

Algebraic Tools for the Performance Evaluation of Discrete Event Systems*

Guy COHEN[†]

Centre d'Automatique et Informatique, École des Mines de Paris,
35 Rue Saint-Honoré — 77305 Fontainebleau Cédex, France[‡]

Pierre MOLLER Jean-Pierre QUADRAT
Rhône-Poulenc, Lyon, France[§] INRIA, Le Chesnay, France

Michel VIOT
École Polytechnique, Palaiseau, France

Abstract

In this paper, it is shown that a certain class of Petri nets called event graphs can be represented as linear "time-invariant" finite-dimensional systems using some particular algebras. This sets the ground on which a theory of these systems can be developed in a manner which is very analogous to that of conventional linear system theory. Part 2 of the paper is devoted to showing some preliminary basic developments in that direction. Indeed, there are several ways in which one can consider event graphs as linear systems: these ways correspond to approaches in the time domain, in the event domain and in a two-dimensional domain. In each of these approaches, a different algebra has to be used for models to remain linear. However, the common feature of these algebras is that they all fall into the axiomatic definition of "dioids". Therefore, Part 1 of the paper is devoted to a unified presentation of basic algebraic results on dioids.

1 Introduction

Definitions and examples of Discrete Event Dynamic Systems (DEDS) will certainly be found elsewhere in this special issue. But there are several aspects of DEDS on which to focuss one's attention. In this work, we are interested in answering such questions as:

*IEEE Proceedings, Vol. 77, pp. 39-58, 1989

[†]Also with INRIA.

[‡]This is the address for correspondence.

[§]On leave from IIASA, Laxenburg, Austria.

- how many events of a particular type will occur in a certain time interval?
- at which time the n^{th} occurrence of an event of a certain type will happen?...

This is what is usually meant by “performance evaluation”. Apart from discrete event computer simulation which probably remains the most widespread practice, performance evaluation is also the scope of queueing theory [11] and *timed* Petri nets [17]. Whereas the former involves a *stochastic* framework and is devoted to *average long-term* evaluation, the latter is rather *deterministic* and it can deal with *transient* behaviour.

Our approach is also deterministic although there has been already some attempt to extend it to stochastic situations [19, 15]. If one considers a manufacturing workshop for example, it may be argued that, on the short term, the probability of machine breakdown is very low (hopefully!), whereas the service times experienced by parts at all machines are likely to be deterministic (time to drill a hole into a piece of iron, ...). The same holds true in such contexts as performance evaluation of dedicated chips in signal processing... However, it is not in our intention to claim that a purely deterministic modelling of such systems is uniformly realistic. The claim is that this deterministic approach *may* sometimes be more adequate than assuming exponential service times for example. If one is ready to admit this statement (so that our theory might have a nonempty field of application), we believe that it is reasonable to start a theory of DEDS with the simplest situation.

Apart from determinism, simplicity also means *linearity*. Linearity is a most welcome property when it holds true. The major message in this paper is that, for a certain class of DEDS, namely those also modelled by *event graphs* (a special class of Petri nets — see Figure 1 hereafter), linearity appears to be an intrinsic property, provided that one accepts the idea of getting familiar with new algebraic tools. These algebraic structures, that we call *dioids* after Gondran and Minoux [9], present some interesting similarities with our familiar linear algebra, vector spaces, etc... Of course, there are also important differences. But it is striking to see how many features and concepts of conventional linear system theory naturally extend to event graphs. Of course, the practical or intuitive meaning of these concepts should be adapted to the nature of applications covered by DEDS, but the algebraic similarity often goes very far.

This paper comes after several other papers on the same topic. In the earliest approach [3], a state space representation was introduced using variables interpreted as dates (later on called *daters*) which were indexed by event numbers (integers). The dioid $(\mathbb{R} \cup \{-\infty\}, \max, +)$ [7] (or $(\mathbb{Z} \cup \{-\infty\}, \max, +)$ if one prefers to deal with integer, rather than real, dates) was the appropriate algebraic tool for this *event domain* representation. Then, in [4], the corresponding input-output representation, or transfer matrix, was proposed using formal polynomials and series with coefficients in the above dioid. Finally, in [6], it was briefly shown that another state space representation could be obtained using

variables indexed by (discrete) time (this is the usual *time domain* representation), those variables being interpreted as numbers of events (they were thus called *counters*). But for this new model to remain linear, it was necessary to appeal to another dioid, namely $(\mathbb{Z} \cup \{+\infty\}, \min, +)$. Again a corresponding input-output representation could also be proposed. But we found that the best representation was an input-output representation, in a *two-dimensional domain* using series in two formal variables (the backward shift operators in dating, say δ , and in counting, say γ) with boolean coefficients. This gave birth to yet another dioid which will be discussed hereafter. For the event graph of Figure 1, the following representation is obtained

$$y = \mathcal{H}_1 u_1 + \mathcal{H}_2 u_2 = \delta^8(\gamma\delta^2)^* u_1 + \gamma\delta^5(\gamma\delta^2)^* u_2 \quad (1)$$

The reader is referred to the second part of this paper to understand the meaning of this expression.

