\downarrow \varnothing}$ and a potential ψ_1 that extends φ from $\varnothing$ to h_1.

We now have

ψ_1, which extends φ from $\varnothing$ to h_1,

ψ_2, which extends φ from h_1 to $h_1 v h_2$,

ψ_3, which extends φ from $h_1 v h_2$ to $h_1 v h_2 v h_3$,

and so on, up to

ψ_n, which extends φ from $h_1 v h_2 v \cdots v h_{n-1}$ to $h_1 v h_2 v \cdots v h_n$.

By statement (iv) of Lemma 6.2, $\psi_1 \otimes \psi_2 \otimes \cdots \otimes \psi_n$ extends φ from $\varnothing$ to $h_1 v h_2 v \cdots v h_n$—i.e., it is a solution of φ on $h_1 v h_2 v \cdots v h_n$.

In some cases, as shall see, the extenders ψ_i may be simple enough that it is feasible to compute $\psi_1 \otimes \psi_2 \otimes \cdots \otimes \psi_n$ directly. In general,this will not be possible, but we can work back out in the hypertree to compute solutions of φ on the individual h_i. To see how to do this, we shift our attention to another aspect of the ψ_i:

ψ_1 extends φ from $\varnothing$ to h_1, (6.1)

ψ_2 extends φ from $h_1 \wedge h_2$ to h_2, (6.2)

ψ_3 extends φ from $h_{b(3)} \wedge h_3$ to h_3, (6.3)

and so on, up to

ψ_n extends φ from $h_{b(n)} \wedge h_n$ to h_n.

Using these facts, we can find solutions of φ on the individual h_i simply by marginalizing and combining the ψ_i within the h_i. This is our outward phase of hypertree computation.

We begin with ψ_1. By (6.1), together with statement (iii) of Lemma 6.2, its marginal $\psi_1{\downarrow}^{h_1 \wedge h_2}$ extends φ from $\varnothing$ to $h_1 \wedge h_2$. We can compute this extender within the domain h_1, and then we can combine it with ψ_2 within the domain h_2; this yields, by (6.2) and statement (iv) of Lemma 6.2, an extender of φ from $\varnothing$ to h_2.

Now we have solutions on h_1 and h_2 (i.e., extenders of φ from $\varnothing$ to h_1.and from $\varnothing$ to h_2). The next step is to find a solution on h_3 (i.e., an extender of φ from $\varnothing$ to h_3). We do this by taking the solution on h_1 or h_2, whichever is the branch of h_3, and marginalizing it to $h_{b(3)} \wedge h_3 \wedge d(\psi_3)$. By statement (iii) of Lemma 6.2, this yields an extender of φ from $\varnothing$ to $h_{b(3)} \wedge h_3$. So by (6.3) and statement (iv) of Lemma 6.2, we can combine it with ψ_3 to obtain an extender of φ from $\varnothing$ to h_3.

Continuing in this way back out the tree, we successively obtain *solutions* on each h_i, working in each case only in the domains h_i and $h_{b(i)}$. The computation at each step is essentially the same as the computation we performed at each step during the inward phase of the computation: marginalization to the meet with a neighbor and then combination with that neighbor, except that now we move from $h_{b(i)}$ to h_i instead of from h_i to $h_{b(i)}$.

6.3. Example

Since continuation is a special case of extension, all the examples of the preceding chapter serve as examples of the theory of this chapter. It is more interesting, however, to see how the theory of this chapter applies to the problem of finding optimal configurations in the example of dynamic programming.

Suppose $x \leq y \leq z$, $x \vee w = y$, and φ is a potential on z. Suppose λ is a map from $\Theta_{x \wedge w}$ to Θ_{y-x} such that for every configuration c of x,
$\varphi{\downarrow}^{y}(c.\lambda(c{\downarrow}^{x \wedge w})) = \max\{\varphi{\downarrow}^{y}(c.c^*) \mid c^* \text{ is a configuration of } y\text{-}x\}$.
Then we say that the potential ψ on w given by

$$\psi(c) = \begin{cases} 1 \text{ if } c{\downarrow}^{y-x} = \lambda(c{\downarrow}^{x \wedge w}) \\ 0 \text{ otherwise} \end{cases}$$

extends φ from x to y.

7. Reduction

We now return to the inward phase of hypertree computation, in order to generalize the theory of Chapter 3 from the case where unique reductions (marginals) exist to the case where a variety of reductions may be possible.

One motivation for this generalization is the example of linear equations, which we studied in Section 3.4.3. As we saw in there, we can define a marginal in this example, but doing so puts our theory at an awkward distance from the example's computational reality. Here we get closer to that reality.

7.1. Reduction Relations

We adopt Axioms A1-A2. We do not adopt Axiom A3 or Axioms M1-M2. Instead we assume the existence of a reduction relation.

