1 Introduction

There are many different formalisms for representing knowledge or information. However a few elementary features are found with every formalism. Knowledge or information comes in chunks or piecewise. These pieces of information are then somehow combined or aggregated to represent the whole of the information. Inference then usually means to extract from the whole of the knowledge the part relevant to a given problem or question. This can be called focusing of information. This observation leads naturally to considering an algebraic structure, composed of a set of "pieces of information" which can be manipulated by two operations: combination to aggregate pieces of information, and focusing to extract the part of information related to a certain domain. The goal of this book is to formally describe this algebraic structure and to study it in some detail.

Probably the most popular instance of such a structure comes from probability theory. A multidimensional probability density over some set of variables can often be decomposed into a product of densities, defined on smaller sets of variables. This factorization reflects the structure of conditional independence between variables. In many cases these independence structures can be represented graphically. Lauritzen and Spiegelhalter in a pioneering work (Lauritzen & Spiegelhalter, 1988) showed how such a factorization can be used to compute marginals of the multidimensional probability density by so-called "local computation". The main advantage of this approach is that the multidimensional density must never be computed in its full dimension, but that rather the computation can be carried out "locally" on the much smaller dimensions of the factors. Only this possibility may make computation feasible. Shenoy and Shafer (Shenoy, 1989; Shenoy & Shafer 1990) introduced for the first time an abstract, axiomatic system capturing the essence for the type of local propagation introduced by (Lauritzen & Spiegelhalter, 1988). They pointed out, that many other formalisms satisfy also the axioms needed for local propagation. The next well-known example of such a formalism is probably provided by belief functions in the sense of Dempster-Shafer theory of evidence. In numerical analysis, sparse matrix techniques can be subsumed under these axioms. Looking through the literature of inference in artificial intelligence and elsewhere, it seems that the structure defined by Shenoy and Shafer is very often implicitly exploited, but mostly without explicit reference to it. Also the associated computational possibilities seem to be recognized over and over again. Therefore, it seems to be the time to formulate this structure explicitly and to study it in detail as a general and generic structure for inference.

A slightly changed version of the axiomatic formulation of Shenoy and Shafer is the starting point for this book. The mathematical structure defined by these axioms is called here *valuation algebra.* It will be studied both from an algebraic point of view as well as from a computational one. Special attention is paid to the idempotent variant of valuation algebras. These structures are called *information algebras.* Idempotency introduces a lot of additional structure. Indeed it places information algebra in the neighborhood of other theories of "information" like relational algebra, domain theory and the theory of information systems.

The elements of valuation algebras are called "valuations". This term is used in mathematics to denote some generalizations of measures, especially probability measures. Here, valuations are introduced in a similar spirit, although in a strict technical sense they do not correspond exactly to the concept used elsewhere in the theory of valuations.

The overview of the book is as follows: In Chapter 2 the axioms of (labeled) valuation algebras are introduced and a few elementary consequences are presented. Essentially, valuation algebras are commutative semigroups with an additional operation of marginalization, representing focusing. These two operations are linked by a kind of distributive law, which is fundamental for the theory. Several examples or instances of valuation algebras are described. These include systems inspired by probability theory like discrete probability potentials as used for example in Bayesian networks, or Gaussian potentials, motivated by normal regression theory, by Kalman filters and the like. In this view combination is represented by multiplication of densities, including conditional densities, and focusing corresponds to marginalization, i.e. summation or integration of densities. But the examples include also non-probabilistic systems. Relational algebra is an important example, and a very basic one as will be seen later. Here combination is the join and focusing corresponds to projection. Possibility potentials and Spohn potentials represent systems related to fuzzy systems. t-norms provide for a variety of combination operators in this field. This shows that valuation algebras indeed cover a wide range of interesting and useful systems for inference. For later reference another, weaker axiomatic system, allowing for partial focusing only, is also introduced.

