An Introduction to Idempotency

Jeremy Gunawardena
BRIMS, Hewlett-Packard Laboratories
Filton Road, Stoke Gifford, Bristol BS12 6QZ, UK
jhcg@hplb.hpl.hp.com

August 15, 1996

appearing in J. Gunawardena (editor), *Idempotency*
Publications of the Newton Institute, Cambridge University Press.
Proceedings of the BRIMS workshop on Idempotency, 3-7 October 1994.

Abstract

The word *idempotency* signifies the study of semirings in which the addition operation is idempotent: $a + a = a$. The best-known example is the *max-plus* semiring, $\mathbb{R} \cup \{-\infty\}$, in which addition is defined as $\max\{a, b\}$ and multiplication as $a + b$, the latter being distributive over the former. Interest in such structures arose in the late 1950s through the observation that certain problems of discrete optimisation could be linearised over suitable idempotent semirings. More recently the subject has established intriguing connections with automata theory, discrete event systems, nonexpansive mappings, nonlinear partial differential equations, optimisation theory and large deviations. The present paper was commissioned as an introduction to the volume of proceedings for the workshop on Idempotency held at Hewlett-Packard's Basic Research Institute in the Mathematical Sciences (BRIMS) in October 1994. It aims to give an introductory survey, from a coherent mathematical viewpoint, of the recent developments in the subject. The major open problems are pointed out and an extensive bibliography is provided.

Contents

1 Introduction **1**

2 First steps **1**

2.1 Introduction 1
2.2 Semirings and dioids 1
2.3 Examples of dioids 2
2.4 Semirings, homomorphisms and semimodules 4
2.5 The partial order in a dioid 5
2.6 Free dioids 6
2.7 Quantales and the Kleene star operation 7
2.8 Matrix dioids and graphs 10
2.9 The max-plus dioid and lattice ordered groups 11
2.10 Finite dioids 12

3 Automata and idempotency **13**

3.1 Quantales and Kleene's theorem 13
3.2 The tropical dioid 16

4 Discrete Event Systems **17**

4.1 Introduction 17
4.1.1 Examples and general questions 17
4.1.2 Mathematical models for DES 17
4.1.3 The basic model 18
4.1.4 Extensions of the basic model 19
4.2 Examples of topical functions 20
4.3 Nonexpansiveness and perioidicity 21
4.4 Cycle times 23

5 Nonlinear partial differential equations **25**

5.1 Introduction 25
5.2 Viscosity solutions 26
5.3 The role of idempotency 28
5.4 Functional analysis over dioids 29
5.4.1 Introduction 29
5.4.2 Boundedly complete dioids 29
5.4.3 Topologies on dioids 31

5.5 Idempotency and viscosity solutions 32

6 Optimisation and large deviations **34**

6.1 Asymptotics 34

6.2 Idempotent measures and integration 35

6.3 Optimisation theory 36

6.4 Large deviations 37

6.5 Topical functions and asymptotics 38

1 Introduction

The word *idempotency* signifies the study of semirings in which the addition operation is idempotent: $a+a=a$. The best-known example is the *max-plus* semiring, $\mathbb{R} \cup \{-\infty\}$, in which addition is defined as $\max\{a,b\}$ and multiplication as $a+b$, the latter being distributive over the former. Interest in such structures arose in the late 1950s through the observation that certain problems of discrete optimisation could be linearised over suitable idempotent semirings. Cuninghame-Green's pioneering book, [27], should be consulted for some of the early references. More recently the subject has established rich connections with automata theory, discrete event systems, nonexpansive mappings, nonlinear partial differential equations, optimisation theory and large deviations and these topics are discussed further in the subsequent sections of this paper. The phrase *idempotent analysis* first appears in the work of Maslov and the Russian school, [82].

The present paper has a twofold purpose. It is intended, firstly, to introduce the reader to some of the main themes within idempotency, with detailed references to the literature, and, secondly, to fill in some background to the remaining papers in this volume. The choice of material has been governed by our desire to present a coherent mathematical account of the subject, avoiding the parochialism into which it has sometimes fallen and bringing to light some of the major open problems. We make no pretence at completeness. A different perspective of the subject is put forward in Litvinov and Maslov's survey paper in this volume, [76].

2 First steps

2.1 Introduction

In this section we take a panoramic look at idempotent semirings. The intention is to give the reader a sense of the scope of the whole subject before looking at more specialised topics in later sections. The approach is largely algebraic.

2.2 Semirings and dioids

We recall that a semigroup is a set with an associative binary operation while a monoid is a semigroup with a distinguished identity element, [73].

Definition 2.1 *A semiring is a set, S, with two binary operations—denoted with the usual conventions for addition ("$+$") and multiplication ("$\times$" or "$.$")—and two distinguished elements, $0,1 \in S$, such that $0 \neq 1$ and*

- $(S, +, 0)$ *is a commutative monoid with identity element* 0,
- $(S, \times, 1)$ *is a monoid with identity element* 1,
- $a.0 = 0.a = 0$ *for all* $a \in S$,
- $a.(b + c) = a.b + a.c$, $(b + c).a = b.a + c.a$, *for all* $a, b, c \in S$.

A dioid, or idempotent semiring, is a semiring, D*, such that:*

- $a + a = a$ *for all* $a \in D$.

A dioid is commutative if $a.b = b.a$ *for all* $a, b \in D$.

The word dioid (or dioïd) is used by Kuntzmann, [74], to mean simply a semiring or *double monoid*, a usage followed by others, [47]. There seems little point in wasting a new word on something well-known and we follow the modern custom, [5, §4.2], of using dioid as a synonym for idempotent semiring.

It is less customary to use $+$ and $\times$ to denote the operations in a dioid and many authors use $\oplus$ and $\otimes$. We shall not do so here because, in our view, formulae involing the latter operations are difficult to read and fail to exploit the analogy with classical algebra. However, our choice does lead to difficulties because some of the best known examples of dioids are based on the classical number systems and it can be unclear whether, for instance, $+$ refers to the dioid addition or to classical addition. To resolve confusion arising from this, we shall use the symbol $:=$ between formulae. This will imply that the symbols on the left hand side refer to the dioid under consideration while the symbols on the right hand side have their classical, or customary, meanings. Hence, $a \times b := a + b$, means that dioid mulitipliation is equal to the customary addition, where *customary* should be clear from the context in which the formula appears.

2.3 Examples of dioids

Before going any further, it is best to see some examples.

1. The Boolean dioid: $\mathbb{B} = \{0, 1\}$, with $0, 1$ thought of as integers and addition and multiplication defined as maximum and minimum respectively.

2. The max-plus dioid: $\mathbb{R}_{\mathsf{max}} = \mathbb{R} \cup \{-\infty\}$ with $a + b := \max\{a, b\}$ and $a \times b := a + b$. Here, $0 := -\infty$ and $1 := 0$. The nonzero elements form a group under multiplication: $\mathbb{R}_{\mathsf{max}}$ is an idempotent semifield. (So too is $\mathbb{B}$ but for rather trivial reasons!) $\mathbb{R}_{\mathsf{max}}$ sometimes appears in its isomorphic form as the min-plus dioid: $\mathbb{R}_{\mathsf{min}} = \mathbb{R} \cup \{+\infty\}$, with $a + b := \min\{a, b\}$ and $a \times b := a + b$. Here, $0 := +\infty$ and $1 := 0$.

3. The tropical dioid: $\mathbb{N}_{\mathsf{min}} = \mathbb{N} \cup \{+\infty\}$, with $a + b := \min\{a, b\}$ and $a \times b := a + b$. Here, $0 := +\infty$ and $1 := 0$.

4. Let S be a set. The set of subsets of S, $\mathcal{P}(S)$, or the set of finite subsets of S, $\mathcal{P}^{Fin}(S)$, with addition and multiplication defined as union and intersection respectively, are both dioids. Here, $0 := \emptyset$ and $1 := S$. Similarly, let T be the set of open sets of any topology with addition and multiplication defined as union and intersection, respectively. Then T is a dioid. Dioids are at least as common as topological spaces!

5. Suppose that $(M, ., 1_M)$ is a monoid. The monoid operation can be used to give $\mathcal{P}(M)$ or $\mathcal{P}^{Fin}(M)$ a different dioid structure to that in the previous example. Addition is still given by union of subsets but multiplication is defined by: $U \times V = \{u.v \mid u \in U,\ v \in V\}$ where $U, V \subseteq M$. Here $0 := \emptyset$ and $1 := \{1_M\}$. There are several interesting examples of such dioids. Take $M = \mathbb{R}^d$ with the standard vector space addition as the monoid operation. The dioid multiplication in $\mathcal{P}(\mathbb{R}^d)$ then corresponds to Minkowski addition of subsets, [58]. Another example is the dioid of formal languages. If A is a set, let A^* be the set of finite strings (words) over A with the monoid operation defined by juxtaposition of strings and the empty string, ϵ, as the identity element. The subsets of A^*, are called (formal) languages over the alphabet A. The dioid $\mathcal{P}(A^*)$ is noncommutative if A has more than one element.

6. Suppose that $(G, +, 0)$ is an abelian group, not necessarily finite. For instance, $G = \mathbb{R}^n$. Let $(\mathbb{R}_{\mathsf{min}})_b(G)$ denote the set of functions $u : G \to \mathbb{R}_{\mathsf{min}}$ which are bounded below. If $u, v \in (\mathbb{R}_{\mathsf{min}})_b(G)$, define addition pointwise, $(u + v)(x) = u(x) + v(x)$, and multiplication as convolution: $(u.v)(x) := \inf_{y \in G}\{u(y) + v(x - y)\}$. Here $x - y$ is calculated in the group G. The boundedness of u and v ensures that this is well-defined. $(\mathbb{R}_{\mathsf{min}})_b(G)$ is a convolution dioid, [94]. The operation $u.v$ is called *inf convolution* in convex analysis, [4, Exercise 1.4].

7. Let R be a commutative ring and let $Spec(R)$ denote the set of ideals of R. $Spec(R)$ becomes a dioid when addition is defined by the sum of ideals, $I + J = \{a + b \mid a \in I,\ b \in J\}$, and multiplication by the product, $I.J = \{a_1.b_1 + \cdots + a_n.b_n \mid a_i \in I,\ b_j \in J\}$, [102, Chapter 12].

8. If D is a dioid, then $M_n(D)$ will denote the dioid of $n \times n$ matrices with entries in D with matrix addition and mulitiplication as the operations.

9. Let D be a dioid and A a set. The dioid of formal power series over D with (non-commuting) variables in A, denoted $D\langle\langle A \rangle\rangle$, is the set of D-valued functions on A^* with addition defined pointwise and multiplication as convolution: if $f, g \in D(A^*)$, then $(f.g)(w) = \sum_{u.v=w} f(u).g(v)$. By restricting to the sub-dioid of functions which are non-zero at only

finitely many points, we recover the polynomials over D with variables in A, denoted $D[A]$.

2.4 Semirings, homomorphisms and semimodules

Many of the basic concepts in the theory of rings can be defined in a similar way for semirings and can hence be specialised to dioids. For the most part, the idempotency plays no special role. We run through some of the most important definitions here; for more details see [44]. Let R, S be semirings. A homomorphism of semirings from R to S, is a function $f : R \to S$ which is a homomorphism of monoids for both addition and multiplication:

$$\begin{array}{rr} f(a+b) = f(a) + f(b) & f(0_R) = 0_S \\ f(a.b) = f(a).f(b) & f(1_R) = 1_S \ . \end{array}$$

We can define a function $\alpha : \mathcal{P}(A^*) \to \mathbb{B}\langle\langle A \rangle\rangle$ by $\alpha(U)(w) = 1$ if $w \in U$, $\alpha(U)(w) = 0$ if $w \notin U$. The reader can check that this is an isomorphism of semirings. That is, there exists a homomorphism of semirings, $\beta : \mathbb{B}\langle\langle A \rangle\rangle \to \mathcal{P}(A^*)$ which is inverse to α: $\alpha\beta = 1_{\mathbb{B}\langle\langle A \rangle\rangle}$ and $\beta\alpha = 1_{\mathcal{P}(A^*)}$. By restriction, β also induces an isomorphism between $\mathbb{B}[A]$ and $\mathcal{P}^{Fin}(A^*)$.

Let $(M, +)$ be a commutative monoid and R a semiring. M is said to be a left R-semimodule if there is an action of R on M, $R \times M \to M$, called scalar multiplication and denoted $r.a$, such that, for all $r, s \in R$ and $a, b \in M$,

$$\begin{array}{rclrcl} r.(a+b) & = & r.(a) + r.(b) & r.(0_M) & = & 0_M \\ (r+s).a & = & r.a + s.a & 0_R.a & = & 0_M \\ r.(s.a) & = & (r.s).a & 1_R.a & = & a \ . \end{array}$$

It is worth noting at this point that if R happens to be a dioid, then M must necessarily have an idempotent addition:

$$a + a = 1.a + 1.a = (1+1).a = 1.a = a \ .$$

A useful source of R-semimodules is provided by the following construction. Let A be any set and let $R(A)$ denote the set of all functions $u : A \to R$. This has a natural structure as a left R-semimodule: addition is defined pointwise, $(u + v)(a) = u(a) + v(a)$, and scalar multiplication by $(r.u)(a) = r.u(a)$. When A is infinite, $R(A)$ contains inside itself another useful R-semimodule. Define the support of u by $supp(u) = \{a \in A \mid u(a) \neq 0\}$ and let $R^{Fin}(A)$ denote the set of functions u with finite support. It is clear that $R^{Fin}(A)$ is a sub-semimodule of $R(A)$.

If M, N are R-semimodules, a homomorphism of R-semimodules from M to N is a homomorphism of monoids $f : M \to N$ which respects the scalar multiplication: $f(r.a) = r.f(a)$. We can construct homomorphisms between the R-semimodules $R^{Fin}(A)$ as follows. If $f : A \to B$ is any set

function, then f can be extended to a function, also denoted f for convenience, $f : R^{Fin}(A) \to R^{Fin}(B)$ where $f(u)(b) = \sum_{f(a)=b} u(a)$. The finiteness of $supp(u)$ ensures that this is well-defined. The reader can check that this is a homomorphism of R-semimodules.

At this point, it will be convenient to make use of the language of category theory, essentially to clarify the nature of $R^{Fin}(A)$. The reader who is unfamiliar with this language but who nevertheless has an intuitive understanding of what it means for an object to be *freely generated*, will not lose much by ignoring the details below. The canonical reference for those wishing to know more is MacLane's book, [77].

The constructions given above define a functor from the category of sets, Set, to the category of R-semimodules, SMod_R, which is left adjoint to the forgetful functor which forgets the R-semimodule structure. That is, if A is a set and M is an R-semimodule, then there is a natural one-to-one correspondence between the respective sets of homomorphisms:

$$\mathsf{Hom}(A, M)_{\mathsf{Set}} \longleftrightarrow \mathsf{Hom}(R^{Fin}(A), M)_{\mathsf{SMod}_R} . \tag{2.1}$$

This expresses the fact that $R^{Fin}(A)$ is the free R-semimodule generated by A: any set function $f : A \to M$ can be extended to a unique homomorphism of R-semimodules $f : R^{Fin}(A) \to M$. If $u \in R^{Fin}(A)$, the extension is given by $f(u) = \sum_{a \in A} u(a).f(a)$, the finiteness of $supp(u)$ again ensuring that this is well-defined.