We call a set R of ordered pairs (φ_0,φ) a reduction relation if Axioms A1-A2 hold and R satisfies the following axioms:

Axiom R1. If $(\varphi_0,\varphi)\varepsilon R$, then $d(\varphi_0) \leq d(\varphi)$.

Axiom R2. If $(\varphi_1,\varphi_2)\varepsilon R$ and $(\varphi_2,\varphi_3)\varepsilon R$, then $(\varphi_1,\varphi_3)\varepsilon R$.

Axiom R3. Suppose φ_x and φ_y are potentials on x and y, respectively, φ_0 is a potential on $x \wedge y$, and $(\varphi_0,\varphi_y)\varepsilon R$. Then $(\varphi_x\otimes\varphi_0,\varphi_x\otimes\varphi_y)\varepsilon R$.

Axiom R4. If $x \leq y$ and φ is a potential on y, then $(\varphi_0,\varphi)\varepsilon R$ for some potential φ_0 on x.

Notice that Axioms R2 and R3 correspond to the Axioms M1 and M2, respectively; R2 is the transitivity axiom, and R3 is the combination axiom.

When $(\varphi_0,\varphi)\varepsilon R$, we say that φ_0 is a reduction of φ. If $d(\varphi_0) = x$, we say also say that φ_0 is a reduction of φ to x.

Axioms R1-R4 clearly constitute a generalization of Axioms M1-M2. If Axioms A1-A2 and M1-M2 are satisfied, then we get a reduction relation by saying that φ_1 is a reduction of φ_2 if and only if it is a marginal of φ_2. On the other hand, if Axioms A1-A2 and R1-R4 are satisfied, and for each potential φ and each subdomain x of $d(\varphi)$, there is only one reduction of φ to x, then we can write $\varphi^{\downarrow x}$ for this reduction, and Axioms M1-M2 will be satisfied.

7.2. Vacuous Extension

Here, as in Chapter 3, if φ_x is a potential on x, and $x \leq y$, we write $\varphi_x{}^{\uparrow y}$ for the product $\varphi_x\otimes\iota_y$, and we call $\varphi_x{}^{\uparrow y}$ the *vacuous extension* of φ_x to y. Since we have retained the labelling and identity axioms, we still have all the conclusions from Lemma 3.2:

(i) If $x \leq y$, then $\iota_x{}^{\uparrow y} = \iota_y$.

(ii) For any domains x and y, $\iota_x\otimes\iota_y = \iota_{x\vee y}$.

(iii) If φ_x is a potential on x, then $\varphi_x\otimes\iota_y = \varphi_x{}^{\uparrow x\vee y}$.

(iv) If φ_x is a potential on x, then $\varphi_x{}^{\uparrow x} = \varphi_x$.

(v) If φ_x is a potential on x, and $y \leq x$, then $\varphi_x\otimes\iota_y = \varphi_x$.

(vi) If φ_x is a potential on x, and $x \leq y \leq z$, then $(\varphi_x{}^{\uparrow y})^{\uparrow z} = \varphi_x{}^{\uparrow z}$.

(vii) If φ_x and φ_y are potentials on x and y, respectively, then $\varphi_x\otimes\varphi_y = (\varphi_x{}^{\uparrow x\vee y})\otimes(\varphi_y{}^{\uparrow x\vee y})$.

(viii) If φ_x is a potential on x, then $(\varphi_x{}^{\uparrow x\vee y})^{\downarrow y} = (\varphi_x{}^{\downarrow x\wedge y})^{\uparrow y}$.

Moreover, we have the following generalization of Lemma 3.3:

Lemma 7.1. If φ_y is a potential on y, and φ is a reduction of φ_y to $x\wedge y$, then $\varphi^{\uparrow x}$ is a reduction of $\varphi_y{}^{\uparrow x\vee y}$ to x.

Proof: Since $\varphi\uparrow^{x} = \iota_x \otimes \varphi$ and $\varphi_y\uparrow^{x \vee y} = \iota_x \otimes \varphi_y$, the lemma is merely the result of substituting ι_x for φ_x in Axiom R3.

7.3. Computational Theory

With Axioms A1-A2 and R1-R4, we can develop essentially the same computational theory as with Axioms A1-A2 and M1-M2. Here, as in Chapter 3, we assume that it is not computationally difficult to find labels or to implement the lattice operations. The computational difficulties lie in finding reductions and perhaps in implementing the semigroup operation. A *feasible* domain is one in which we can express a potential explicitly, find its product with other potentials on the same domain, and find its reductions to subdomains.