In the following Chapter 3 the algebraic theory of valuation algebras is developed to some extent. Some concepts from universal algebra are introduced. In particular, there is an important congruence, which allows to group together valuations representing the "same" information. The corresponding quotient algebra gives us then an alternative way to represent the valuation algebra, in a "domain-free" form. Inversely, from the domain-free version we can reconstruct the original "labeled" algebra. This provides us with two equivalent ways to look at a valuation algebra, which proves very valuable. Essentially, the labeled point of view is more appropriate for computational and also for some semantical issues. The domain-free variant is generally more convenient for theoretical considerations. An important issue is the question of division, which, in general, is not defined in semigroups. But from semigroup theory we know that there are commutative semigroups which are a union of groups. We adapt in Chapter 3 this semigroup theory to *regular* valuation algebras. More generally, there are semigroups which are embedded (as a semigroup) in a semigroup which is a union of groups. This generalizes to *separative* valuation algebras. The issue of division is important for computation, but also for the concept of independence, as will be seen in Chapters 4 and 5. It is shown for example that, depending on the t-norm, possibility potentials may or may not allow for some form of division. And this makes a lot of difference both from a computational as well as from a semantical point of view.

There are several architectures for local computation known from the literature, especially for Bayesian networks. These will be presented in Chapter 4. Some of these architectures are valid for any valuation algebra. Others use some form of division. It is shown that these latter architectures can be used especially for regular valuation algebras. In the case of separative algebras, the additional problem arises that in the embedding union of groups, marginalization is only partially defined. It turns out that this partial marginalization is sufficient to apply the architectures with division also in the case of separative valuation algebras. In these algebras scaling is usually needed for semantical reasons. The architectures with division allow for an efficient organization of scaling.

The local character of computation in probability theory is closely related to conditional independence of variables. This concept can be generalized to valuation algebras in general and to regular and separative algebras in particular. This is discussed in Chapter 5. In probability theory conditional independence is also closely related to conditional probability, that is to division. This indicates that the concept of conditional independence depends very much on the structure of the valuation algebra. In a valuation algebra without division not very much can be said about conditional independence. Regular algebras on the other hand maintain many of the properties of conditional independence known from probability theory. In particular a concept of *conditional* can be defined which resembles a conditional density in probability theory. Separative valuation algebras are somewhere in between. Conditionals may also be defined. But they do not necessarily have all the properties of a conditional density. This explains for example why conditional belief functions are not of much interest in Dempster-Shafer theory of evidence.

The last two chapters are devoted to valuation algebras which are *idempotent*. This means that a valuation combined with a focused version of itself does not change the first valuation. This is an essential ingredient of "information". A piece of information combined with part of it gives nothing new. That is why these idempotent valuation algebras are called *information algebras* (Chapter 6). Idempotency allows to introduce a partial order between pieces of information, representing the relation of more (or less) informative information. This order is very essential for the theory. Information algebras become thus semilattices. With the aid of this partial order we can express the idea of "finite" elements, which serve to approximate general, "non-finite" elements. This leads to *compact* information algebras, which are in fact algebraic lattices (but with an additional operation of focusing). And this brings information algebras into the realm of domain theory. In fact, we show that an *information system*, a concept introduced by Scott into domain theory and adapted here to the needs of our theory, induces an information algebra. Inversely, any information algebra determines an information system. Thus information systems are an alternative way to represent information algebras. And they provide for a very important approach to information algebras, especially in practice. Propositional logic, systems of linear equations or linear inequalities are examples of information systems. In fact, information systems link information algebras to logic. Via the information algebras they induce, they can be treated by architectures of local computation as introduced in Chapter 4. Indeed, since idempotency makes division trivial, the architectures can even be simplified for information algebras. Information algebras can, on the other hand, also be related to relational algebra in general. For this purpose an abstract notion of tuple and relation is introduced. Information algebras can then be embedded into an abstract relational algebra over abstract relations. We call this a *file system.* So file systems provide for a second alternative representation of information algebras. In short: a piece of information may be looked at as a file (set) of tuples, that is as a relation. Or it may be looked at as a set of sentences expressed in some logic. Information can thus alternatively be described in a relational or in a logical way.