If M is an R-semimodule then the set $\mathsf{Hom}(M, M)_{\mathsf{SMod}_R}$ has both an addition, $(f + g)(a) = f(a) + g(a)$, and a multiplication given by composition of functions $(f.g)(a) = f(g(a))$. This defines a semiring, over which M is a left semimodule under $f.a = f(a)$. For finitely generated free modules, $\mathsf{Hom}(M, M)$ is well known. Let A be a finite set, so that $R^{Fin}(A) = R(A)$. By (2.1), an element of $\mathsf{Hom}(R(A), R(A))$ is uniquely determined by a function $A \to R(A)$. This, in turn, may be uniquely specified by a function $A \times A \to R$. In other words, by an element of $R(A \times A)$, or by an $n \times n$ matrix over R. It is easy to show, after choosing an ordering, $A = \{a_1, \cdots, a_n\}$, that this gives an isomorphism of semirings, $\mathsf{Hom}(R(A), R(A)) \longrightarrow M_n(R)$.

2.5 The partial order in a dioid

Our constructions have, so far, been valid for arbitrary semirings. The reader might be forgiven for thinking that the theory of dioids has no special character of its own. This is not the case. The idempotency gives rise to a natural partial order in a dioid which differentiaties the theory from that of more general semirings.

Proposition 2.1 ([62, §I.1.3]) *Let A be a commutative semigroup with an idempotent addition. Define $a \preceq b$ whenever $a + b = b$. Then $(A, \preceq)$ is a sup-*

semilattice (a partially ordered set in which any two elements have a least upper bound). Furthermore,

$$\max\{a, b\} = a + b \tag{2.2}$$

Conversely, if $(A, \preceq)$ is a sup-semilattice and $+$ is defined to satisfy (2.2), then $(A, +)$ is an idempotent semigroup. These two constructions are inverse to each other.

It follows that any dioid, D, has a natural ordering, denoted $\preceq$ or $\preceq_D$. Addition is a monotonic operation and 0 is the least element. Distributivity implies that left and right multiplication are semilattice homomorphisms; in particular, they are monotonic. For the dioids based on the classical number systems, the dioid ordering is related to the usual ordering (for the tropical semiring, for which $0 := +\infty$, $a \preceq b$ if, and only if, $a \geq b$). For the dioids constructed using $\mathcal{P}$ or $\mathcal{P}^{Fin}$, it corresponds to inclusion of subsets.

We see from Proposition 2.1 that a dioid may be thought of as a semilattice-ordered semigroup. There is a continuing tradition of work on idempotency from this perspective, [27, 109, 14].

2.6 Free dioids

We turn now to an examination of the dioids arising as sets of subsets (examples 4 and 5 in §2.3). Once again, it is convenient to use the language of category theory.

The constructions $\mathcal{P}(M)$ and $\mathcal{P}^{Fin}(M)$ extend to functors from the category of monoids, Monoid, to the category of dioids, Dioid: if $f : M \to N$ is a homomorphism of monoids, then define $f : \mathcal{P}(M) \to \mathcal{P}(N)$ by $f(U) = \{f(u) \mid u \in U\}$ and similarly for $\mathcal{P}^{Fin}(-)$.

Proposition 2.2 *The functor $\mathcal{P}^{Fin} : \mathsf{Monoid} \to \mathsf{Dioid}$ is left adjoint to the forgetful functor from Dioid to Monoid which forgets the dioid addition.*

Proof: Let M be a monoid and D a dioid. It is required to show that there is a natural one-to-one correspondence between the respective sets of homomorphisms:

$$\mathsf{Hom}(M, D)_{\mathsf{Monoid}} \longleftrightarrow \mathsf{Hom}(\mathcal{P}^{Fin}(M), D)_{\mathsf{Dioid}} \ .$$

Let $q : M \to D$ be a homomorphism of monoids. Define $\theta(q) : \mathcal{P}^{Fin}(M) \to D$ by $\theta(q)(U) = \sum_{u \in U} q(u)$. It is an exercise to check that θ establishes the required correspondence.

QED

In other words, $\mathcal{P}^{Fin}(M)$ is the free dioid generated by the monoid M. The construction A^* can also be extended to a functor from Set to Monoid which is left adjoint to the forgetful functor which forgets the monoid multiplication. A^* is hence the free monoid generated by A. These constructions can also be specialised to the commutative case. The free commutative monoid generated by A is simply $\mathbb{N}^{Fin}(A)$. By putting these remarks together, we obtain the following characterisations.

Proposition 2.3 *If A is a set, then $\mathcal{P}^{Fin}(A^*)$ and $\mathcal{P}^{Fin}(\mathbb{N}^{Fin}(A))$ are, respectively, the free dioid and the free commutative dioid generated by A.*

For finite A, the commutative case was proved in [96, Theorem 4.1].

2.7 Quantales and the Kleene star operation

As we saw in §2.4, the difference between $\mathcal{P}(M)$ and $\mathcal{P}^{Fin}(M)$ is similar to the difference between power series and polynomials. In the former, an infinite number of additions can be meaningfully performed. Dioids of this type have been extensively studied under another name.

Definition 2.2 *A quantale, or complete dioid* ([5, Definition 4.32]), *Q, is a dioid in which the underlying sup-semilattice has arbitrary suprema over which the multiplication distributes: $\forall S \subseteq Q$ and $\forall a \in Q$,*

- *there exists a least upper bound* $\sup_{x\in S} x$*;*
- $a.(\sup_{x\in S} x) = \sup_{x\in S}(a.x)$, $(\sup_{x\in S} x).a = \sup_{x\in S}(x.a)$.

The word *quantale* was coined by Mulvey in his study of the constructive foundations of quantum theory, [84]. His definition does not require an identity element for multiplication.

If $(V, \leq)$ is any partial order and $S \subseteq V$ any subset of V, then it is easy to see that

$$\inf_{x\in S} = \sup\{y \in V \mid y \leq x,\ \forall x \in S\} \ . \tag{2.3}$$

If either side exists, so does the other, and both are equal. Hence, every quantale has a greatest element, $\mathsf{T} = \inf \emptyset$ and any subset has an infimum. However, multiplication does not necessarily distribute over infima, even though it does over suprema.

We regard a quantale Q as endowed with an infinitary addition given by $\sum_{x\in S} x = \sup_{x\in S} x$, following the identification in (2.2). Homomorphisms of quantales are required to preserve the multiplication and the infinitary addition. The category of quantales, Quant, forms a subcategory of Dioid.

$\mathcal{P}(-)$ defines a functor from Monoid to Quant which is left adjoint to the forgetful functor which forgets the quantale multiplication. Hence, we obtain the following characterisations.

Proposition 2.4 *If A is a set, then $\mathcal{P}(A^*)$ is the free quantale generated by A* ([1, Theorem 2.1]) *and $\mathcal{P}(\mathbb{N}^{Fin}(A))$ is the free commutative quantale generated by A.*

Topological spaces form another source of quantales, as pointed out in example 4 in §2.3. These quantales are special because multiplication corresponds to minimum. Quantales of this type are called frames, [62, Chapter 2] and are characterised by having 1 as the greatest element and an idempotent multiplication, $a^2 = a$, [64, Proposition III.1]. Because they provide an extended concept of topological space, frames have proved important in formulating an appropriate abstract setting—Grothendieck's notion of *topos*—for modern algebraic geometry. In the course of proving structure theorems for topoi, Joyal and Tierney have developed the theory of complete semimodules over frames, [64].

Quantales have been studied for a number of reasons. On the one hand the infinitary addition allows the definition of the binary operation $a \rightarrow b$, characterised by $c \preceq a \rightarrow b$ if, and only if, $c.a \preceq b$. (This is best understand as an application of the Adjoint Functor Theorem, [62, §I.4.2].) This operation is referred to variously as a residuation, [11, 5], implication or pseudocomplement, [62, §I.1.11]. When Q is a frame, the operation $\neg x = x \rightarrow 0$ has many of the features of a negation, except that the law of the excluded middle, $x + \neg x = \mathsf{T}$ may fail. Logicians have consequently studied frames as models for intuitionistic logic, [102]. More recently, quantales have served the same purpose for Girard's linear logic. They are also used as semantic models in computer science. For details, see the paper by Mascari and Pedicini in this volume, [78], or [1].

On the other hand, despite the infinitary operation, quantales and frames are still algebraic theories, in the sense that they can be defined in terms of generators and relations, [62, §II.1.2]. For frames, this leads to the study of topological spaces from an algebraic viewpoint, sometimes called *pointless topology*, [63].

For the purposes of the present paper, interest in quantales arises for the following reason. The discrete dynamical system $x_{i+1} = a.x_i + b$ appears in several situations. In general a may be an element of a dioid D and x_i and b elements of some left D-semimodule. The problem is to understand the asymptotic behaviour of the sequence $x_0, x_1, \cdots$. A useful first step is to determine the equilibrium points of the system, where

$$x = a.x + b\,. \tag{2.4}$$

(In this volume, Walkup and Borriello consider the more general problem of solving $a.x + b = c.x + d$ when a and c are max-plus matrices, [106].)

Definition 2.3 *If Q is a quantale and $a \in Q$, the Kleene star of a, denoted a^*, is defined by*

$$a^* = 1 + a + a^2 + a^3 + \cdots = \sup_{0 \leq i} a^i \ . \tag{2.5}$$

It is sometimes helpful to use also $a^+ = a + a^2 + \cdots$. By the infinitary distributive law, $a.a^* = a^+$.

Proposition 2.5 ([21, Theorem III.2]) *Let Q be a quantale and $a, b \in Q$. Then $a^*.b$ is the least solution of (2.4).*

Proof: $a.(a^*.b) + b = a^+.b + b = a^*.b$. Hence $a^*.b$ is a solution. Since multiplication and addition are both monotonic, if $x \preceq y$, then $a.x + b \preceq a.y + b$. Now let s be any other solution. Since $0 \preceq s$, it follows that $b \preceq s$. By induction, $(1 + a + a^2 + \cdots + a^n).b \preceq s$. It is now not difficult to show that $a^*.b \preceq s$, as required.

QED

We see from this that a^* has a "rational" flavour: it is analogous to $(1 - a)^{-1}$ in customary algebra. The star operation satisfies many identities, [21, Chapter 3], of which we mention only one:

$$(a + b)^* = (a^*.b)^*.a^* \ . \tag{2.6}$$

The left hand side is a sup of terms of the form $a^{i_1}b^{j_1} \cdots a^{i_m}b^{j_m}$. On the right hand side, $(a^*.b)^*$ is a sup of similar terms except for those ending in powers of a. The product with a^* supplies the missing terms. The problem of finding a complete set of identities for $+$, $\times$ and * is very subtle, [21, Chapter 12].

The uniqueness of the solution given in Proposition 2.5 appears to be an open problem. It is a classical result in automata theory, Arden's Lemma, [60, Exercise 2.22], that, in the free quantale, $\mathcal{P}(A^*)$, if $1 \not\preceq a$, then $a^*.b$ is the unique solution of (2.4). This is not true in general. In any frame, $a^2 = a$ and 1 is the greatest element. Hence the function $f(x) = a.x + b$ is itself idempotent, $f^2 = f$, implying that any element of the form $a.x + b$ is a solution of (2.4). On the other hand, in any quantale, if $1 \preceq a$, it is easy to see that $a^*.(u + b)$ is a solution of (2.4) for any $u \in Q$. Another relevant result is [5, Theorem 4.76].

It is convenient to mention here a technical trick which is well known in certain quarters. It is used by Walkup and Borriello in their paper in this volume, [106], and it will be the key ingredient in the proof of Theorem 3.1. The proof is left as an exercise for the reader, who will need to make use of (2.6) and Proposition 2.5. There are several equivalent formulations: [72, Theorem 2.5],[19, Lemma 2].

Lemma 2.1 ([21, Theorem III.4]) *Let Q be a quantale. The following identity holds for any block representation of a matrix over Q.*

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^* = \begin{pmatrix} (A+BD^*C)^* & A^*B(D+CA^*B)^* \\ D^*C(A+BD^*C)^* & (D+CA^*B)^* \end{pmatrix}$$

2.8 Matrix dioids and graphs

If the dioid R is not a quantale, it may still be the case that a^* exists and that Proposition 2.5 continues to hold. Indeed, in any dioid, if $a \preceq 1$ then $a^* = 1$. In a matrix dioid, the existence of A^* is related to the eigenvalues of A, in analogy with classical results. For dioids such as $M_n(\mathbb{R}_{\mathsf{max}})$, this has been completely worked out, [5, Theorem 3.17].

The case of matrix dioids is of particular interest for problems of discrete optimisation. This arises through the intimate relationship between matrices and graphs, which becomes particularly attractive in the idempotent context. Let R be a semiring and $T \in M_n(R)$ a matrix.

Definition 2.4 *The graph of T, $\mathcal{G}(T)$, is a directed graph on the vertices $\{1, \cdots, n\}$ with edges labelled by elements of R. There is an edge from i to j if, and only if, $T_{ij} \neq 0$. If there is such an edge then its label is T_{ij}.*

This gives a one-to-one correspondence between $n \times n$ matrices over R and directed graphs on $\{1, \cdots, n\}$ with edge labels in R.

Taking powers of A corresponds to building longer paths. A path, p of length m from i to j is a sequence of vertices $i = v_0, \cdots, v_m = j$ such that $A_{v_i v_{i+1}} \neq 0$. We can write

$$A^m_{ij} = \sum_p |p|_{\mathsf{w}} \ ,$$

where p runs over all paths of length m from i to j and $|p|_{\mathsf{w}}$ is the weight of the path: the product of the labels on the edges in the path, $|p|_{\mathsf{w}} = A_{v_0 v_1}.\cdots.A_{v_{m-1} v_m}$. For a general semiring, it is hard to deduce much from this. If R is a dioid, however, the sum corresponds to taking a maximum with respect to $\preceq_R$ and A^m_{ij} is the maximum weight among paths of length m from i to j. By choosing the dioid appropriately, a variety of discrete optimisation problems can be formulated in terms of matrices.

For instance, we can answer questions about the existence of paths in a directed (unlabelled) graph, G. Assume the vertices are labelled $\{1, \cdots, n\}$. Take $R = \mathbb{B}$ and label each edge of the graph with $1 \in \mathbb{B}$. Let A be the matrix in $M_n(\mathbb{B})$ corresponding to G. Then A^* (which exists since $\mathbb{B}$ and $M_n(\mathbb{B})$ are both quantales) gives the transitive closure of the edge relation in G: $A^*_{ij} = 1$ if, and only if, there is some path in G from i to j. Problems of enumeration,

shortest paths, critical paths, reliability, etc, in graphs or networks can be formulated using other dioids, often as solutions to (2.4), [47, Chapter 2].

To calculate A^* efficiently, classical algorithms for computing the inverse matrix—Jacobi, Gauss-Seidel, Jordan—can be adapted to the idempotent setting, [15], [47, Chapter2]. For finding longest paths (for which the appropriate dioid is $\mathbb{R}_{\max}$) these correspond to well-known algorithms such as those of Bellman, Ford-Fulkerson and Floyd-Warshall, [15].

It is interesting to ask if properties of periodic graphs, which are infinite graphs with $\mathbb{Z}^n$-symmetry, can be studied by similar methods. Backes, in his thesis, has given a formula for longest paths in a periodic graph, [7]. Can this be interpreted in idempotent terms?

2.9 The max-plus dioid and lattice ordered groups

The max-plus dioid, $\mathbb{R}_{\max}$, has been of great importance for idempotency and deserves special mention. For a general reference see [5, Chapter 3]. Cuninghame-Green's survey paper, [28], discuses some topics that we omit.