The combination axiom has the same significance as in Chapter 3. If x and y is feasible, then we can find a reduction of $\varphi_x \otimes \varphi_y$ to x. We first find a reduction of φ_y to $x \wedge y$ (this only requires us to work in y) and then we combine it with φ_x (this only requires us to work in x).

Next, we can reduce $\varphi_x \otimes \varphi_y$ to a subdomain of x.

Lemma 7.2. Suppose $z \leq x$, φ_x and φ_y are potentials on x and y, respectively, φ is a reduction of φ_y to $x \wedge y$, and φ' is a reduction of $\varphi_x \otimes \varphi$ to z. Then φ' is also a reduction of $\varphi_x \otimes \varphi_y$ to z.

Proof: By the combination axiom, $\varphi_x \otimes \varphi$ is a reduction of $\varphi_x \otimes \varphi_y$ to x. So by the transitivity axiom, φ' is a reduction of $\varphi_x \otimes \varphi_y$ to z.

And we can take advantage of twigs.

Lemma 7.3. Suppose $\{\varphi_h\}_{h \varepsilon H}$ is a collection of potentials on the hypergraph H, t is a twig in H, and b is a branch for t. Suppose φ is a reduction of φ_t to $t \wedge b$. Then

(i) $(\otimes\{\varphi_h | h \varepsilon H^{-t}\}) \otimes \varphi$ is a reduction of $\otimes\{\varphi_h | h \varepsilon H\}$ to $\vee H^{-t}$.

Moreover, if we set $\varphi_b{}^{-t} = \varphi_b \otimes \varphi$ and $\varphi_h{}^{-t} = \varphi_h$ for all other h in H^{-t}, then

(ii) $\otimes\{\varphi_h{}^{-t} | h \varepsilon H^{-t}\}$ is a reduction of $\otimes\{\varphi_h | h \varepsilon H\}$ to $\vee H^{-t}$, and

(iii) if $z \leq \vee H^{-t}$ and φ' is a reduction of $\otimes\{\varphi_h{}^{-t} | h \varepsilon H^{-t}\}$ to z, then φ' is also a reduction of $\otimes\{\varphi_h | h \varepsilon H\}$ to z.

Proof: By the labelling axiom, $\otimes\{\varphi_h{}^{-t} | h \varepsilon H^{-t}\}$ is a potential on $\vee H^{-t}$. So (i) follows directly from the combination axiom. To prove (ii), we merely rearrange the factors on the right-hand side of (i). Then we get (iii) by the transitivity axiom.

Here, as in Lemma 3.4, statement (iii) shows how to simplify the the task of reducing a product on H to the smaller task of reducing a product on $\vee H^{-t}$, while statement (ii) opens the way to exploiting the successive removal of twigs.

Suppose, indeed, that H is a feasible hypertree, and we want to compute $\varphi\downarrow^{x}$, where $\varphi = \otimes\{\varphi_h | h \varepsilon H\}$, and x is a subdomain of some domain in H, say h_1. We can do this following the same steps that we followed in Chapter 3. we choose a hypertree construction ordering for H that begins with h_1, say $h_1,h_2,\ldots,h_n$, and we choose a branching b(i) for this construction ordering. For i = 1,2,...,n, we set

$$H^i = \{h_1,h_2,\ldots,h_i\}.$$

This is a sequence of hypertrees, each larger than the last; $H^1 = \{h_1\}$ and $H^n = H$. The domain h_i is a twig in H^i. So we can work backwards in this sequence, using the idea of Lemma 7.3 each time. At the step from H^i to H^{i-1}, we go from a collection $\{\varphi_h{}^i|h\varepsilon H^i\}$, say, which has a reduction of $\varphi^{\vee H^i}$ to $\vee H^i$ as its product, to a collection $\{\varphi_h{}^{i-1}|h\varepsilon H^{i-1}\}$, which has a reduction of $\varphi^{\vee H^{i-1}}$ to $\vee H^{i-1}$ its product. To go from $\{\varphi_h{}^i|h\varepsilon H^i\}$ to $\{\varphi_h{}^{i-1}|h\varepsilon H^{i-1}\}$, we omit h_i and change the potential on $h_{b(i)}$ from $\varphi_{h_{b(i)}}{}^i$ to

$$\varphi_{h_{b(i)}}{}^i \otimes (\varphi_{h_i}{}^i)^{h_{b(i)} \wedge h_i},$$

where $(\varphi_{h_i}{}^i)^{h_{b(i)} \wedge h_i}$ is a reduction of $\varphi_{h_i}{}^i$ to $h_{b(i)} \wedge h_i$. The collection with which we begin, $\{\varphi_h{}^n|h\varepsilon H^n\}$, is simply $\{\varphi_h|h\varepsilon H\}$, and the collection with which we end, $\{\varphi_h{}^1|h\varepsilon H^1\}$, consists of the single potential, which is a reduction of φ to h_1. One more marginalization within h_1 reduces this to what we want, $(\otimes\{\varphi_h|h\varepsilon H\})^{\downarrow x}$.