Information may be uncertain. So it is natural to ask how uncertainty can be represented in information algebras. This can be done by random variables with values in information algebras (Chapter 7). Random variables represent sources of evidence or information. Accordingly an operation of combination and another one of focusing can be defined. Not surprisingly, this leads to an information algebra of random variables. If we look at the distribution of these random variables, we find belief functions (here called support functions) in the sense of Dempster-Shafer theory of evidence. Therefore, we claim that information algebras are the natural mathematical framework for Dempster-Shafer theory of evidence. The usual set-theoretic framework of this theory is only a particular case of information algebras. But for example belief functions on linear manifolds (or systems of linear equations) are better treated in the framework of information algebra than in a purely set-theoretic setting. If information systems are used to express uncertainty, then this leads to assumption-based reasoning and *probabilistic argumentation systems.* Therefore, this is another approach to Dempster-Shafer theory of evidence, and a very practical one indeed. If an appropriate notion of "independence" between sources of evidence is introduced, then combination becomes the well-known rule of Dempster (expressed in information algebras of course). The corresponding algebra of "independent" belief functions is a valuation algebra.

This book depends on many publications and also on numerous personal discussions during different European research projects and other contacts. I want to give credit to the most important documents which helped to shape this book: The axioms and the first part of Chapter 3 (domain-free algebras) are largely based on an unpublished paper by Shafer (Shafer, 1991). The second part of Chapter 3 related to division has been motivated by the paper (Lauritzen & Jensen, 1997). There the author found the references to semigroup theory which are essential for the development of regular and separative valuation algebras. The chapter on local computation, Chapter 4, is based on the various original papers, especially (Lauritzen & Spiegelhalter, 1988; Jensen, Lauritzen & Olesen, 1990), where the different architectures were presented (for the case of probability networks) and on many personal discussions with Prakash Shenoy. Part of this chapter is also based on a chapter "Computation in Valuation Algebras", written by Prakash Shenoy and the author (Kohlas & Shenoy, 2000), in (Gabbay & Smets, 2000). The chapter on conditional independence, Chapter 5, is motivated by (Shenoy, 1997 a). It has been adapted to the axiomatic system used in this book and makes use of the results about regular and separative valuation algebras, as well as of the concept of valuation algebras with partial marginalization. Although these parts draw heavily on former work, the author hopes that there are sufficient new elements in this book to make these chapters interesting even for the reader which knows already the papers mentioned above.

Chapters 6 and 7 draw largely on unpublished material. Special credit is due to Robert Staerk, who contributed to the development of information algebra, and who, among other things, invented the file systems (Kohlas & Staerk, 1996). We remark that the cylindric algebras treated in (Henkin, Monk & Tarski, 1971) are special classes of information algebras, related to first order logic. Furthermore, classification domains as introduced and discussed in (Barwise & Seligman, 1997) seem to bear interesting connections to information algebras. Structures similar to information algebras are used also to study modules and modularity (Bergstra, et. al., 1990; Renardel de Lavalette, 1992). (Mengin & Wilson, 1999) discuss the use of the structure of information algebras for logical deduction.

For the uncertainty in information systems, Chapter 7, the basic literature on Dempster-Shafer theory of evidence was of course important, especially (Dempster, 1967; Shafer, 1973; Shafer, 1976; Shafer, 1979). We mention that there is an alternative, non-probabilistic approach to evidence theory (Smets, 1998). Partially this chapter is based on some former papers of the author (Kohlas, 1993; Kohlas, 1995; Besnard & Kohlas, 1995; Kohlas, 1997). These papers however were not based on information algebras.

This book is, as far as the author knows, the first systematic treatment of valuation algebras from an algebraic point of view. This does of course not mean that the subject is treated in an exhaustive way. Not nearly so. Many questions remain open. Here are only a few of them: What is the full structure theory of valuation and information algebras (what types of these algebras exist and how are they characterized)? What is the exact relation between information algebras and logic, which logic lead to information algebras? Which valuation algebras, representing uncertainty formalisms, can be induced from an algebra of random variables with values in an information algebra? How is Shannon's theory of information and algorithmic information theory related to information algebra, and especially to Dempster-Shafer theory? It is the author's hope that this book may arouse interest in the subject and serve to unify and promote efforts in developing inference schemes in different fields, using different formalisms.

http://www.springer.com/978-1-85233-689-9

Information Algebras
Generic Structures For Inference
Kohlas, J.
2003, X, 265 p., Softcover
ISBN: 978-1-85233-689-9