As remarked in §2.3, $\mathbb{R}_{\max}$ is a semi-skewfield. The group operation endows such structures with natural symmetry. The reader should have no difficulty proving the following result by using the remarks in §2.5.

Lemma 2.2 *If D is a semi-skewfield then $(D, \preceq_D)$ is a lattice and, for $a, b \in D\backslash\{-\infty\}$, $\min\{a, b\} = (a^{-1} + b^{-1})^{-1}$.*

In the language of ordered algebraic structures, semi-skewfields are therefore lattice ordered groups and these sometimes form a convenient generalisation of $\mathbb{R}_{\max}$. They correspond to the *blogs* of [27], a name which has, understandably, not survived.

The most extensively studied aspect of max-plus is linear algebra: finitely generated free semimodules and their endomorphisms. There is a close analogy between this and classical linear algebra over $\mathbb{R}$ and this has stimulated much of the interest, [8, 9, 14, 46, 86]. The spectral theory of max-plus matrices has been of particular interest because of the role of eigenvalues as a performance measure for discrete event systems; see §4.4. The basic discovery is that eigenvalues correspond to maximum mean circuit weights in the associated graph.

Let us consider elements of $(\mathbb{R}_{\max})(\{1, \cdots, n\})$ as column vectors. Recall that u is an eigenvector of a matrix $A \in M_n(\mathbb{R}_{\max})$, with eigenvalue $\lambda \in \mathbb{R}_{\max}$, if $Au = \lambda u$. If g is a circuit in $\mathcal{G}(A)$—a path that returns to its starting vertex—let $|g|_\ell$ be its length: the number of edges in the circuit. Note that in $\mathbb{R}_{\max}$, the polynomial equation $x^k = a$ has a unique solution: $x = a^{1/k} := a/k$. (Dioids with this property are *radicable* in [27] or *algebraically complete* in

[82].) Hence, the mean weight of a circuit, $\mu(g) := |g|_{\mathsf{w}}/|g|_\ell$, is a *bona fide* element of $\mathbb{R}_{\mathsf{max}}$.

Proposition 2.6 *Suppose that $A \in M_n(\mathbb{R}_{\mathsf{max}})$ and $Au = \lambda u$. Then, in $\mathbb{R}_{\mathsf{max}}$, $\lambda = \sum_g \mu(g)$, where g ranges over circuits of $\mathcal{G}(A)$ whose vertices lie in $supp(u)$. In particular, any two eigenvectors with the same support have the same eigenvalue.*

Results of this type go back to [26] and [92]. When $supp(u) = \{1, \cdots, n\}$, a description of the eigenvectors, valid for radicable semi-skewfields, appears in [27]. The general case for $\mathbb{R}_{\mathsf{max}}$ is discussed in [107], and later in [39], and for dioids more general than semi-skewfields in [34]. The formulation above is taken from [52, Lemma 4.5]. There is a strong analogy here between the spectral theory of max-plus matrices and that of nonnegative matrices, [83], which seems too good to be an accident. We will return to this point in §6.5.

Cuninghame-Green has investigated beyond the linear framework by studying polynomials and rational functions over $\mathbb{R}_{\mathsf{max}}$. We mention only the following result, a max-plus analogue of the fundamental theorem of algebra.

Theorem 2.1 ([29]) *Let $p(x) \in (\mathbb{R}_{\mathsf{max}})[\{x\}]$ be a polynomial over max-plus. Then p can be factored into linear factors,*

$$p(x) = a_0 \prod_{i=1}^{n} (x + a_i), \tag{2.7}$$

and this representation is unique up to order of the factors.

This result is best understood in terms of the graph of $p(x)$, which is the upper envelope of functions of the form $b.x^k := kx+b$, where $k \in \mathbb{N}$ and $b \in \mathbb{R}$. The quantities a_i in (2.7), for $1 \leq i \leq n$ and $a_i \neq -\infty$, are the x-coordinates of the corners of the graph. See [25] for an application.

Over a field, every finitely generated module is free but this is not the case for semimodules over an idempotent semifield. Wagneur has been one of the few to study non-free finitely generated semimodules, [105]. (See also [64].) He has found a structure theorem for finitely generated semimodules over dioids such as $\mathbb{R}_{\mathsf{max}}$, which is analogous to that for finitely generated abelian groups: every such semimodule is the sum of a *free* part and a *torsion* part, [104, Theorem 3]. Further results in this direction appear in Wagneur's paper in this volume, [103].

2.10 Finite dioids

As we have seen, dioids are very plentiful. However, the examples so far considered fall into a small number of families. Shubin, who seems to have

been one of the few to take a systematic approach, has constructed all finite commutative dioids having at most 4 elements, [96, §2]. (Conway enumerates certain specialised dioids in [21, Chapter 12].) Of the examples discussed above, only $\mathbb{B}$ and $\mathcal{P}(\mathbb{Z}_2)$ fall into this set. As Shubin points out, even with the imposed restrictions, the examples appear too numerous for systematic classification: there are 14 pairwise non-isomorphic commutative dioids with 4 elements. However, this profusion can be misleading.

Proposition 2.7 *The only dioid which is both a quantale and a semi-skewfield is the Boolean dioid,* $\mathbb{B}$. *In particular,* ([96, Theorem 3.1]) *there are no finite idempotent semi-skewfields other than* $\mathbb{B}$.

Proof: Let D be a dioid satisfying the hypotheses and let g be the greatest element of D. Since $1 \preceq g$, it follows that $g \preceq g^2$. Hence $g = g^2$ and so $g = 1$. Now suppose that $s \neq 0$. Since $s^{-1} \preceq 1$, it follows that $1 \preceq s$ and so $s = 1$. Hence $D = \mathbb{B}$.

QED

This negative result is an appropriate place to bring this section to a close. We hope to have given some idea of the scope of the dioid concept. It would be fair to say that most of the interesting questions about general dioids remain unanswered, if not unasked. In the subsequent sections we shall be concerned with specific dioids and will not touch on such general matters again.

3 Automata and idempotency

Automata may be thought of as language recognisers. That is, as rules for recognising strings of symbols from some (finite) alphabet. By defining appropriate rule schemes, computer scientists have defined classes of automata of varying computing power: finite automata, push-down automata, stack automata, Turing machines, etc, [60]. Because languages are elements of the free quantale, $\mathcal{P}(A^*)$, it should not come as a surprise that automata theory has something to tell us about idempotent semirings. In this paper, we shall only consider finite automata, where the connections have been most studied. For a good overview see [87].

Our main objective in this section is to prove a generalised version of Kleene's Theorem. We then briefly discuss the role of the tropical dioid.

3.1 Quantales and Kleene's theorem

It will be convenient to identify an element $u \in \mathbb{B}(\{1, \cdots, n\})$ with a column vector, as in §2.9, and to make use of the customary operations on vectors and matrices. For instance, u^{t} will denote the transpose of u:

$u^{\mathsf{t}} = (u(1), \cdots, u(n))$. Since the Boolean dioid, $\mathbb{B}$, can be regarded as a subdioid of any dioid, D, vectors and matrices over $\mathbb{B}$ can always be regarded as vectors and matrices over D.

Definition 3.1 *Let Q be a quantale and $S \subseteq Q$. An element $q \in Q$ is recognisable over S if, for some n, there exists a matrix $T \in M_n(S)$, and vectors $\iota, \phi \in \mathbb{B}(\{1, \cdots, n\})$ such that $q = \iota^{\mathsf{t}} T^* \phi$. The set of elements recognisable over S will be denoted* $\mathrm{Rec}(S)$.

Let A be a finite set and let $Q = \mathcal{P}(A^*)$, the free quantale generated by A. The elements of Q are the languages over A. Let $S = \{q \in Q \mid q \preceq A\}$, where A is treated as an element of Q by identifying each $a \in A$ with the corresponding string of length 1. The elements of S can then be identified with the subsets of A. The data (ι, T, ϕ) over S correspond to a nondeterministic finite automaton, [88, §5.1]. The states of the automaton are $\{1, \cdots, n\}$ and the subsets ι and ϕ are the initial states and final states, respectively. T encodes the transitions of the automaton: if $T_{ij} = q$, where $q \subseteq A$, then if the automaton is in state i and encounters any symbol from the subset q, it may make a transition to state j. The possibility that the same symbol could occur in both T_{ij} and T_{ik}, where $j \neq k$, or that $T_{ij} = \emptyset$, allows for nondeterminism.

In the usual definition of a finite automaton, the effect of individual symbols is separated out. Let $\mu : A \to M_n(\mathbb{B})$ be given by: $\mu(a)_{ij} = 1$ if, and only if, $a \in T_{ij}$. The matrix $\mu(a)$ identifies those transitions which recognise the symbol a. It follows that, in $M_n(\mathcal{P}(A^*))$,

$$T = \sum_{a \in A} (aI)\mu(a),$$

where I is the identity matrix. Since $M_n(\mathbb{B})$ is a multiplicative monoid, the function μ can be extended to a homomorphism of monoids $\mu : A^* \to M_n(\mathbb{B})$. Hence, ignoring initial and final states, we can think of an automaton in different ways: as a representation of the free monoid over finitely generated free $\mathbb{B}$ modules, or as a finitely generated semigroup (or monoid) of matrices generated by the subset $\{\mu(a) \mid a \in A\} \subseteq M_n(\mathbb{B})$. The slogan *automata are semigroups of matrices* is helpful to keep in mind, particulary when it comes to defining more general types of automata, as in §3.2 and §4.1.4.

What does the element $q = \iota^{\mathsf{t}} T^* \phi$ correspond to in terms of automata? Recalling the discussion in §2.8 the reader can check that T_{ij}^m is the set of strings of length m which lead from state i to state j. It follows that q is the set of strings of any length which lead from some initial state to some final state. Hence, q is the language recognised by the automaton, [88, §5.1].

Kleene's original result, [65], which is the starting point of formal language theory, characterised the languages of finite automata and led to the introduction of the * operation.

Definition 3.2 *Let Q be a quantale and $S \subseteq Q$. The rational closure of S, S^*, is the smallest subset of Q which contains S and is closed under $+$, $\times$ and *.*

Theorem 3.1 *Let Q be a quantale and $\{0,1\} \subseteq S \subseteq Q$. Then* $\mathrm{Rec}(S) = S^*$.

Proof: We first show that $S^* \subseteq \mathrm{Rec}(S)$. Choose $s \in S$ and form the data below, which we may do since $0 \in S$.

$$\iota = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad T = \begin{pmatrix} 0 & s \\ 0 & 0 \end{pmatrix} \quad \phi = \begin{pmatrix} 0 \\ 1 \end{pmatrix} .$$

Evidently, $\iota^{\mathsf{t}} T^* \phi = s$ and so $S \subseteq \mathrm{Rec}(S)$. Now suppose that $q_1 = \iota_1 T_1^* \phi_1$ and $q_2 = \iota_2 T_2^* \phi_2$ and form the three sets of block matrix data below, which we may do since $\{0,1\} \subseteq S$.

$$\begin{array}{lll} \iota = \begin{pmatrix} \iota_1 \\ \iota_2 \end{pmatrix} & T = \begin{pmatrix} T_1 & 0 \\ 0 & T_2 \end{pmatrix} & \phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \\ \iota = \begin{pmatrix} \iota_1 \\ 0 \end{pmatrix} & T = \begin{pmatrix} T_1 & \phi_1 \iota_2^{\mathsf{t}} \\ 0 & T_2 \end{pmatrix} & \phi = \begin{pmatrix} 0 \\ \phi_2 \end{pmatrix}, \\ \iota = \begin{pmatrix} 0 \\ 1 \end{pmatrix} & T = \begin{pmatrix} T_1 & \phi_1 \\ \iota_1^{\mathsf{t}} & 0 \end{pmatrix} & \phi = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \end{array}$$

(It is instructive to picture these constructions for finite automata.) The reader can now check, using Lemma 2.1, that $\iota^{\mathsf{t}} T^* \phi$ for each set of data is, respectively, $q_1 + q_2$, $q_1 . q_2$ and q_1^*. It follows by induction that $\mathrm{Rec}(S)$ is closed under $+$, $\times$ and *. Hence, $S^* \subseteq \mathrm{Rec}(S)$, since S^* is the smallest set containing S with that property.

For the other way round, if $q = \iota T^* \phi$, where $T \in M_n(S)$, then, by Lemma 2.1, $T^* \in M_n(S^*)$. Hence $q \in S^*$. It follows that $\mathrm{Rec}(S) = S^*$, as required.

QED

There is a mild embarassment with this result: in the case of finite automata, where $Q = \mathcal{P}(A^*)$ and $S = \{q \preceq A\}$, $1 \notin S$! However, in this case, S has additional properties that allow the same result to go through.

Corollary 3.1 *Let Q be a quantale and $S \subseteq Q$. If $0 \in S$ and S is closed under $+$, then* $\mathrm{Rec}(S) = S^*$.

Proof: By the Theorem, $\mathrm{Rec}(S \cup \{1\}) = (S \cup \{1\})^*$. Choose $q \in S \cup \{1\}$ and suppose that $q = \iota^{\mathsf{t}} T^* \phi$ where $T \in M_n(S \cup \{1\})$. We can write $T = C + P$ where $C \in M_n(\mathbb{B})$ and $P \in M_n(S)$. By (2.6),

$$q = \iota^{\mathsf{t}} (C + P)^* \phi = \iota^{\mathsf{t}} (C^* P)^* (C^* \phi).$$

Since S is closed under $+$, $C^*P \in M_n(S)$ and, evidently, $C^*\phi \in \mathbb{B}(\{1,\cdots,n\})$. Hence, $q \in \mathrm{Rec}(S)$ and so $\mathrm{Rec}(S \cup \{1\}) = \mathrm{Rec}(S)$. Furthermore, since $0 \in S$, it is easy to see that $(S \cup \{1\})^* = S^*$. Hence, $\mathrm{Rec}(S) = S^*$.

QED

The treatment we have given follows Conway, [21], and was also influenced by Kuich, [72], who states the result in even greater generality. The essential ideas, but not the theorem itself, can be found in [19], which inexplicably fails to make any reference to the automata theory literature.

3.2 The tropical dioid

The star operation is not a simple one because of its infinitary nature. It is interesting to ask when it can be defined in a finite way. That is, whether

$$a^* = 1 + a + \cdots + a^m, \text{ for some } m. \tag{3.1}$$

In a dioid which is not a quantale, this gives a way of constructing a^*; Gondran and Minoux introduced the notion of m-regularity, $a^{m+1} = a^m$, for just this reason, [46, Definition 1]. In automata theory, (3.1) is known as the finite power property. In 1966, Brzozowski raised the question of whether it was decidable if a given rational set (ie: the language of a finite automaton) had this property.

This celebrated problem was solved in the affirmative independently by Simon, [97], and Hashiguchi, [57]. Simon's proof introduced the tropical dioid, $\mathbb{N}_{\mathsf{min}}$, into automata theory and initiatied a deep exploration of decision problems related to the * operation. For an early survey, see [98]. The basic idea is to use automata with multiplicities in $\mathbb{N}_{\mathsf{min}}$ or, equivalently, semigroups of matrices in $M_n(\mathbb{N}_{\mathsf{min}})$, to reformulate the finite power property as a Burnside problem. Recall that the original Burnside problem asks if a finitely generated group must necessarily be finite if each element has finite order. This is true for groups of matrices over a commutative ring but is false in general. An essential step in Simon's proof is to show that it is also true for semigroups of matrices over $\mathbb{N}_{\mathsf{min}}$, [97, Theorem C].