It follows that if $H \cup \{x\}$ is a feasible hypertree, then we find a reduction to x of $\otimes\{\varphi_h|h\varepsilon H\}$. And by using vacuous extension and identities, we can draw the same conclusion in the more general case where $H \cup \{x\}$ has a feasible hypertree cover.

7.4. Example

Since reduction is a generalization of marginalization, all the examples of marginalization in Section 3.4 are examples of reduction. The theory of reduction allows us to improve one of the examples, however—the example of sparse linear equations. Indeed, if we are working with Axioms R1-R6, then we do not have to resort to equivalence classes of systems of linear equations. Instead, we move closer to the computational reality by taking the systems themselves as the potentials.

Here is how it goes. A potential on x is a pair (A,x), where A is a finite (possibly empty) set of equations of the form

$$\sum_{\xi \varepsilon w} a_\xi \xi + a = 0,$$

where $w \leq x$, and a and the a_ξ are real numbers. (If $x = \varnothing$, then A can contain only equations of the form a = 0.) If a configuration c of x satisfies all the equations in A, then we call c a solution of (A,x).

We combine potentials by taking their union: $(A,x)\otimes(B,y) = (A \cup B, x \vee y)$. Since $\cup$ and $\vee$ are commutative and associative, $\otimes$ is as well, and Axiom A1 holds by definition. The identity ι_x is the potential $(\varnothing,x)$, and Axiom A2 also holds.

If (A,x) and (B,y) are potentials, $y \leq x$, and $c^{\downarrow y}$ is a solution of (B,y) whenever c is a solution of (A,x), then we say that (B,y) is a reduction to y of (A,x). Thus Axiom R1 holds by definition.

We find reductions, of course, by eliminating variables. Axiom R2 is one aspect of the fact that the order in which we eliminate variables does not affect the solutions of the variables that remain. Axiom R3 holds because when we eliminate variables, we do not change equations in which these variables do not appear. Indeed, suppose (A,x) and (B,y) are potentials on x and y, respectively. We obtain a reduction $(B',y \wedge x)$ of (B,y) by eliminating the variables in y-x from all the equations in B. The potential $(A \cup B',y)$ will then be a reduction of $(A \cup B, x \vee y)$, because the same steps will also eliminate these variables from $A \cup B$. Axiom R4 holds because we can always eliminate variables.

If (A,x) has solutions, then $(\varnothing,\varnothing)$ and $(\{0 = 0\},\varnothing)$ will be reductions of (A,x) to $\varnothing$. If (A,x) does not have any solutions, then any reduction of (A,x) to $\varnothing$ will contain false equations (such as 6 = 0). Thus we will find out whether a potential has solutions by reducing it to $\varnothing$.

8. Resolution

This chapter combines the idea of reduction (Chapter 7) with the idea of extension (Chapter 8). It simulataneously axiomatizes these two ideas.

The axioms given here are the most general of this paper, except that we do require modularity and the existence of a zero, which are not, as we saw in Chapters 3 and 7, required for reduction alone. All the examples we studied in Chapter 3 satisfy the general axioms given here. These axioms also allow us to deal straightforwardly with examples such as systems of linear equations, where neither reduction nor extension are unique, but where they must be coordinated.

8.1. Axioms

We adopt Axioms A1-A3. We do not adopt Axioms M1-M2. Instead we assume the existence of a resolution relation.

We call a set S of ordered triplets (φ_0,ψ,φ) a resolution relation if Axioms A1-A3 hold and S satisfies the following axioms:

Axiom S1. If $(\varphi_0,\psi,\varphi)\varepsilon S$, then $d(\varphi) = d(\varphi_0) \vee d(\psi)$.

Axiom S2. If $(\varphi_1,\psi_1,\varphi_2)\varepsilon S$ and $(\varphi_2,\psi_2,\varphi_3)\varepsilon S$, then $(\varphi_1,\psi_1\otimes\psi_2,\varphi_3)\varepsilon S$.

Axiom S3. If $(\varphi_1,\psi,\varphi)\varepsilon S$ and φ_2 is a potential such that $d(\varphi_2)\wedge d(\varphi) \leq d(\varphi_1)$, then $(\varphi_1\otimes\varphi_2,\psi,\varphi\otimes\varphi_2)\varepsilon S$.

Axiom S4. If $x \leq y$ and φ is a potential on y, then there exist potentials φ_0 and ψ (possibly not unique) such that $d(\varphi_0)=x$ and $(\varphi_0,\psi,\varphi)\varepsilon S$.

Axiom S5. If $(\varphi_0,\psi,\varphi)\varepsilon S$ and $d(\varphi_0) \leq x \leq d(\varphi)$, then there exist potentials ψ_1, φ_1, and ψ_2 (possibly not unique) such that $d(\varphi_1) = x$, $(\varphi_0,\psi_1,\varphi_1)\varepsilon S$ and $(\varphi_1,\psi_2,\varphi)\varepsilon S$.

Axioms S2 and S3 are the crucial axioms here. Axiom S2 expresses the transitivity axiom for reduction and the combination axiom for extension. Axiom S3 expresses the combination axiom for reduction and the lifting axiom for extension. For simplicity, we will call Axiom S2 the axiom of combination for extension, and we will call Axiom S3 the axiom of combination for reduction.

[NOTE: Should I strengthen Axiom S5 by requiring that $(\psi_1,\psi_2,\psi)\varepsilon S$ and/or that $\psi = \psi_1 \otimes \psi_2$?]

The idea of a resolution relation generalizes both the idea of a reduction relation and the idea of an extension relation. If we omit mention of the extender, then the resolution relation becomes a reduction relation. If this reduction relation has unique reductions (marginals), then the extenders constitute an extension relation. On the other hand, if we have marginals with an extension relation, then they together form a resolution relation. This is made precise by the next few lemmas.

First, if we omit mention of the extenders, the resolution relation becomes a reduction relation:

Lemma 8.1. Suppose S is a resolution relation—i.e., a set of ordered triplets satisfying Axioms S1-S5. Then the binary relation

$$\{ (\varphi_0,\varphi) \mid (\varphi_0,\psi,\varphi) \text{ is in S for some } \psi \}$$

is a reduction relation—i.e., it satisfies Axioms R1-R4.

Proof: Axioms R1-R4 follow directly from Axioms S1-S4, respectively.

Next, if the reduction is a marginalization, then the extenders constitute an extension relation:

Lemma 8.2. Suppose S is a resolution relation—i.e., a set of ordered triplets satisfying Axioms S1-S5. Suppose further that the reduction relation defined in Lemma 8.1 is a marginalization—i.e., for each potential φ and each subdomain x of $d(\varphi)$, there is only one potential φ_0 on x such that φ_0 is a reduction of φ to x. If we write $\varphi^{\downarrow x}$ for this unique reduction, then the relation

$$\{ (x,\psi,\varphi) \mid x \leq y,\ d(\varphi)=y, \text{ and } (\varphi^{\downarrow x},\psi,\varphi) \text{ is in S for some } \psi \}$$

is an extension relation—i.e., it satisfies Axioms E1-E5.

Finally, if a marginalization and an extension relation determine a resolution relation:

Lemma 8.3. Suppose Axioms A1-A3 and M1-M2 are satisfied, and suppose E is an extension relation. Then

$$\{ (\varphi^{\downarrow x},\psi,\varphi) \mid (x,\psi,\varphi)\varepsilon E \}$$

is a resolution relation.

Proof:

With these lemmas in mind, we will talk freely about reduction and extension when we are working with a resolution relation. Whenever $(\varphi_0,\psi,\varphi)\varepsilon S$, we will say that (φ_0,ψ) is a resolution of φ to $d(\varphi_0)$, φ_0 is a reduction of φ to $d(\varphi_0)$, and ψ is an extender from φ_0 to φ.

We call a resolution of φ to $\varnothing$ a solution of φ. This differs slightly from the usage of Chapter 6. There a solution was an extension from $\varnothing$. Here a solution is a pair, a reduction to $\varnothing$ together with an extension from $\varnothing$.

8.2. Computational Theory

Here we combine what we have learned in Chapters 6 and 7.

A feasible domain is now a domain in which we can implement combination and also find the resolutions whose existence is guaranteed by Axioms S4 and S5. A feasible hypertree is one all of whose domains are feasible.

Suppose H is a feasible hypertree, and we want to find a solution of φ on x, where $\varphi = \otimes\{\varphi_h | h \varepsilon H\}$, and x is a subdomain of some domain in H, say h_1. As usual, we choose a hypertree construction ordering $h_1, h_2, \ldots, h_n$ and a branching b(i). We set $H^i = \{h_1, h_2, \ldots, h_i\}$, so that h_i is a twig in H^i. We work backwards from H^n to H^1.