For further details, the reader cannot do better than turn to the papers in this volume devoted to this subject. The tutorial by Pin, [88], explains in more depth how the tropical dioid enters the picture, while that of Krob, [71], surveys a number of decision problems related to (3.1). D'Alessandro and Sakarovitch show that the finite power property also holds for rational subsets of the free group, [30]. Leung, in his thesis, introduced topological ideas into the study of * problems over the tropical dioid. The topology clarifies the behaviour of q^k as $k \to \infty$ for matrices $q \in M_n(\mathbb{N}_{\mathsf{min}})$. His paper in this volume gives a new and simplified treatment of this approach, [75].

Gaubert has initiated the study of automata with multiplicities in $\mathbb{R}_{\mathsf{max}}$ and has shown that the Burnside problem has a positive answer for semigroups

of matrices in $M_n(\mathbb{R}_{\mathsf{max}})$, [38]. The rationale for introducing such automata is that, just as the tropical dioid can be used to estimate frequencies (how often?), max-plus can be used to estimate durations (how long?). This leads naturally to the next section which studies problems of performance analysis. We move, correspondingly, from $\mathcal{P}(A^*)$ and $\mathbb{N}_{\mathsf{min}}$ to $\mathbb{R}_{\mathsf{max}}$.

4 Discrete Event Systems

4.1 Introduction

4.1.1 Examples and general questions

A discrete event system is one whose behaviour consists of the repeated occurrence of events, [59, Scanning the Issue]. For example: a distributed computing system, in which an event might be the receipt of a message; a digital circuit, in which an event might be a voltage change on a wire; or a manufacturing process, in which an event might be the delivery of a part to a machine.

Our main interest will be in the long-term behaviour of the system. If a denotes an event and $t_i(a)$ denotes the time at which the i-th occurrence of this event takes place then we shall study the asymptotic behaviour of the sequence $t_1(a), t_2(a), \cdots$. Note that, depending on the system, this may have to be formulated stochastically.

It may not be immediately clear that idempotency has anything to contribute to this. In fact, discrete event systems which can be modelled by max-plus matrices have been studied repeatedly, [26, 91, 90, 17, 13, 89, 37], but rarely have the authors directly exploited the idempotency. In this volume, Gürel, Pastravanu and Lewis use max-plus matrices to study manufacturing systems, [56]; Cofer and Garg exploit the partial order structure of $\mathbb{R}_{\mathsf{max}}$ to study supervisory control, [16]; and Cuninghame-Green uses the max-plus polynomials of §2.9 to study the realisability problem, [25].

There are many questions that can be asked about the long-term behaviour of discrete event systems. We shall consider only a few. Do there exist steady or periodic regimes? What form do they take? On average, how quickly does the next occurrence of a take place?

4.1.2 Mathematical models for DES

To answer questions such as those above, it is necessary to have a mathematical description of the system, which specifies a mechanism for determining when events occur. A variety of models have been proposed, [59]. The automata of the previous section are convenient for modelling systems with

state. They allow easy specification of how different choices of event can occur, depending on the current state of the system. To specify temporal behaviour, they can be augmented with information on the duration of events, as in the model proposed by Glasserman and Yao, [42]. A different approach is to specify how events are causally related to each other. An example of this is the classical Gantt chart, or PERT diagram. A somewhat related model is the task-resource model, studied by Gaubert and Mairesse in this volume, [40], in which tasks are specified by their durations and the resources they require. This is sometimes called a tetris model by analogy with the computer game of that name.

In the tetris model, resources are renewable and can be used repeatedly, like a machine in a factory. For consumable resources, a more complex model is required, which combines both state and causal aspects. Petri nets have proved popular in this respect. A Petri net can be defined in terms of a finite set of tasks, Q, and a finite set of resource types, P. Each task, q, consumes and produces a basket of resources, which are specified as elements ${}^\bullet q, q^\bullet$, respectively, of $\mathbb{N}(P)$. The state of the system is given by an element, $c \in \mathbb{N}(P)$, which specifies the current availability of resources of each type. The only tasks, q, which can proceed in state c are those for which there are sufficient resources, ie: ${}^\bullet q \leq c$; if q does proceed, the state of the system changes to $c - {}^\bullet q + q^\bullet$. (Inequality and addition being defined, as usual, pointwise in the semimodule $\mathbb{N}(P)$.) Temporal behaviour can be incorporated by giving each resource a *holding time* for which it must be kept, before becoming available for consumption by a task. Timed Petri nets are the subject of the paper by Cohen, Gaubert and Quadrat in this volume, [18].

One of the difficulties with studying discrete event systems is the variety of different models; there is no consensus on a single fundamental model, [59, Scanning the Issue]. Our approach here will be to extract certain features which appear both conceptually reasonable and common to many systems, and to study a basic model with just these features. Although this model is quite restricted, it can be extended in several ways and used as a building block to model a wide class of discrete event systems. As we shall see, there are many unanswered questions about the basic model itself.

4.1.3 The basic model

Let $\mathbb{R}^n$ denote the space of functions $\mathbb{R}(1, \cdots, n)$, which we think of as column vectors. We use x, y, z for vectors and x_i instead of $x(i)$ for the i-th component of x. $\mathbb{R}^n$ acquires the usual pointwise ordering from the ordering on $\mathbb{R}$: $x \leq y$ if, and only if, $x_i \leq y_i$ for $1 \leq i \leq n$. If $x \in \mathbb{R}^n$ and $h \in \mathbb{R}$ then $x + h$ will denote the vector y for which $y_i = x_i + h$: the operation is performed on each component of the vector. This *vector-scalar* convention extends to equations and inequalities: $x = h$ will mean that $x_i = h$ for $1 \leq i \leq n$.

Definition 4.1 ([54]) *A topical function is a function, $F : \mathbb{R}^n \to \mathbb{R}^n$, such that, if $x, y \in \mathbb{R}^n$ and $h \in \mathbb{R}$, then the following properties hold:*

- monotonicity $x \leq y \implies F(x) \leq F(y)$ M
- (homogeneity) $F(x + h) = F(x) + h$. H

A topical function can be thought of as modelling a system with n events in which $x \in \mathbb{R}^n$ specifies the time of occurrence of each event and $F(x)$ specifies the times of next occurrences. The sequence $x, F(x), F^2(x), \cdots$ is then the sequence of occurrence times of events, whose asymptotic behaviour is the object of study. Definition 4.1 can be understood as follows: if the times of occurrences of some events are increased, then the times of next occurrences of all events cannot decrease; if the times of occurrences are all increased, or decreased, by exactly the same amount for each event, then so too, respectively, are the times of next occurrences. These properties appear very reasonable and can be seen to hold in some form for most discrete event systems.

Functions satisfying the conditions of Definition 4.1 have been introduced and studied independently by other authors, [66], but the material discussed below, as well as the name *topical*, is based on joint work of this author with Keane and Sparrow, [54, 55].

4.1.4 Extensions of the basic model

There are practical systems which can be modelled directly as topical functions, [49, 101], but such systems are rather restricted. They must be closed, or autonomous, in that they require only an initial condition to generate a sequence of occurrence times. Open systems, in contrast, require input to be regularly provided. To model this, it is necessary to work, not with vectors in $\mathbb{R}^n$, but with appropriate functions $\mathbb{R} \to \mathbb{R}^n$, which represent input, or output, histories. This requires an extension of the theory of topical functions to infinite dimensional spaces, a problem studied in [66]. Another extension is to study semigroups of topical functions, which allows the introduction of choice between alternative behaviours, as in an automaton. As remarked at the end of §3.2, Gaubert has made initial investigations in this direction, [38]. Finally, stochastic systems can be modelled by random variables taking values in the space of topical functions. In this volume, Baccelli and Mairesse give an extensive discussion of ergodic theorems for random topical functions, [6], while Gauajal and Jean-Marie study the problem of computing asymptotic quantities in stochastic systems, [41].

These three extensions bring a much wider class of discrete event systems within our scope. It remains an open problem to understand exactly how

much wider. For instance, how can Petri nets be incorporated in this framework? What about the model of Glasserman and Yao?

Before going further, we need to see some concrete examples. We shall then discuss the properties of topical functions, with reference to the questions raised in §4.1.1.

4.2 Examples of topical functions

Let $\mathsf{Top}(n,n)$ denote the set of topical functions, $F : \mathbb{R}^n \to \mathbb{R}^n$. This set has a rich structure, which we need some notation to explain. If a, b are elements of some partially ordered set, let $a \vee b$ and $a \wedge b$ denote the least upper bound and greatest lower bound, respectively, when these exist. The set of functions, $\mathbb{R}^n \to \mathbb{R}^n$ has a natural partial order defined pointwise from that on $\mathbb{R}^n$. Finally, let F^- denotes the function $-F(-x)$.

Lemma 4.1 ([54, Lemma 1.1]) *Let $F, G \in \mathsf{Top}(n,n)$. Let $\lambda, \mu \in \mathbb{R}$ satisfy $\lambda, \mu \geq 0$ and $\lambda + \mu = 1$. Let $c \in \mathbb{R}^n$. Then $FG, F \vee G, F \wedge G, F + c, F^-$ and $\lambda F + \mu G \in \mathsf{Top}(n,n)$.*

This result is an immediate consequence of Definition 4.1. $\mathsf{Top}(n,n)$ is a distributive lattice under $\vee$ and $\wedge$. It is almost a Boolean algebra with F^- as complement but lacks top and bottom elements.

Definition 4.2 *A function $F : \mathbb{R}^n \to \mathbb{R}^n$ is said to be simple if each component, $F_i : \mathbb{R}^n \to \mathbb{R}$, can be written as $F_i(x) = x_j + a$, for some j and some $a \in \mathbb{R}$.*

Simple functions are clearly topical. Lemma 4.1 now provides a mechanism for building topical functions which are not simple.

Definition 4.3 ([52]) *A function $F : \mathbb{R}^n \to \mathbb{R}^n$ is said to be min-max if it can be constructed from simple functions by using only $\vee$ and $\wedge$. Such a function is max-only if it can be built using only $\vee$ and min-only if it can be built using only $\wedge$.*

Max-only functions provide the link to idempotency. Any max-only function can be placed in normal form:

$$F_i(x) := (x_1 + A_{i1}) \vee \cdots \vee (x_n + A_{in})$$

where $A \in M_n(\mathbb{R}_{\mathsf{max}})$. We may then write $F(x) = Ax$. Hence, max-only functions correspond to matrices over $\mathbb{R}_{\mathsf{max}}$ which satisfy the non-degeneracy condition:

$$\forall i,\ \exists j, \text{ such that } A_{ij} \neq -\infty\ . \tag{4.1}$$

A min-only function, dually, corresponds to a matrix over $\mathbb{R}_{\mathsf{min}}$. Min-max functions are nonlinear in the idempotent sense, although in a different way to the polynomials of §2.9.

Lemma 4.2 ([55]) *Choose $F \in \mathsf{Top}(n,n)$ and $a \in \mathbb{R}^n$. There exists a max-only function F_a such that $F(a) = F_a(a)$ and $F \leq F_a$. A dual statement holds for min-only functions.*

Corollary 4.1 ([55]) *Let $F \in \mathsf{Top}(n,n)$. Then F can be written in the two forms:*

$$\bigwedge_{i \in I} G_i = F = \bigvee_{j \in J} H_j \tag{4.2}$$

where G_i are max-only, H_j are min-only and I and J may be uncountably infinite.

Min-max functions are, in a strict sense, finite topical functions. If either of the index sets, I or J, is finite, then F is a min-max function. In this case, the representations in (4.2) can be reduced to essentially unique normal forms, [52, Theorem 2.1]. Min-max functions capture some of the dynamical features of topical functions; see Corollary 4.2.

The infinite topical functions include some well-known functions in disguise. Let $\mathbb{R}^+$ denote the positive reals and $\mathbb{R}^{+0}$ the nonnegative reals. Let $\exp : \mathbb{R}^n \to (\mathbb{R}^+)^n$ and $\log : (\mathbb{R}^+)^n \to \mathbb{R}^n$ be defined componenentwise: $\exp(x)_i = \exp(x_i)$ and $\log(x)_i = \log(x_i)$. These are mutually inverse bijections between $\mathbb{R}^n$ and $(\mathbb{R}^+)^n$. Let $A : (\mathbb{R}^+)^n \to (\mathbb{R}^+)^n$ be any function on the positive cone and let $\mathcal{E}(A) : \mathbb{R}^n \to \mathbb{R}^n$ denote the function $\log(A(\exp))$. The functional $\mathcal{E}$ allows us to transport functions on the positive cone to functions on $\mathbb{R}^n$. Moreover, $\mathcal{E}(AB) = \mathcal{E}(A)\mathcal{E}(B)$, so A and $\mathcal{E}(A)$ have equivalent dynamic behaviour.

Now suppose that A is a nonnegative matrix, $A \in M_n(\mathbb{R}^{+0})$, which satisifes a similiar non-degeneracy condition to (4.1): $\forall i$, $\exists j$, such that $A_{ij} \neq 0$. Then A preserves the positive cone, when elements of $(\mathbb{R}^+)^n$ are interpreted as column vectors in the usual way. The reader can easily check that $\mathcal{E}(A)$ is a topical function. It follows that the theory of topical functions is a generalisation of Perron-Frobenius theory, [83]. It can further be shown, using Lemma 4.1, that a number of problems considered in the optimisation theory literature also fall within the theory of topical functions, [55]. These include problems of deterministic optimal control, Markov decision processes and Leontief substitution systems.

4.3 Nonexpansiveness and perioidicity

If $x \in \mathbb{R}^n$, let $\| x \|$ denote the ℓ^∞ norm of x: $\| x \| = \vee_{1 \leq i \leq n} |x_i|$. A function $F : \mathbb{R}^n \to \mathbb{R}^n$ is nonexpansive in the ℓ^∞ norm if

- $\| F(x) - F(y) \| \leq \| x - y \|$. N

The following observation was first made by Crandall and Tartar, [24]; see also [54, Proposition 1.1] for a proof adapted to the present context.

Proposition 4.1 *If $F : \mathbb{R}^n \to \mathbb{R}^n$ satisfies* H*, then* M *is equivalent to* N*.*

In particular, topical functions are nonexpansive. This constrains their dynamics in ways that are still not understood. We draw the reader's attention here to one result which has significance for discrete event systems.

One of the general questions raised in §4.1.1 concerned the existence of a periodic regime. In the light of the suggested interpretion of topical functions in §4.1.3, a periodic regime can be reasonably formulated as a generalised periodic point: a vector x such that $F^p(x) = x + h$ for some $p > 0$ and some $h \in \mathbb{R}$. (Recall that we are using the vector-scalar convention of §4.1.3.) The system returns after p occurrences with a shift of h in each occurrence time. Because of property H, this behaviour persists. We can, without loss of generality, consider only ordinary periodic points, because $F^p(x) = x + h$ if, and only if, $(F - h/p)^p(x) = x$ and, by Lemma 4.1, $F - h/p$ is topical.

The least p for which $F^p(x) = x$ is the period of F at x. What periods are possible for discrete events systems modelled by topical functions? It turns out, surprisingly, that there is a universal bound on the size of periods which depends only on the dimension of the ambient space.

Theorem 4.1 ([10]) *If $F : \mathbb{R}^n \to \mathbb{R}^n$ is nonexpansive in the ℓ^∞ norm and if p is the period of a periodic point of F, then $p \leq (2n)^n$.*

Results of this form originate in the work of Sine, [99]. An up to date discussion, as well as complete references, can be found in the survey paper by Nussbaum in this volume, [85]. The bound in Theorem 4.1 is not tight: Nussbaum has conjectured that $p \leq 2^n$, and this can be shown to be best possible. The Nussbaum Conjecture has been proved only for $n \leq 3$ and remains the outstanding open problem in this area.