On the first step, we use Axiom S4 in h_n to find a resolution $(\varphi_{h_n}{}^{h_{b(n)} \wedge h_n}, \psi_n)$ of φ_{h_n} to $h_{b(n)} \wedge h_n$. Then we form a collection $\{\varphi_h{}^{n-1} | h \varepsilon H^{n-1}\}$ by omitting h_n, changing the potential on $h_{b(n)}$ from $\varphi_{h_{b(n)}}$ to

$$\varphi_{h_{b(n)}} \otimes \varphi_{h_n}{}^{h_{b(n)} \wedge h_n},$$

and leaving the potentials on the other h_i unchanged. By Axiom S3, $(v\{\varphi_h{}^{n-1} | h \varepsilon H^{n-1}\}, \psi_n)$ is a resolution of φ. Continuing in this way, we obtain a sequence

$$\{\varphi_{h1}{}^1\} = \{\varphi_h{}^1 | h \varepsilon H^1\}, \{\varphi_h{}^2 | h \varepsilon H^2\}, \ldots, \{\varphi_h{}^n | h \varepsilon H^n\} = \{\varphi_h | h \varepsilon H\},$$

of collections of potentials, and a sequence

$$\psi_1, \psi_2, \quad , \psi_n,$$

of potentials such that for i=2,...,n, ψ_i extends $(\varphi_{h_i}{}^i)^{h_{b(i)} \wedge h_i}$ on $h_{b(i)} \wedge h_i$ to $\varphi_{h_i}{}^i$ on h_i, and also extends $v\{\varphi_h{}^{i-1} | h \varepsilon H^{i-1}\}$ on $h_1 v h_2 v \cdots v h_{i-1}$ to $v\{\varphi_h{}^i | h \varepsilon H^i\}$ on $h_1 v h_2 v \cdots v h_i$. (The first extender, ψ_1 extends a potential φ_0 on x to the potential $\varphi_{h1}{}^1$ on h_1.) By Axiom S2, $\psi_1 \otimes \psi_2 \otimes \cdots \otimes \psi_n$ extends φ_0 on x to φ on $h_1 v h_2 v \cdots v h_n$. More generally, $\psi_1 \otimes \psi_2 \otimes \cdots \otimes \varphi_i$ extends φ_0 to $v\{\varphi_h{}^i | h \varepsilon H^i\}$, and $\psi_i \otimes \psi_2 \otimes \cdots \otimes \varphi_n$ extends $v\{\varphi_h{}^i | h \varepsilon H^i\}$ to φ.

If $x = \varnothing$, then $(\varphi_0, \psi_1 \otimes \psi_2 \otimes \cdots \otimes \psi_n)$ is a solution of φ.

In the example we consider below, $\psi_1 \otimes \psi_2 \otimes \cdots \otimes \psi_n$ can be computed relatively easily. We now explain, however, how to find reductions to individual domains while working within domains.

References

Beeri, C., R. Fagin, D. Maier, and M. Yannakakis (1983). On the desirability of acyclic database schemes. Journal of the Association for Computing Machinery 30 479-513.

Berge, Claude (1973). Graphs and Hypergraphs. North-Holland.

Bertelè, Umberto, and Francesco Brioschi (1972). *Nonserial Dynamic Programming.* Academic Press.

Birkhoff, Garrett (1967). *Lattice Theory.* Third Edition. American Mathematical Society.

Cowell, R.G. (1991). Combinatorial enumeration on decomposable weight functions. Technical report, University College London.

Cowell, R.G., and A.P. Dawid (1991). Fast retraction of evidence in a probabilistic expert system. Technical report, University College London.

Dawid, A.P. (1991). Applications of a general propagation algorithm for probabilistic expert systems. Technical report, University College London.

Dechter, Rina, Avi Dechter, and Judea Pearl (1990). Optimization in constraint networks. Pp. 411-425 of *Influence Diagrams, Belief Nets, and Decision Analysis*, edited by R.M. Oliver and J.Q. Smith, Wiley.

Dempster, Arthur P. (1990). Normal belief functions and the Kalman filter. Unpublished manuscript.

Dempster, Arthur P., and Augustine Kong (1988). Uncertain evidence and artificial analysis. *Journal of Statistical Planning and Inference* **20** 355-368. Reprinted on pp. 522-528 of Shafer and Pearl (1990).

Golumbic, M.C. (1980). *Algorithmic Graph Theory and Perfect Graphs.* Academic Press.

Jensen, Finn V., Steffen L. Lauritzen, and Kristian Olesen (1990). Bayesian updating in causal probabilistic networks by local computations. *Computational Statistics Quarterly.*

Kong, Augustine (1986). Multivariate belief functions and graphical models. Doctoral Dissertation, Department of Statistics, Harvard University, USA.

Kong, Augustine (1987). Quasi Belief Functions. Technical Report, Department of Statistics, Harvard University, USA.

Lauritzen, Steffen L., and David J. Spiegelhalter (1988). Local computation with probabilities on graphical structures and their application to expert systems (with discussion). *Journal of the Royal Statistical Society, Series B* **50** 157-224. Reprinted on pp. 415-448 of Shafer and Pearl (1990).

Lauritzen, Steffen L. (1990). Propagation of probabilities, means and variances in mixed graphical association models. Report R 90-18, Institute for Electronic Systems, Department of Mathematics and Computer Science, University of Aalborg, Denmark.

Maier, D. (1983). *The Theory of Relational Databases.* Computer Science Press.

Ndilikilikesha, Pierre (1991). Potential influence diagrams. Working Paper, School of Business, University of Kansas USA.

Petrich, Mario (1973). *Introduction to Semigroups.* Charles E. Merrill Publishing Company. Columbus, Ohio.

Rose, D.J. (1970). Triangulated graphs and the elimination process. *Journal of Mathematical Analysis and Its Applications* **32** 597-609.

Shachter, Ross D., Stig K. Andersen, and Kim L. Poh (1990). Directed reduction algorithms and decomposable graphs. *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence.* July 27-29, Cambridge, Massachusetts.

Shafer, Glenn (1990). Perspectives on the theory and practice of belief functions. *International Journal of Approximate Reasoning* **4** 323-362.

Shafer, Glenn, and Judea Pearl, eds. (1990). *Readings in Uncertain Reasoning.* Morgan Kaufmann. San Mateo, California.

Shafer, Glenn, and Prakash Shenoy (1988). Local computation in hypertrees. Working Paper 201, School of Business, University of Kansas, USA.

Shafer, Glenn, and Prakash Shenoy (1990). Probability propagation. *Annals of Mathematics and Artificial Intelligence* **2** 327-352.

Shafer, Glenn, Prakash Shenoy, and Khaled Mellouli (1987). Propagating belief functions in qualitative Markov trees. *International Journal of Approximate Reasoning* **1** 349-400.

Shenoy, Prakash (1990). Valuation-based systems for Bayesian decision analysis, Working Paper 220, School of Business, University of Kansas, USA.

Shenoy, Prakash, and Glenn Shafer (1990). Axioms for probability and belief-function propagation. Pp. 169-198 of *Uncertainty in Artificial Intelligence 4,* edited by R.D. Shachter, T.S. Levitt, L.N. Kanal, and J.F. Lemmer. North-Holland. Reprinted on pp. 575-610 of Shafer and Pearl (1990).

The University of Kansas
School of Business Working Paper Series

Copies available from: School of Business
203 Summerfield
University of Kansas
Lawrence, Kansas 66045

No. 182 "Managing Uncertainty in Expert Systems, A Perspective for Auditing," Glenn Shafer and Rajendra Srivastava (January 1986).

No. 183 "The Construction of Probability Arguments," Glenn Shafer (January 1986).

No. 184 "Propogating Belief Functions with Local Computations," Prakash P. Shenoy and Glenn Shafer (February 1986).

No. 185 "The Empirical Content of the Normative Interpretation of Subjective Expected Utility," Glenn Shafer (February 1986).

No. 186 "Varieties of Computer Based Instruction for the Business Administration Classroom: An Emphasis on Case Courses," Marilyn L. Taylor and Charles F. Douds (April 1987).

No. 187 "A Geometric Insight into Simplex Tableaus," I. S. Chien and P. L. Yu (May 1987).

No. 188 "Habitual Domain Analysis and Effective Goal Setting for Better Performance," P. L. Yu and I. S. Chien (May 1987).

No. 189 "The Bayesian and Belief-Function Fomalisms: I. A General Perspective for Auditing," Glenn Shafer and Rajendra Srivastava (June 1987).

No. 190 "Propagating Belief Functions in Qualitative Markov Trees," Glenn Shafer, Prakash P. Shenoy and Khaled Melouli (June, 1987).

No. 191 "Linear Optimal Designs and Optimal Contingency Plans," Y. R. Lee, Y. Shi and P. L. Yu (June 1987).

No. 192 "Bayesian and Belief-Function Propagation," Glenn Shafer and Prakash Shenoy (April 1988).

No. 193 "Evidential Reasoning Using DELEIF," Debra K. Zarley, Yen-Teh Hsia, and Glenn Shafer (April 1988).

No. 194 "Comments on 'An Inquiry into Computer Understanding', by Peter Cheeseman. Prepared for Computational Intelligence, An International Journal," Glenn Shafer (April 1988).

No. 195 "Combining AI and OR," Glenn Shafer (April 1988).

No. 196 "On the Propagation of Beliefs in Networks Using the Dempster-Shafer Theory of Evidence," Khaled Mellouli (April 1988).

No. 197 "AUDITOR'S ASSISTANT: A Knowledge Engineering Tool for Audit Decision," Glenn Shafer, Prakash Shenoy, and Rajendra Srivastava (April 1988).

No. 198 "Studies on Finding Hypertree Covers for Hypergraphs," Zhang Lianwen (April 1988).