For topical functions, more can be said. The following is an immediate consequence of Lemma 4.2.

Corollary 4.2 ([55]) *Let F be a topical function and $S \subseteq \mathbb{R}^n$ any finite set of vectors. There exists a min-max function G such that $F(s) = G(s)$ for all $s \in S$. In particular, any period of a topical function must be the period of a min-max function.*

It follows that it is sufficient to consider only min-max functions in determining the best upper bound for the periods of topical functions. By augmenting min-max functions with $F(-x)$, where F is min-max, it is possible

to give a similar reduction for general nonexpansive functions, [55]. Gunawardena and Sparrow, in unpublished work, have shown that there are min-max functions in dimension n with period ${}^nC_{[n/2]}$ and conjecture that this is the best upper bound for topical functions.

Fixed points, where $F(x) = x$, are of particular importance for discrete event systems. They represent equilibria, as in (2.4). The existence of fixed points for nonexpansive functions is a classical problem, [43]. For topical functions, Kolokoltsov has given a sufficient condition in terms of a game theoretic representation of topical functions, [66, Theorem 9]. For min-max functions, much stronger results are thought to hold, as in Theorem 4.3. This turns out to be related to the other question raised in §4.1.1, which forms the subject of the next section.

4.4 Cycle times

The rate at which events occur in a discrete event system is an important measure of its performance. The average elapsed time between occurrences, starting from the initial condition $x \in \mathbb{R}^n$, is given by $(F^k(x) - F^{k-1}x + \cdots + F(x) - x)/k$. The asymptotic average, as $k \to \infty$, is then

$$\lim_{k\to\infty} F^k(x)/k \; . \tag{4.3}$$

It is not at all clear that this limit exists in general. Suppose, however, that it does exist for a given F at some initial condition x and that y is some other initial condition. Since F is nonexpansive, $\| F^k(x) - F^k(y) \| \leq \| x - y \|$. Hence, if the limit (4.3) exists anywhere, it must exist everywhere, and must have the same value.

Definition 4.4 *Let $F \in \mathsf{Top}(n,n)$. The cycle time vector of F, $\chi(F) \in \mathbb{R}^n$, is defined as the value of (4.3) if that limit exists for some $x \in \mathbb{R}^n$, and is undefined otherwise.*

If F has a generalised fixed point, $F(x) = x + h$ then, by property H, $F^k(x) = x + kh$. Hence, $\chi(F) = h$. If F is a max-only function, and $A \in M_n(\mathbb{R}_{\mathsf{max}})$ is the corresponding max-plus matrix, then a generalised fixed point of F is simply an eigenvector of A and h is the corresponding eigenvalue. By Proposition 2.6, h is the largest mean circuit weight in $\mathcal{G}(A)$. Similarly, if $F = \mathcal{E}(A)$, where $A \in M_n(\mathbb{R}^{+0})$, then $\exp(x) \in (\mathbb{R}^+)^n$ is an eigenvector of A with eigenvalue $\exp(h)$. In this case $\exp(h)$ is the classical spectral radius of A. We see from this that $\chi(F)$ is a vector generalisation of the notion of eigenvalue, suitable for an arbitrary topical function.

However, if $A \in M_n(\mathbb{R}_{\mathsf{max}})$ is the max-plus matrix corresponding to the max-only function F, then not all eigenvectors of A can be generalised fixed points of F. They must lie in $\mathbb{R}^n$ and hence have no component equal to

$-\infty$. To bring the other eigenvectors into the picture, requires some form of compactification of $\mathbb{R}^n$, analogous to putting the boundary on the positive cone $(\mathbb{R}^+)^n$. It also suggests the existence of other cycle time vectors which give information on fixed points lying in different parts of this boundary. To make sense of this for topical functions remains an open problem.

When does $\chi(F)$ exist? The first indication that this is a difficult question came from the following result of Gunawardena and Keane.

Theorem 4.2 ([54]) *Let $\{a_i\}$, $i \geq 1$, be any sequence of real numbers drawn from the unit interval $[0,1]$. There exists a function $F \in \mathsf{Top}(3,3)$, such that $F^i(0,0,0)_2 = a_1 + \cdots + a_i$.*

In particular, $\chi(F)$ does not always exist and the result indicates the extent of the departure from convergence.

On the other hand, there is strong evidence that $\chi(F)$ does exist for min-max functions. We can formulate this by asking how $\chi(F)$ should behave with respect to the operations which preserve min-max functions. Let $\mathsf{MM}(n,n) \subseteq \mathsf{Top}(n,n)$ denote the set of min-max functions $\mathbb{R}^n \to \mathbb{R}^n$. It is easy to see that $\mathsf{MM}(n,n)$ is closed under all the operations of Lemma 4.1, with the exception of convex combination: $\lambda F + \mu G$. In particular, $\mathsf{MM}(n,n)$ is a distributive lattice. $\mathsf{MM}(n,n)$ also has a Cartesian product structure in which each $F \in \mathsf{MM}(n,n)$ is decomposed into its separate components $(F_1, \cdots, F_n)$. If $S \subseteq A_1 \times \cdots \times A_n$ is a subset of some such product, let $r(S)$ denote its rectangularisation:

$$r(S) = \{u \in A_1 \times \cdots \times A_n \mid \pi_k(u) \in S\},$$

where $\pi_k : A_1 \times \cdots \times A_n \to A_k$ is the projection on the k-th factor. It is always the case that $S \subseteq r(S)$ and if $S = r(S)$ we say that S is a rectangular subset. For example, if $F, G \in \mathsf{MM}(2,2)$ then

$$r\{F,G\} = \{(F_1,F_2),(F_1,G_2),(G_1,F_2),(G_1,G_2)\}.$$

The lattice operations on $\mathsf{MM}(n,n)$ behave well with respect to the Cartesian product structure: if $S \subseteq \mathsf{MM}(n,n)$, then

$$\bigvee_{F \in S} F = \bigvee_{G \in r(S)} G \quad \text{and} \quad \bigwedge_{F \in S} F = \bigwedge_{G \in r(S)} G. \tag{4.4}$$

Conjecture 4.1 (The duality conjecture) $\chi : \mathsf{MM}(n,n) \to \mathbb{R}^n$ *always exists and is a homomorphism of lattices on rectangular subsets: if $S \subseteq \mathsf{MM}(n,n)$ is a finite, rectangular subset, then*

$$\begin{aligned} \chi(\bigvee_{F \in S} F) &= \bigvee_{F \in S} \chi(F) \\ \chi(\bigwedge_{F \in S} F) &= \bigwedge_{F \in S} \chi(F)\,. \end{aligned}$$

Theorem 4.3 ([50]) *Let $F \in \mathsf{MM}(n,n)$. If the duality conjecture is true in dimension n then $F(x) = x + h$ for some x, if, and only if, $\chi(F) = h$.*

Conjecture 4.1 was first stated in a different but equivalent form in [50]. By virtue of (4.4), it gives an algorithm for computing $\chi(F)$ for any min-max function in terms of simple functions, for which χ can be trivally calculated. However, because of the rectangularisation required in (4.4), this algorithm is very infeasible. This raises issues of complexity about which little is known.

The conjecture is known to be true when S consists only of simple functions. That is, when $\bigvee_{F\in S} F$ is a max-only function, and $\bigwedge_{F\in S} F$ is a min-only function. The conjecture has also been proved in dimension 2, where the argument is already non-trivial, [51]. Sparrow has shown that χ exists for min-max functions in dimension 3 but his methods do not establish the full conjecture, [100]. The duality conjecture remains the fundamental open problem in this area and a major roadblock to further progress in understanding topical functions.

The existence of χ does not depend on the finiteness of min-max functions: it also exists for functions $\mathcal{E}(A)$ where $A \in M_n(\mathbb{R}^{+0})$, [55]. At present, we lack even a conjecture as to which topical functions have a cycle time.

We have sketched some of the main results and open problems for topical functions, which we believe are fundamental building blocks for discrete event systems. Topical functions also provide a setting in which max-plus linearity, $A \in M_n(\mathbb{R}_{\mathsf{max}})$, and classical linearity, $A \in M_n(\mathbb{R}^{+0})$, coexist. We shall discuss the curious relationship between these in §6.5. Before that, we must venture into infinite dimensions.

5 Nonlinear partial differential equations

5.1 Introduction

In this section we shall be concerned with scalar nonlinear first order partial differential equations of the form

$$F(x, u(x), Du(x)) = 0 \tag{5.1}$$

where $F : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$. Here, u is a real-valued function on some open subset $X \subseteq \mathbb{R}^n$, $u : X \to \mathbb{R}$, and $Du(x) \in \mathbb{R}^n$ can be thought of as the derivative of u at x, $Du = (\partial u/\partial x_1, \cdots, \partial u/\partial x_n)$.

It may seem implausible that idempotency has anything to say about differential equations: operations like $\max(u, v)$ do not preserve differentiable functions. However, remarkable advances have taken place in our understanding of nonlinear partial differential equations which enable us to give meaning

to solutions of (5.1) which may not be differentiable anywhere. From one perspective, this can be viewed as borrowing ideas from convex analysis. Indeed, the origins of the differential calculus itself go back to Fermat's observation that the maximum of a function $f : \mathbb{R} \to \mathbb{R}$ occurs at points where $df/dx = 0$. Efforts to generalise this to non-differentiable functions have led to new notions of differentiability, [4, Chapter 4], which are closely related to the ideas discussed below, [22, Definition 1]. For ease of exposition, we take a different approach, but convex analysis forms a backdrop to much of what we say and its relationship to idempotency is badly in need of further investigation. Aubin's book provides an excellent foundation, [4].

5.2 Viscosity solutions

The advances mentioned above centre around the concept of *viscosity solutions*, introduced by Crandall and Lions in a seminal paper in 1983. For references, see the more recent *user's guide* of Crandall, Ishii and Lions, [23], which discusses extensions of the theory to second order equations. The earlier survey by Crandall, Evans and Lions, [22], is a model of lucid exposition and can be read, even by non-users, with pleasure and insight.

The word *viscosity* refers to a method of obtaining solutions to (5.1) as limits of solutions of a second order equation with a small parameter—the viscosity—as that parameter goes to zero. We discuss asymptotics further in §6.1. Solutions obtained by this method of vanishing viscosity can be shown to be viscosity solutions in the sense of Crandall and Lions, [22, Theorem 3.1].

Non-differentiable solutions to (5.1) arise in a number of applications. For instance, the initial value problem

$$\begin{aligned} u_t + H(Du) &= 0 \\ u(x,0) &= u_0(x) \end{aligned} \tag{5.2}$$

where $u : \mathbb{R}^{n-1} \times [0,\infty) \to \mathbb{R}$, arises in the context of optimisation. In mechanics it is known as the Hamilton-Jacobi equation, in optimal control as the Bellman equation and in the theory of differential games as the Isaacs equation. The function u represents the optimal value under the appropriate optimisation requirement. In mechanics, the Hamiltonian H and the initial conditions u_0 are often sufficiently smooth to give uniqueness and existence of smooth solutions, at least for small values of t. However, in the other contexts, neither the Hamiltonian nor the initial conditions need be differentiable and are sometimes not even continuous. See, for instance, the Hamiltonian found by Kolokoltsov and Maslov in their paper in this volume, [69], for multicriterial optimisation problems. It becomes important, therefore, to give a convincing account of what it means for u to be a solution of (5.2) or (5.1), when u is not differentiable.

An obvious approach is to restrict solutions to be non-differentiable only on a set of measure zero. Unfortunately, there are usually lots of these. Consider the very simple example of (5.1) with $n = 1$ and $F(x, u, p) = p^2 - 1$. In other words, the equation $(du/dx)^2 = 1$. Suppose that we seek solutions on the interval $[-1, 1]$ which are zero on the boundary. Then $u_1(x) = 1 - |x|$ is differentiable except at $x = 0$ and satisfies the boundary conditions. However, so too does $u_2(x) = -u_1(x)$ and the reader will see that there are infinitely many piecewise differentiable solutions of this form.

The viscosity method cuts through this problem by specifying conditions on the local behaviour of u which isolate one solution out of the many possibilities. The idea is beautifully simple. Let X be an open subset of $\mathbb{R}^n$ and suppose that $\phi : X \to \mathbb{R}$ is a C^1 function such that $u - \phi$ has a local maximum at x_0. If u is differentiable at x_0, then $Du(x_0) = D\phi(x_0)$. If u is not differentiable, why not use $D\phi(x_0)$ as a surrogate for $Du(x_0)$?

Definition 5.1 *An upper semi-continuous (respectively, lower semi-continuous) function $u : X \to \mathbb{R}$ is said to be a viscosity subsolution (respectively, supersolution) of (5.1) if, for all $\phi \in C^1(X)$, whenever $u - \phi$ has a local maximum (respectively, minimum) at $x_0 \in X$, then $F(x_0, u(x_0), D\phi(x_0)) \leq 0$ (respectively, ≥ 0). A viscosity solution is both a subsolution and a supersolution.*

We recall that $u : X \to \mathbb{R}$ is upper semi-continuous at $x \in X$ if $\inf_{U \ni x} \sup_{y \in U} f(y) \leq f(x)$ and lower semi-continuous if $\sup_{U \ni x} \inf_{y \in U} f(y) \geq f(x)$, where U runs through neighbourhoods of x, [4, Chapter 1]. Definition 5.1 interleaves Definition 2 of [22], for first order equations, with Definition 2.2 of [23], for semi-continuous solutions.

Definition 5.1 conceals many subtleties, [23, §2]. For instance, in the example above, u_1 is both a subsolution and a supersolution. There are, in fact, no C^1 functions, ϕ, such that $u_1 - \phi$ has a local minimum at 0 and so the supersolution condition is vacuously satisfied at 0. However, u_2, while it is a subsolution for the same reason, is not a supersolution: the function $u_2 + 1$ has a minimum at 0 but $p^2 - 1 \not\geq 0$ when $p = -1$.

Definition 5.1 leads to elegant and powerful existence, uniqueness and comparison theorems which form the heart of the theory of viscosity solutions, [23]. We mention only the following point, which is relevant to the discussion in §5.5. It is easy to see that for $a \in \mathbb{R}$ and $b \in \mathbb{R}^n$, $u(x, t) = a + b.x - H(b)t$ is a smooth solution of (5.2). Here $b.x$ denotes the standard inner product in $\mathbb{R}^n$. Hopf showed how these linear solutions could be combined to give a generalised global solution of (5.2).

Theorem 5.1 ([61, Theorem 5a]) *Suppose that H is strictly convex and that*

$H(p)/|p| \to \infty$ as $|p| \to \infty$. Suppose further that u_0 is Lipschitz. Then,

$$u(x,t) = \inf_{y \in \mathbb{R}^{n-1}} \left(u_0(y) + \sup_{z \in \mathbb{R}^{n-1}} \{z.(x-y) - tH(z)\} \right) \tag{5.3}$$

satisfies (5.2) almost everywere in $\mathbb{R}^{n-1} \times [0,\infty)$. ([36, Theorem 6.1]) *If u_0 is also bounded then (5.3) is a viscosity solution of (5.2).*

The reader familiar with convex analysis will note that the innermost term in (5.3) can be rewritten as $tL((x-y)/t)$, where $L : \mathbb{R}^{n-1} \to \mathbb{R}_{\min}$, is the Legendre-Fenchel transform of H, [4, Definition 3.1]. L is called the Lagrangian in mechanics. Hence,

$$u(x,t) = \inf_{y \in \mathbb{R}^{n-1}} \{u_0(y) + tL((x-y)/t)\} . \tag{5.4}$$

Kolokoltsov and Maslov have shown that this formula gives a C^1 solution of (5.2) throughout $\mathbb{R}^{n-1} \times [0,\infty)$ under weaker hypothses on H but stronger hypotheses on u_0, [70, §4].

5.3 The role of idempotency

What, then, are the contributions of idempotency to this area? A key intution has been that equations of the form (5.2) should be considered as linear equations over $\mathbb{R}_{\max}$ or $\mathbb{R}_{\min}$. This idea was first put forward by Maslov in the Russian literature in 1984 and later in English translation in [80]. There is a clue to it already in Hopf's *basic lemma*, [61, §2], and it becomes manifest in the early work on viscosity solutions.

Proposition 5.1 ([22, Proposition 1.3(a)]) *Let u, v be viscosity subsolutions (respectively, supersolutions) of (5.1). Then* $\max(u,v)$ *(respectively,* $\min(u,v)$*) is also a viscosity subsolution (respectively, supersolution).*

It follows that for equations in which F is independent of the second variable u—such as (5.2)—the space of viscosity subsolutions is almost a semimodule over $\mathbb{R}_{\max}$; it lacks only a zero element, $u(x) = -\infty$. Notice that supersolutions are linear over $\mathbb{R}_{\min}$ and to get access to viscosity solutions one must work with both dioids.

This linearity is very appealing. It suggests that the Hopf formula (5.4), from an idempotent viewpoint over $\mathbb{R}_{\min}$, is a Green's function representation with kernel $tL((x-y)/t)$. The kernel should arise when the initial condition is a Dirac function at 0 and the solution for general u_0 should then be obtained as a convolution with the kernel, in the usual way. Of course, the notions of Dirac function and convolution must be understood in the idempotent sense. While this idea is very suggestive, it has yet to be fully realised in a convincing

manner. We try to unravel what has been done in this direction in §5.5. More recently, idempotent methods have begun to clarify the nature of viscosity solutions outside the continuous regime. In the next sub-section we develop some of the language needed to discuss these ideas.

5.4 Functional analysis over dioids

5.4.1 Introduction

Let D be a dioid. The aim of this section is to study spaces of functions, $X \to D$, and their linear transformations, when X is an arbitrary set. This is a natural development of the work of earlier sections, in which X was finite. In infinite dimensions, limiting conditions must be imposed, as in classical functional analysis.

5.4.2 Boundedly complete dioids

To begin with, let us use the partial order of the dioid to control the infinite behaviour. We have already had a taste of this with quantales in §2.7 and it is in keeping with the lattice theoretic nature of idempotency. We compare this with more conventional topological methods in §5.4.3.

Definition 5.2 *A boundedly complete (BC) dioid, D, is a dioid in which every subset which is bounded above has a least upper bound, over which the multiplication distributes: $\forall S \subseteq Q$ such that $\exists d \in Q$, for which $x \preceq d$ for all $x \in S$,*

- *there exists a least upper bound* $\sup_{x\in S} x$*;*
- $a.(\sup_{x\in S} x) = \sup_{x\in S}(a.x), \;\; (\sup_{x\in S} x).a = \sup_{x\in S}(x.a)$.

As with quantales, we write $\sum_{x\in S} x = \sup_{x\in S} x$. Let D be a BC dioid and X an arbitrary set. Let $D_b(X)$ denote the set of functions to D which are bounded above: $D_b(X) = \{f : X \to D \mid \exists d \in D,\ f(x) \preceq d,\ \forall x \in X\}$. $D_b(X)$ is a sub-semimodule of $D(X)$. If D is a BC dioid, then $D_b(X)$ inherits bounded completeness from D, which allows us to write any element $u \in D_b(X)$ as a linear combination of characteristic functions. Let $\delta_A \in D_b(X)$ denote the characteristic function of $A \subseteq X$:

$$\delta_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise .} \end{cases}$$

For $a \in X$, $\delta_{\{a\}}$ can be thought of as the idempotent Dirac function at a. When $D = \mathbb{R}_{\min}$, it corresponds to the indicator function of a point as used in convex analysis, [4, Definition 1.2].

Choose $u \in D_b(X)$. Since $u(x) \preceq d$ for some $d \in D$, the functions $u(a).\delta_{\{a\}}(x)$ all lie in $D_b(X)$ and the set of such functions is bounded above by the constant function d. Hence we can write, in $D_b(X)$,

$$u = \sum_{a \in X} u(a).\delta_{\{a\}} \ . \tag{5.5}$$

We are interested in the linear transformations on $D_b(X)$ which preserve infinite sums of this form.

Definition 5.3 *Let D be a BC dioid and M a semimodule over D. M is a boundedly complete (BC) semimodule if every subset of M which is bounded above has a least upper bound over which the scalar multiplication distributes. A homomorphism of BC semimodules, $f : M \to N$, is a homomorphism of semimodules which preserves the least upper bound of any bounded subset.*

As remarked above, $D_b(X)$ is a BC semimodule over D. It would seem from (5.5) that $D_b(X)$ is the free BC semimodule generated by X but this is not quite right. To clarify its properties, consider the following more general construction. Let M be any BC semimodule and let $D_b(X, M) = \{f : X \to M \mid f(x) \preceq m, \text{ for some } m \in M\}$. $D_b(X, M)$ is also a BC semimodule over D and $D_b(X) = D_b(X, D)$. For fixed X, this defines a functor from the category $\mathsf{BC\text{-}SMod}_D$ to itself.

Let $\lambda : D_b(X) \to M$ be a homomorphism of BC semimodules. For each $a \in X$, the value of λ on $\delta_{\{a\}}$ defines a function $\theta_M(\lambda) : X \to M$. Since $\delta_{\{a\}} \preceq \delta_X$, it follows that $\lambda(\delta_{\{a\}}) \preceq \lambda(\delta_X)$, since λ is a homomorphism of semimodules. By definition, $\theta_M(\lambda)(a) = \lambda(\delta_{\{a\}})$ and so $\theta_M(\lambda)$ is bounded above. Hence we have a function $\theta_M : \mathsf{Hom}(D_b(X), M)_{\mathsf{BC-SMod}_D} \to D_b(X, M)$.

Proposition 5.2 *θ_M is an isomorphism of BC semimodules.*

This follows by what category theorists call *general nonsense* and can safely be left to the reader, as can the proofs of the corollaries below.

Corollary 5.1 *Choose $f, g \in D_b(X)$. The pairing $\langle f, g \rangle = \theta_D^{-1}(g)(f)$ defines a nonsingular, boundedly complete inner product $D_b(X) \times D_b(X) \to D$ whose value is given by $\langle f, g \rangle = \sum_{a \in X} f(a).g(a)$.*

The implication is that $D_b(X)$ is self-dual as a BC semimodule. We can also describe the structure of homomorphisms between function spaces. Let $\Lambda : D_b(X) \to D_b(Y)$ be a homomorphism of BC semimodules and choose $(x, y) \in X \times Y$. Let $\eta : \mathsf{Hom}(D_b(X), D_b(Y)) \to D_b(X \times Y)$ be defined by $\eta(\Lambda)(x, y) = \theta_{D_b(Y)}(\Lambda)(x)(y)$.

Corollary 5.2 *η is an isomorphism of BC semimodules. Furthermore, for $f \in D_b(X)$ and $y \in Y$, $\Lambda(f)(y) = \sum_{x \in X} f(x).\eta(\Lambda)(x, y)$.*

If $D = \mathbb{R}_{\min}$ and $X = Y = \mathbb{R}^n$, then the Hopf formula (5.4) has exactly this structure, emphasising once again the linearity of the solutions of (5.2).

Corollary 5.2 is essentially Shubin's kernel representation theorem, [96, Theorem 7.1]. We have reformulated his treatment to bring out its algebraic nature and to highlight BC dioids and BC semimodules. Shubin refers to BC dioids as *regular* and homomorphisms of BC semimodules as *normal*. Akian refers to *locally complete* dioids but does not imply by this that multiplication distributes over infinite sums, [2].

5.4.3 Topologies on dioids

Consider any BC dioid or semimodule. If follows from (2.3) that any *nonempty* subset has an infimum. Hence, the upper and lower limits of a bounded sequence can be defined in the usual way and we can say what it means for a sequence to converge. This defines a topology, which we call the BC topology. (For technical reasons, one must work with generalised sequences indexed over a directed set and not just with sequences indexed over $\mathbb{N}$, [68].)

Lemma 5.1 ([68]) *Let* $\rho : \mathbb{R}_{\min} \times \mathbb{R}_{\min} \to [0, \infty)$ *be defined by* $\rho(x, y) := |\exp(-x) - \exp(-y)|$ *with* $\exp(-\infty) := 0$. *Then* ρ *is a metric whose underlying topology coincides with the BC topology on* $\mathbb{R}_{\min}$.

The metric can extended to function spaces over $\mathbb{R}_{\min}$ in the usual way. Let $f, g \in (\mathbb{R}_{\min})_b(X)$ and define $\rho_X(f, g) := \sup_{x \in X} \rho(f(x), g(x))$. The boundedness of f and g ensures that $\rho_X(f, g) \in [0, \infty)$ and gives a valid metric. We now have two topologies on $(\mathbb{R}_{\min})_b(X)$: the BC topology, corresponding to pointwise convergence, and the metric topology of ρ_X, corresponding to uniform convergence. These are undoubtedly different but the author knows of no results or examples comparing them in the literature.

The treatment of function spaces in §5.4.2, due essentially to Shubin, is satisfying and elegant. It is therefore embarassing to note that homomorphisms of BC semimodules appear not to be continuous in the BC topology. (Nor presumably in the metric topology.) Once again, this issue has not been studied in the literature. However, kernel representations as in Corollary 5.2 have been found for semimodule homomorphisms which are continuous in the metric topology, [70, §2]. The marked discrepancy in hypothesis between results like this and Corollary 5.2 indicates that we are still some way from a full understanding of function spaces and their linear transformation. We can formulate the kernel problem as follows. Let V, W be sub-semimodules of $(\mathbb{R}_{\min})_b(X)$ and $\lambda : V \to W$ a homomorphism of $\mathbb{R}_{\min}$ semimodules. Under what conditions on V, W and λ does λ have a kernel representation as in Corollary 5.2?

5.5 Idempotency and viscosity solutions

From now on we shall work entirely with $\mathbb{R}_{\min}$, equipped with the BC topology. For the remainder of this section, **we shall only use the customary notations in** $\mathbb{R}_{\min}$.

The pairing introduced in Corollary 5.1 provides an elegant way of comparing functions $X \to \mathbb{R}_{\min}$. Let $\Phi \subseteq (\mathbb{R}_{\min})_b(X)$ be a set of test functions. If $f, g \in (\mathbb{R}_{\min})_b(X)$, we say that f and g are equivalent, $f \approx g$, if $\langle f, \phi \rangle = \langle g, \phi \rangle$ for all $\phi \in \Phi$. The equivalence class of f will be denoted $[f]$ and the set of equivalence classes Φ^*.

Each equivalence class contains a unique smallest element, which can be identified as follows. Let $P_\Phi : (\mathbb{R}_{\min})_b(X) \to (\mathbb{R}_{\min})_b(X)$ be given by

$$P_\Phi(f)(x) = \sup_{\phi \in \Phi} \{\langle f, \phi \rangle - \phi(x)\} \tag{5.6}$$

which is well-defined provided $\Phi \neq \emptyset$.

Lemma 5.2 ([45, Theorem 1]) *Suppose that $\Phi \neq \emptyset$. Then $P_\Phi(f) \approx f$ and, if $f \approx g$, then $P_\Phi(f) \leq g$ in $(\mathbb{R}_{\min})_b(X)$.*

Proof: Suppose that $g \approx f$. Choose $\phi \in \Phi$ and $x \in X$. Then, by definition of the pairing, $\langle g, \phi \rangle \leq g(x) + \phi(x)$. Hence, $\langle g, \phi \rangle - \phi(x) \leq g(x)$ and so $\langle f, \phi \rangle - \phi(x) \leq g(x)$, since $g \approx f$. This holds for any ϕ and any $x \in X$ and so $P_\Phi(f) \leq g$. Now choose $\phi \in \Phi$ and $x \in X$ as before. Then, by definition of P_Φ, $\langle f, \phi \rangle - \phi(x) \leq P_\Phi(f)(x)$. Hence, $\langle f, \phi \rangle \leq P_\Phi(f)(x) + \phi(x)$. Since this holds for all $x \in X$, it follws that $\langle f, \phi \rangle \leq \langle P_\Phi(f), \phi \rangle$. But, by the second assertion proved above, $P_\Phi(f) \leq f$, and so by linearity of the pairing, $\langle P_\Phi(f), \phi \rangle \leq \langle f, \phi \rangle$. Hence, $\langle P_\Phi(f), \phi \rangle = \langle f, \phi \rangle$. Since this holds for any $\phi \in \Phi$ it follws that $\mathcal{P}_\Phi(f) \approx f$, as required.

QED

For appropriate sets of test functions, the projector P_Φ corresponds to functionals which are well known in convex analysis, [45].

Let $\lambda : \Phi \to \Phi$ be any function. The pairing $\langle f, g \rangle$ enables us to dualise λ to a function $\lambda^* : \Phi^* \to \Phi^*$ as follows. For any $\phi \in \Phi$ and any $f \in (\mathbb{R}_{\min})_b(X)$, let $\langle \lambda^*[f], \phi \rangle = \langle f, \lambda(\phi) \rangle$. This defines λ^* unambiguously. Note that if $\lambda(\Phi) \not\subseteq \Phi$, then λ^* is not well defined on Φ^*, and may be multiple-valued. λ^* is analogous to the adjoint or transpose in classical linear analysis. If $\Phi = (\mathbb{R}_{\min})_b(X)$ and λ has a kernel representation with kernel $k(x, y)$, then λ^* has kernel $k(y, x)$.

Kolokoltsov and Maslov have introduced a notion of *generalised weak solution* of (5.2). To describe this, we first need to explain the test functions used by them. Let X be a locally compact topological space and let $c^\infty(X) \subseteq (\mathbb{R}_{\min})_b(X)$ denote the set of functions $f : X \to \mathbb{R}_{\min}$ which are bounded below, continuous with respect to the BC topology on $\mathbb{R}_{\min}$ and

which tend to $+\infty$ at infinity. This last condition means simply that, for any $0 < d$, no matter how large, there exists a compact subset $K \subseteq X$, such that $d < f(x)$ whenever $x \notin K$. This is equivalent to f being lower semi-compact (or inf compact as in [45]) in the sense of convex analysis, [4, Definition 1.6].

Lemma 5.3 ([70, §2]) *$P_{c^\infty(X)}(f)$ is the lower semi-continuous closure of f: $P_{c^\infty(X)}(f) = \sup\{g \in (\mathbb{R}_{\min})_b(X) \mid g \text{ continuous and } g \leq f\}$. Furthermore, if f, g are both lower semi-continuous and $f \approx g$ then $f = g$.*

Consider now the initial value problem (5.2). For $0 < t$, let $\lambda_t : (\mathbb{R}_{\min})_b(\mathbb{R}^{n-1}) \to (\mathbb{R}_{\min})_b(\mathbb{R}^{n-1})$ be the time evolution defined by the Hopf formula (5.4):

$$\lambda_t(u)(x) = \inf_{y \in \mathbb{R}^{n-1}} \{u(y) + tL((x-y)/t)\}. \tag{5.7}$$

According to Kolokoltsov and Maslov, a generalised weak solution of (5.2), with initial condition $u_0 \in (\mathbb{R}_{\min})_b(\mathbb{R}^{n-1})$, is $\lambda_t([u_0]) \in c^\infty(X)^*$, [70, §4]. For this to make sense, it is necessary that $\lambda_t^*(c^\infty(X)) \subseteq c^\infty(X)$ where λ_t^* is the transposed operator with kernel $tL((y-x)/t)$. Curiously, this is neither explicitly stated nor proved in [70].