No. 199 "An Axiomatic Framework for Bayesian and Belief-function Propagation," Prakash P. Shenoy and Glenn Shafer (July 1988).

No. 200 "Probability Propagation," Glenn Shafer and Prakash P. Shenoy (August 1988).

No. 201 "Local Computation in Hypertrees," Glenn Shafer and Prakash P. Shenoy (August 1988).

No. 202 "Axioms for Constraint Propagation and Optimization Using Local Computation," Prakash P. Shenoy and Glenn Shafer (August 1988).

No. 203 "Valuation Systems: A Language for Knowledge-based Systems," Prakash P. Shenoy (August 1988).

No. 204 "Riciprocal Cost Allocation Model Reinvestigated," Rajendra P. Srivastava (September 1988).

No. 205 "A Matrix Approach to Signal Extraction for ARIMA Component Time Series Models," William R. Bell and Steven C. Hillmer (September 1988).

No. 206 "An Evidential Reasoning System," Debra K. Zarley (November, 1988).

No. 207 "Discrete Optimization Using Local Computation," Prakash P. Shenoy and Glenn R. Shafer.

No. 208 "Constraint Propagation," Prakash P. Shenoy and Glenn R. Shafer.

No. 209 "Axioms for Probability and Belief-Function Propagation," Prakash P. Shenoy and Glenn R. Shafer.

No. 210 "The Unity of Probability," Glenn Shafer (February 1989).

No. 211 "MacEvidence: A Visual Environment for Constructing and Evaluating Evidential Systems," Yen-Teh Hsia and Prakash P. Shenoy (March 1989).

No. 212 "Audit Risk: A Belief-Function Approach," Rajendra P. Srivastava, Prakash P. Shenoy and Glenn R. Shafer (April 1989).

No. 213 "On Spohn's Rule for Revision of Beliefs," Prakash P. Shenoy (July 1989).

No. 214 "Size, Liquidity, and the Cost of Equity," William L. Beedles (October 1989).

No. 215 "Studies on Hypergraphs (I) Hyperforests" Lianwen Zhang (January 1990).

No. 216 "Consistency in Valuation-Based Systems" Prakash P. Shenoy (February 1990).

No. 217 "On the History and Future of Statistics": 1. The Unity and Diversity of Probability; 2. Review of "The Empire of Chance"; 3. Review of "The Rise of Statistical Thinking," Glenn Shafer (February 1990).

No. 218 "Perspectives on the Theory and Practice of Belief Functions," Glenn Shafer (April 1990).

No. 219 "A Belief-Function Approach to Audit Risk: I. Structure of Audit Evidence and Its Aggregation," Rajendra P. Srivastava and Glenn R. Shafer (April 1990).

No. 220 "Valuation-Based Systems for Bayesian Decision Analysis," Prakash P. Shenoy (April 1990).

No. 221 "Valuation-Based Systems for Discrete Optimization," Prakash P. Shenoy (June 1990).

No. 222 "A Comparative Evaluation of 15-Second and 30-Second Transformational and Informational Television Commercials Over Repeated Exposures," Surendra N. Singh and Catherine A. Cole (June 1990).

No. 223 "A New Method for Representing and Solving Bayesian Decision Problems," Prakash P. Shenoy (September 1990).

No. 224 "Selecting Optimal Linear Production Systems in Multiple Criteria Environments," Yong Shi and Po-Lung Yu (October 1990).

No. 225 "Why Should Statisticians Be Interested in Artificial Intelligence?" Glenn Shafer (November 1990).

No. 226 "Valuation-Based Systems: A Framework for Managing Uncertainty in Expert Systems," Prakash P. Shenoy (March 1991).

No. 227 "Valuation Networks, Decision Trees, and Influence Diagrams: A Comparison," Prakash P. Shenoy (June 1991).

No. 228 "The Early Development of Mathematical Probability," Glenn Shafer (August 1991).

No. 229 "What is Probability?" Glenn Shafer (August 1991).

No. 230 "Can the Various Meanings of Probability be Reconciled?" Glenn Shafer (August 1991).

No. 231 "Response to the Discussion of Belief Functions," Glenn Shafer (August 1991).

No. 232 "An Axiomatic Study of Computation in Hypertrees," Glenn Shafer (August 1991).

No. 233 "Using Possibility Theory in Expert Systems," Prakash P. Shenoy (September 1991).

No. 234 "Using Dempster-Shafer's Belief-Function Theory in Expert Systems," Prakash P. Shenoy (September 1991).

No. 235 "Potential Influence Diagrams," Pierre Ndilikilikesha (September 1991).

No. 236 "Independence in Valuation-Based Systems," Prakash P. Shenoy (September 1991).