Leaving this issue aside—it is a property of H and not of equation (5.2)—the implication here is that initial conditions with the same lower semi-continuous closure should have the same evolution under (5.2). This does appear to capture an essential feature of (5.2) but the definition itself is unconvincing. The rationale behind it appears to be an analogy with generalised solutions in the classical linear theory. The difficulty is that the operator $\partial_t(-) + H(D(-))$ is not itself defined on Φ^*; it is only the evolution operator (5.7) which is defined there. This differs from the classical theory where differential operators are themselves dualised to act on spaces of distributions, [35, §2.1.6], and one is left in no doubt as to what it means for a distribution to be a solution. It is also not clear why the set of test functions $c^\infty(X)$ should play such a fundamental role with respect to solutions of (5.2).

Maslov and Samborskiĭ have made important progress towards resolving these difficulties for equations of the form $H(Du) = 0$, [81]. They define a notion of derivative on elements of $c^\infty(X)^*$ and prove existence and uniqueness results for an appropriate boundary value problem, [81, Theorems 1 and 2]. In this volume, Samborskiĭ clarifies the structure of the space on which $H(D(-))$ can be considered to act and states further uniqueness results, [93], while Kolokoltsov uses the idea of generalised weak solutions to study a stochastic version of (5.2), [67].

Maslov and Samborskiĭ remark that *"viscosity solutions coincide"* with those given by their results, [81, Remark 4], but do not state the conditions under which such a comparison can be proved. Very recently, Gondran has announced a characterisation of viscosity solutions, [45, Theorem 6], based

on the ideas in [70] and [81]. This is an important development which finally suggests a precise connection between viscosity solutions and idempotent methods. Unfortunately, the announcement contains no proofs! We limit ourselves to a brief statement of Gondran's result, as our final comment on this fascinating subject.

Definition 5.4 *Let X be a set and suppose that $\Phi \subseteq (\mathbb{R}_{\min})_b(X)$ is non empty. The sequence $u_k \in (\mathbb{R}_{\min})_b(X)$ is said to converge weakly from below to $u \in (\mathbb{R}_{\min})_b(X)$ with respect to Φ, denoted $\liminf^w_{k\to\infty} u_k = u$, if $\liminf_{k\to\infty}\langle u_k, \phi\rangle = \langle u, \phi\rangle$, for all $\phi \in \Phi$.*

Weak convergence from above, $\limsup^w_{k\to\infty} u_k$, and weak convergence, $\lim^w_{k\to\infty} u_k$, can be defined in a similar way. Weak convergence was introduced by Maslov, [80].

Theorem 5.2 ([45, Theorem 6]) *Let $X = \mathbb{R}^n$ and $\Phi = c^\infty(\mathbb{R}^n)$. Suppose that, $u_k \in (\mathbb{R}_{\min})_b(\mathbb{R}^n)$ is a sequence of C^1 functions such that $\lim^w_{k\to\infty} u_k = u$, for some $u \in (\mathbb{R}_{\min})_b(\mathbb{R}^n)$. Suppose further that the image sequence $F(x, u_k, Du_k)$ has the property that $\liminf^w_{k\to\infty} F(x, u_k, Du_k) = 0$. Then u is a viscosity supersolution of (5.1). A dual statement over $\mathbb{R}_{\max}$ characterises subsolutions.*

6 Optimisation and large deviations

6.1 Asymptotics

In the previous section we studied nonlinear equations in their own right. We touched on asymptotics in explaining the origins of the word viscosity in §5.2. In fact, asymptotics was a fundamental motivation behind Maslov's work and throws up some intruiguing questions about the nature of idempotency.

The following example, which is taken from [80], gives a good illustration of how asymptotics enters the picture. Consider the heat equation in one dimension, with a small parameter, $0 < h$: $\partial_t = h\partial_x^2 u$. This is a linear equation and if u_1, u_2 are solutions, then so is $a_1u_1 + a_2u_2$ for $a_1, a_2 \in \mathbb{R}$. We can, however, apply a nonlinear transformation, such as $u = \exp(-w/h)$, and thereby arrive at a nonlinear equation

$$\partial_t w + (\partial_x w)^2 - h\partial_x^2 w = 0 \,,$$

which still satisfies a superposition principle. If w_1, w_2 are solutions of this then so is $-h\log(\exp(-(a_1 + w_1)/h) + \exp(-(a_2 + w_2)/h))$. Now let $h \to 0$. We obtain the nonlinear first order equation

$$\partial_t w + (\partial_x w)^2 = 0 \,,$$

which the reader will recognise as a Hamilton-Jacobi equation of the form (5.2) with $H(p) = p^2$. We see further that

$$\lim_{h \to 0^+} -h \log(\exp(-a/h) + \exp(-b/h)) = \min(a, b) \,, \tag{6.1}$$

so that the superposition principle reduces to $\min(a_1 + w_1, a_2 + w_2)$. Once again we have stumbled across the intuition that solutions of (5.2) are linear over $\mathbb{R}_{\mathsf{min}}$.

Limits of the form $\lim_{h \to 0} h \log(\exp(F(h)))$ are studied in a number of areas but particularly in the theory of large deviations; see §6.4. Calculations like that above have given rise to another intuition, arising out of asymptotics, that *idempotency appears in the large deviation limit*, at least over $\mathbb{R}_{\mathsf{max}}$ and $\mathbb{R}_{\mathsf{min}}$. How do we make sense of this?

Maslov, in a seminal book published in 1987 in French translation, [79], put forward the idea of constructing foundations for optimisation theory on the basis of idempotent measures: measures taking values in $\mathbb{R}_{\mathsf{min}}$ and satisfying an appropriate additivity property in $\mathbb{R}_{\mathsf{min}}$. Recent work suggests that this idempotent optimisation theory is the large deviation limit of classical probability theory. In the remaining sections we briefly discuss this circle of ideas.

6.2 Idempotent measures and integration

Let X be a set and choose $c \in (\mathbb{R}_{\mathsf{min}})_b(X)$. If $A \subseteq X$, let $c(A) = \langle \delta_A, c \rangle := \inf_{a \in A} c(a)$. This is the canonical example of an idempotent measure. If $\{A_i \subseteq X \mid i \in I\}$ is any family of subsets, then it follows easily that

$$c\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} c(A_i) \,. \tag{6.2}$$

Unlike conventional measures, which are based on geometric concepts of area, idempotent measures are based on economic concepts of cost and the subsets A_i need not be pairwise disjoint. It is also not necessary, in order for (6.2) to hold, that the index set I be countable, although countable additivity always seems to be assumed.

More generally, one can seek an idempotent measure, k, on a restricted family, $\mathcal{F}$, of subsets, corresponding to a σ-algebra in measure theory. There is no obvious need for $\mathcal{F}$ to be closed under either complementation or intersection, although these properties are often assumed. It is now no longer necessary that a function $c \in (\mathbb{R}_{\mathsf{min}})_b(X)$ exists such that, for each $A \in \mathcal{F}$, $k(A) = \langle \delta_A, c \rangle$. If c does exist, it is called a *density* for k. By Lemma 5.2, there is a unique least such density, given by $k_* = P_\Phi(c)$ where $\Phi = \{\delta_A \mid A \in \mathcal{F}\}$. Formula (5.6) shows that $k_*(x) = \sup_{A \ni x} k(A)$, which gives us a candidate for a density, should one exist. For measures defined on the open sets of

a topology, densities always exist if the topological space is reasonable: for instance, a separable metric space, [2, Theorem 34].

Another way to think about measure is through integration. If $(X, \mathcal{F}, k)$ is a reasonable idempotent measure space and $f \in (\mathbb{R}_{\mathsf{min}})_b(X)$, then the idempotent integral is given by $\int f(x)dk = \langle f, k_* \rangle$. When k is the uniform measure, whose density is the characteristic function of X, this is written $\int f(x)dx$. This integral notation is widely used. It suggests analogoues of various classical theorems on integration, such as those of Riesz, Fubini, etc. For the most part, these are not deep. Fubini's Theorem, for instance, is equivalent to the triviality that $\inf_{a\in A} \inf_{b\in B} f(a,b) = \inf_{b\in B} \inf_{a\in A} f(a,b)$, where $f \in (\mathbb{R}_{\mathsf{min}})_b(X \times Y)$. Akian's result, mentioned in the previous paragraph, tells us that idempotent measures on sufficiently nice topological spaces are always absolutely continuous with respect to the uniform measure. This is different from the classical Radon-Nikodym theorem.

6.3 Optimisation theory

Maslov's conception of an optimisation theory built on idempotent measures has been taken up and developed in the PhD theses of Bellalouna and del Moral and by the Max-Plus Working Group. We will not discuss it in detail here; the reader should refer to the papers in this volume of Akian, Quadrat and Viot, [3], del Moral, [31], and del Moral and Salut, [32]. Quadrat's presentation to the International Congress in Zürich in 1994, [48], includes a useful summary. The theory is in a state of vigorous development; there is not always agreement on terminology and many of the foundational definitions have not stabilised. To what extent it provides a true foundation for optimisation, as probability theory does for the study of random phenomena, is a question that must be addressed elsewhere.

Optimisation theory provides a conceptual framework which runs parallel to probability theory: cost variable, as opposed to random variable; value, as opposed to mean; independence of cost variables; conditional cost; Bellman chains, as opposed to Markov chains; etc. The analogoue of the Gaussian distribution turns out to be $(x-a)^2/2b$, which is stable under inf convolution, as the customary Gaussian is under ordinary convolution, [48]. Analogoues of the laws of large numbers can be formulated, [3, Theorem 6.6], based on various notions of convergence of cost variables.

Weak convergence of cost measures, as in [3, Definition 5.1], is defined to be weak convergence of their densities in the sense of Definition 5.4 but with respect to the set of continuous functions $f \in (\mathbb{R}_{\mathsf{min}})_b(X)$. This is larger than the set of test functions used for viscosity solutions in Theorem 5.2, which had prescribed behaviour at infinity. The differences between various sets of test functions is another foundational question which has not been adequately studied. Weak convergence of cost measures is exactly analogous

to weak convergence of probability measures, [108, Chapter 17].

6.4 Large deviations

Let $\mathbf{r}_i$ be a sequence of independent and identically distributed random variables. One of the basic questions in probability theory concerns the behaviour of the average $\mathbf{s}_k = (\mathbf{r}_1 + \cdots + \mathbf{r}_k)/k$. Under reasonable conditions, $\mathbf{s}_k$ concentrates at the mean of $\mathbf{r}_1$, as $k \to \infty$. Large deviation theory deals with the asymptotics of this process. If $0 < a$, how fast does $\mathsf{P}(\mathbf{s}_k \geq a)$ go to zero as $k \to \infty$? Cramér showed in 1938 that, roughly speaking, it goes as $\exp(-kI(a))$, where $I : \mathbb{R} \to \mathbb{R}$ is a certain *rate function*, [33, §2.2]. I can be constructed from $\mathbf{r}_1$ by the Cramér transform: the Legendre-Fenchel transform of the logarithm of the moment generating function:

$$I(x) = \sup_{t \in \mathbb{R}} \{xt - \log(\mathsf{E}(\exp(t\mathbf{r}_1)))\} \ .$$

More generally, we have the *large deviation principle.*

Definition 6.1 ([33, §1.2]) *Let $\{\mu_\epsilon \mid 0 < \epsilon\}$ be a family of probability measures on the Borel σ-algebra of a topological space X. Let $I \in (\mathbb{R}_{\mathsf{min}})_b(X)$ be lower semi-continuous and satisfy $0 \leq I(x)$ for all $x \in X$. Then $\{\mu_\epsilon\}$ satisfies the large deviation principle with rate function I, if, for any closed set F and open set G of X,*

$$\begin{array}{rcl} \limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(F) & \leq & -\inf_{x \in F} I(x) \\ \liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon(G) & \geq & -\inf_{x \in G} I(x) \ . \end{array} \tag{6.3}$$

The papers of Akian *et al*, [3], and del Moral, [31], shed light on the connection between large deviations and idempotency, but contain no detailed proofs. Theorem 5.2 of [3] gives a necessary and sufficient condition for weak convergence of measures on a metric space, which is strikingly similiar to (6.3). This suggests that the large deviation principle is a form of weak convergence to the rate function. Proposition 6.8 of [3] is asserted to be equivalent to the Gärtner-Ellis theorem, one of the main results of large deviation theory, [33, Theorem 2.3.6]. Proposition 6.8 is analogous to the Lévy convergence theorem: weak convergence of distribution functions corresponds to pointwise convergence of characteristic functions, [108, §18.1]. (Characteristic function is used here in the sense of probability theory, [108, §16.1].) As in probability theory, the central limit theorem for cost variables, [3, Theorem 4.6], falls out as a corollary of Proposition 6.8.

As this account suggests, the results of optimisation theory appear to encapsulate and reformulate the large deviation asymptotics of probability theory.

6.5 Topical functions and asymptotics

There is another way to approach the emergence of idempotency in the large deviation limit. It uses the topical functions of §4.2 and formula (6.1). We noted in §2.9 the close analogy between the spectral theory of max-plus matrices and nonnegative matrices. This seems very similar to the analogy between probability theory and optimisation theory discussed in the previous section but without the large deviation mechanism for moving from the former to the latter.

Let $A \in M_n(\mathbb{R}_{\max})$. For each $0 < h \leq 1$, define the functional $\mathcal{E}^h : M_n(\mathbb{R}_{\max}) \to \mathsf{Top}(n,n)$ by

$$\left(\mathcal{E}^h(A)(x_1,\cdots,x_n)\right)_i := h\log(\sum_{j=1}^{n}\exp((A_{ij}+x_j)/h)) \; .$$

It is easy to check, using Lemma 4.1, that $\mathcal{E}^h(A)$ is a topical function for all $h \in (0,1]$. Furthermore, $\mathcal{E}^1(A) = \mathcal{E}(\exp(A))$, where $\exp(A) \in M_n(\mathbb{R}^{+0})$ is given by $\exp(A)_{ij} = \exp(A_{ij})$, and $\mathcal{E}$ is the functional introduced in §4.2. At the other end, for any fixed $x \in \mathbb{R}^n$, $\lim_{h\to 0^+} \mathcal{E}^h(A)(x) = Ax$, by (6.1).

We see that we have a deformation, within the space of topical functions, between $\mathcal{E}(\exp(A))$, representing the nonnegative matrix $\exp(A)$, and the max-plus matrix A. The class of topical functions provides us with a context within which both extremes, and the intervening space, can be explored.

How do the dynamics and spectral theory of $\mathcal{E}^h(A)$ vary with h? In particular, how do the dynamics and spectral theory of A emerge in the limit as $h \to 0$? Perhaps this can be viewed as a non-commutative version, appropriate for matrices, of the problems discussed in §6.4. On this enigmatic note, we bring our survey of idempotency to a close.

References

[1] S. Abramsky and S. Vickers. Quantales, observational logic and process semantics. *Mathematical Structures in Computer Science*, 3:161–227, 1993.

[2] M. Akian. Densities of idempotent measures and large deviations. Rapport de Recherche 2534, INRIA, April 1995.

[3] M. Akian, J.-P. Quadrat, and M. Viot. Duality between probability and optimization. Appears in [53].

[4] J.-P. Aubin. *Optima and Equilibria*, volume 140 of *Graduate Texts in Mathematics*. Springer-Verlag, 1993.

[5] F. Baccelli, G. Cohen, G. J. Olsder, and J.-P. Quadrat. *Synchronization and Linearity*. Wiley Series in Probability and Mathematical Statistics. John Wiley, 1992.

[6] F. Baccelli and J. Mairesse. Ergodic theorems for stochastic operators and discrete event systems. Appears in [53].

[7] W. Backes. *The structure of longest paths in periodic graphs*. PhD thesis, Universität des Saarlandes, 1994.

[8] R. Bapat. Permanents, max algebra and optimal assignment. *Linear Algebra and its Applications*, 226-228:73–86, 1995.

[9] R. Bapat, D. P. Stanford, and P. van den Driessche. Pattern properties and spectral inequalities in max algebra. *SIAM Journal of Matrix Analysis and Applications*, 16:964–976, 1995.

[10] A. Blokhuis and H. A. Wilbrink. Alternative proof of Sine's theorem on the size of a regular polygon in $\mathbb{R}^n$ with the ℓ^∞ metric. *Discrete Computational Geometry*, 7:433–434, 1992.

[11] T. S. Blyth and M. F. Janowitz. *Residuation Theory*, volume 102 of *International Series of Monographs in Pure and Applied Mathematics*. Pergammon, 1972.

[12] R. E. Burkard, R. A. Cuninghame-Green, and U. Zimmermann, editors. *Algebraic and Combinatorial Methods in Operations Research*, volume 19 of *Annals of Discrete Mathematics*. North-Holland, 1984.

[13] S. M. Burns. *Performance Analysis and Optimization of Asynchronous Circuits*. PhD thesis, California Institute of Technology, 1990.

[14] P. Butkovic. Strong regularity of matrices—a survey of results. *Discrete Applied Mathematics*, 48:45–68, 1994.

[15] B. A. Carré. An algebra for network routing problems. *Journal of the Institute of Mathematics and its Applications*, 7:273–294, 1971.

[16] D. D. Cofer and V. K. Garg. Idempotent structures in the supervisory control of discrete event systems. Appears in [53].

[17] G. Cohen, D. Dubois, J.-P. Quadrat, and M. Viot. A linear-system-theoretic view of discrete-event processes and its use for performance evaluation in manufacturing. *IEEE Transactions on Automatic Control*, 30(3):210–220, 1985.

[18] G. Cohen, S. Gaubert, and J.-P. Quadrat. Algebraic system analysis of timed Petri nets. Appears in [53].

[19] G. Cohen, P. Moller, J.-P. Quadrat, and M. Viot. Algebraic tools for the performance evaluation of discrete event systems. *Proceedings of the IEEE*, 77(1):39–58, 1989.

[20] G. Cohen and J.-P. Quadrat, editors. *11th International Conference on Analysis and Optimization of Systems*. Springer LNCIS 199, 1994.

[21] J. H. Conway. *Regular Algebra and Finite Machines*. Chapman and Hall, 1971.

[22] M. G. Crandall, L. C. Evans, and P.-L. Lions. Some properties of viscosity solutions of Hamilton-Jacobi equations. *Transactions of the AMS*, 282:487–502, 1984.

[23] M. G. Crandall, H. Ishii, and P.-L. Lions. User's guide to viscosity solutions of second order partial differential equations. *Bulletin of the AMS*, 27:1–67, 1992.

[24] M. G. Crandall and L. Tartar. Some relations between nonexpansive and order preserving maps. *Proceedings of the AMS*, 78(3):385–390, 1980.

[25] R. A. Cuninghame-Green. Maxpolynomials and discrete event dynamic systems. Appears in [53].

[26] R. A. Cuninghame-Green. Describing industrial processes with interference and approximating their steady-state behaviour. *Operational Research Quarterly*, 13(1):95–100, 1962.

[27] R. A. Cuninghame-Green. *Minimax Algebra*, volume 166 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, 1979.

[28] R. A. Cuninghame-Green. Minimax algebra and applications. *Advances in Imaging and Electron Physics*, 90:1–121, 1995.

[29] R. A. Cuninghame-Green and P. F. J. Meijer. An algebra for piecewise-linear minimax problems. *Discrete Applied Mathematics*, 2:267–294, 1980.

[30] F. d'Alessandro and J. Sakarovitch. The finite power property for rational sets of the free group. Appears in [53].

[31] P. del Moral. Maslov optimisation theory: topological aspects. Appears in [53].

[32] P. del Moral and G. Salut. Random particle methods in $(\max, +)$ optimization problems. Appears in [53].

[33] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Jones and Bartlett, 1993.

[34] P. I. Dudnikov and S. N. Samborskiĭ. Endomorphisms of finitely generated free semimodules. Appears in [82].

[35] Y. V. Egorov and M. A. Shubin, editors. *Partial Differential Equations I*, volume 30 of *Encyclopaedia of Mathematical Sciences*. Springer-Verlag, 1992.

[36] L. C. Evans. Some min-max methods for the Hamilton-Jacobi equation. *Indiana University Mathematics Journal*, 33:31–50, 1984.

[37] S. Even and S. Rajsbaum. Unison, canon and sluggish clocks in networks controlled by a synchronizer. *Mathematical Systems Theory*, 28:421–435, 1995.

[38] S. Gaubert. On the Burnside problem for semigroups of matrices in the (max,+) algebra. To appear in Semigroup Forum.

[39] S. Gaubert. *Théorie des Systèmes Linéaires dans les dioides*. PhD thesis, L'École Nationale Supérieure des Mines de Paris, 1992.

[40] S. Gaubert and J. Mairesse. Task resource models and (max,+) automata. Appears in [53].

[41] B. Gaujal and A. Jean-Marie. Computational issues in recursive stochastic systems. Appears in [53].

[42] P. Glasserman and D. D. Yao. *Monotone Structure in Discrete Event Systems*. Wiley Series in Probability and Mathematical Statistics. John Wiley, 1994.

[43] K. Goebel and W. A. Kirk. *Topics in Metric Fixed Point Theory*, volume 28 of *Cambridge Studies in Advanced Mathematics*. Cambride University Press, 1990.

[44] J. S. Golan. *The Theory of Semirings with applications in mathematics and theoretical computer science*, volume 54 of *Pitman Monographs and Surveys in Pure and Applied Mathematics*. Lonman Scientific and Technical, 1992.

[45] M. Gondran. Analyse MINPLUS. To appear in C. R. Acad. Sci. Paris, 1996.

[46] M. Gondran and M. Minoux. Linear algebra in dioids: a survey of recent results. Appears in [12].

[47] M. Gondran and M. Minoux. *Graphs and Algorithms.* Wiley-Interscience Series in Discrete Mathematics. John Wiley, 1984.

[48] Max-Plus Working Group. Max-plus algebra and applications to system theory and optimal control. In *Proceedings of the International Congress of Mathematicians, Zürich, 1994.* Birkhäuser, 1995.

[49] J. Gunawardena. Timing analysis of digital circuits and the theory of min-max functions. In *TAU'93, ACM International Workshop on Timing Issues in the Specification and Synthesis of Digital Systems*, September 1993.

[50] J. Gunawardena. Cycle times and fixed points of min-max functions. In G. Cohen and J.-P. Quadrat, editors, *11th International Conference on Analysis and Optimization of Systems*, pages 266–272. Springer LNCIS 199, 1994.

[51] J. Gunawardena. A dynamic approach to timed behaviour. In B. Jonsson and J. Parrow, editors, *CONCUR'94: Concurrency Theory*, pages 178–193. Springer LNCS 836, 1994.

[52] J. Gunawardena. Min-max functions. *Discrete Event Dynamic Systems*, 4:377–406, 1994.

[53] J. Gunawardena, editor. *Idempotency.* Publications of the Isaac Newton Institute. Cambridge University Press, 1996. Proceedings of the 1994 BRIMS Workshop on Idempotency.

[54] J. Gunawardena and M. Keane. On the existence of cycle times for some nonexpansive maps. Technical Report HPL-BRIMS-95-003, Hewlett-Packard Labs, 1995.

[55] J. Gunawardena, M. Keane, and C. Sparrow. Topical functions. In preparation, 1996.

[56] A. Gürel, O. C. Pastravanu, and F. L. Lewis. A system-theoretic approach to discrete event control of manufacturing systems. Appears in [53].

[57] K. Hashiguchi. A decision procedure for the order of regular events. *Theoretical Computer Science*, 8:69–72, 1979.

[58] H. J. A. M. Heijmans. Mathematical morphology: a modern approach in image processing based on algebra and geometry. *SIAM Review*, 37(1):1–36, 1995.

[59] Y. C. Ho, editor. *Special issue on Dynamics of Discrete Event Systems.* Proceedings of the IEEE, 77(1), January 1989.

[60] J. E. Hopcroft and J. D. Ullman. *Introduction to Automata Theory, Languages and Computation.* Addison Wesley, 1979.

[61] E. Hopf. Generalized solutions of non-linear equations of first-order. *Journal of Mathematics and Mechanics*, 14:951–973, 1965.

[62] P. T. Johnstone. *Stone Spaces*, volume 3 of *Studies in Advanced Mathematics*. Cambridge University Press, 1982.

[63] P. T. Johnstone. The point of pointless topology. *Bulletin American Mathematical Society*, 8(1):41–53, 1983.

[64] A. Joyal and M. Tierney. An extension of the Galois theory of Grothendieck. *Memoirs of the AMS*, 51(309), 1984.

[65] S. C. Kleene. Representation of events in nerve nets and finite automata. Appears in [95].

[66] V. N. Kolokoltsov. On linear, additive and homogeneous operators in idempotent analysis. Appears in [82].

[67] V. N. Kolokoltsov. Stochastic HJB equation and WKB method. Appears in [53].

[68] V. N. Kolokoltsov and V. P. Maslov. *Idempotent Analysis.* To be published by Kluwer.

[69] V. N. Kolokoltsov and V. P. Maslov. New differential equation for the dynamics of the Pareto sets. Appears in [53].

[70] V. N. Kolokoltsov and V. P. Maslov. Idempotent analysis as a tool of control theory and optimal synthesis I. *Functional Analysis and Applications*, 23(1):1–14, 1989.

[71] D. Krob. Some automata theoretic aspects of min-max-plus semirings. Appears in [53].

[72] W. Kuich. The Kleene and the Parikh theorem in complete semirings. In *Proceedings ICALP*, pages 212–225. Springer LNCS 267, 1987.

[73] W. Kuich and A. Salomaa. *Semirings, Automata, Languages*, volume 5 of *EATCS Monographs on Theoretical Computer Science.* Springer-Verlag, 1986.

[74] J. Kuntzmann. *Thèorie des Rèseaux.* Dunod, 1972.

[75] H. Leung. The topological approach to limitedness problem on distance automata. Appears in [53].

[76] G. L. Litvinov and V. P. Maslov. Correspondence principle for idempotent calculus and some computer applications. Appears in [53].

[77] S. MacLane. *Categories for the Working Mathematician.* Graduate Texts in Mathematics. Springer-Verlag, 1971.

[78] G. Mascari and M. Pedicini. Types and dynamics in partially additive categories. Appears in [53].

[79] V. P. Maslov. *Méthodes Opératorielles.* Éditions MIR, 1987.

[80] V. P. Maslov. On a new principle of superposition for optimization problems. *Russian Math Surveys*, 42(3):43–54, 1987.

[81] V. P. Maslov and S. N. Samborskiĭ. Stationary Hamilton-Jacobi and Bellman equations (existence and uniqueness of solutions). Appears in [82].

[82] V. P. Maslov and S. N. Samborskiĭ, editors. *Idempotent Analysis*, volume 13 of *Advances in Soviet Mathematics.* American Mathematical Society, 1992.

[83] H. Minc. *Nonnegative Matrices.* Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley, 1988.

[84] C. J. Mulvey. &. *Rend. Circ. Mat. Palermo*, 12:99–104, 1986.

[85] R. D. Nussbaum. Periodic points of nonexpansive maps. Appears in [53].

[86] G. J. Olsder and C. Roos. Cramer and Cayley-Hamilton in max algebra. *Linear Algebra and its Applications*, 101:87–108, 1988.

[87] D. Perrin. Finite automata. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science Volume B*, pages 1–57. Elsevier, 1990.

[88] J.-E. Pin. Tropical semirings. Appears in [53].

[89] S. Rajsbaum and M. Sidi. On the performance of synchronized programs in distributed networks with random processing times and transmission delays. *IEEE Transactions on Parallel and Distributed Systems*, 5:939–950, 1994.

[90] C. V. Ramamoorthy and G. S. Ho. Performance evaluation of asynchronous concurrent systems using Petri nets. *IEEE Transactions on Software Engineering*, SE-6(5):440–449, 1980.

[91] R. Reiter. Scheduling parallel computations. *Journal of the ACM*, 15(4):590–599, 1968.

[92] I. V. Romanovskiĭ. Optimization of stationary control of discrete deterministic process in dynamic programming. *Cybernetics*, 3, 1967.

[93] S. N. Samborskiĭ. Lagrange problem from the point of view of idempotent analysis. Appears in [53].

[94] S. N. Samborskiĭ and A. A. Tarashchan. The Fourier transform and semirings of Pareto sets. Appears in [82].

[95] C. E. Shannon and J. McCarthy, editors. *Automata Studies*. Annals of Mathematics Studies. Princeton University Press, 1956.

[96] M. A. Shubin. Algebraic remarks on idempotent semirings and the kernel theorem in spaces of bounded functions. Appears in [82].

[97] I. Simon. Limited subsets of a free monoid. In *Proceedings 19th IEEE Symposium on FOCS*, pages 143–150. IEEE, 1978.

[98] I. Simon. Recognizable sets with multiplicities in the tropical semiring. In *Proceedings MFCS*, pages 107–120. Springer LNCS 324, 1988.

[99] R. Sine. A nonlinear Perron-Frobenius theorem. *Proceedings of the AMS*, 109:331–336, 1990.

[100] C. Sparrow. Existence of cycle time vectors for minmax functions of dimension 3. Technical Report HPL-BRIMS-96-008, Hewlett-Packard Labs, 1996.

[101] T. Szymanski and N. Shenoy. Verifying clock schedules. In *Digest of Technical Papers of the IEEE International Conference on Computer-Aided Design of Integrated Circuits*, pages 124–131. IEEE Computer Society, 1992.

[102] S. Vickers. *Topology via Logic*, volume 5 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 1989.

[103] E. Wagneur. The geometry of finite dimensional pseudomodules. Appears in [53].

[104] E. Wagneur. Subdirect sum decomposition of finite dimensional pseudomodules. Appears in [20].

[105] E. Wagneur. Moduloïds and pseudomodules. *Discrete Mathematics*, 98:57–73, 1991.

[106] E. A. Walkup and G. Borriello. A general linear max-plus solution technique. Appears in [53].

[107] C. Wende, Q. Xiangdong, and D. Shuhui. The eigen-problem and period analysis of the discrete event system. *Systems Science and Mathematical Sciences*, 3:243–260, 1990.

[108] D. Williams. *Probability with Martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, 1991.

[109] U. Zimmermann. *Linear and Combinatorial Optimization in Orderd Algebraic Structures*, volume 10 of *Annals of Discrete Mathematics*. North-Holland, 1981.