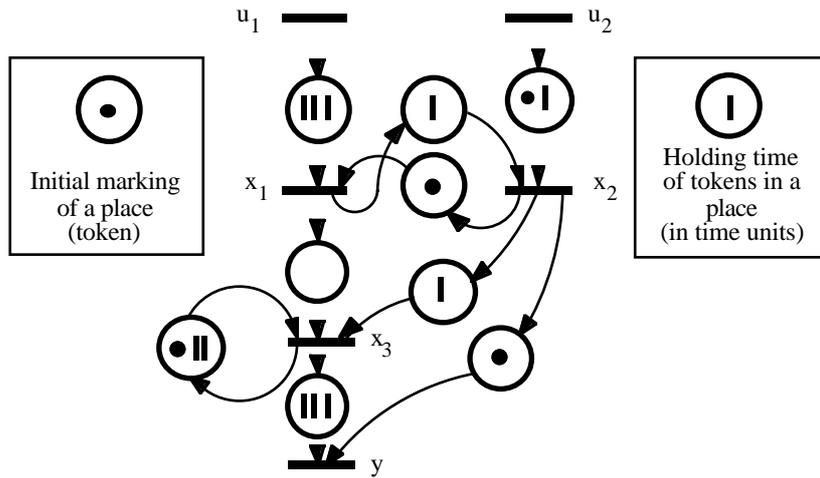


Figure 1: An event graph

From these considerations we draw the following conclusion: event graphs are linear systems on various dioids; therefore, there is an incentive to study these algebraic structures in the most general manner, avoiding the use of properties or axioms which are too specific of one particular dioid and which are not likely to carry over other modelling approaches. It is one objective of this paper to present such an abstract study in a sufficiently self-contained way, so as to serve as a possible basis for further work in this direction. This does not mean that

all the results presented in the first part of this paper are original: a few of them appear here for the first time, as far as we know, but most of them can be found scattered in the literature on dioids (see [9, 7] and references therein), lattice theory and lattice-ordered semigroups [8], and probably in other fields we are not even aware of. It is a difficulty of this emerging theory of (some) DEDES to have to appeal to mathematical tools which are not as familiar to engineers as conventional linear algebra.

The second part of the paper will be devoted to discussing the various modelling approaches of DEDES alluded to hereabove, to summarizing the earlier developments on elements of system theory and to discussing their practical significance. The reader should wait until this second part for finding more concrete applications of the tools and results abstractly introduced in the first part (or he may first scan the second part before coming back to the first). In a conclusion, we try to discuss open problems and directions of future research for this emerging theory.

Part I

Dioid theory

2 Axiomatic definition and basic properties

2.1 Dioid definition

Definition 1 *A set \mathcal{D} supplied with two inner operations denoted by \oplus and \otimes (called “sum” or “addition”, and “product” or “multiplication”) is a dioid if the following axioms are fulfilled:*

Axiom 1 (Associativity)

$$\forall a, b, c \in \mathcal{D}, (a \oplus b) \oplus c = a \oplus (b \oplus c) \quad \text{and} \quad (a \otimes b) \otimes c = a \otimes (b \otimes c)$$

Axiom 2 (Commutativity of addition)

$$\forall a, b \in \mathcal{D}, a \oplus b = b \oplus a$$

Axiom 3 (Distributivity)

$$\forall a, b, c \in \mathcal{D}, (a \oplus b) \otimes c = (a \otimes c) \oplus (b \otimes c)$$

This is right distributivity of product over sum. Since commutativity of product is not a priori required, we do require left distributivity too.

Axiom 4 (Null and identity elements)

$$\exists \varepsilon \in \mathcal{D} : \forall a \in \mathcal{D}, a \oplus \varepsilon = a$$

$$\exists e \in \mathcal{D} : \forall a \in \mathcal{D}, a \otimes e = e \otimes a = a$$

Axiom 5 (Absorbing null element)

$$\forall a \in \mathcal{D}, a \otimes \varepsilon = \varepsilon \otimes a = \varepsilon$$

Axiom 6 (Idempotency of addition)

$$\forall a \in \mathcal{D}, a \oplus a = a$$

□

Definition 2 (Commutative dioid) *A dioid is commutative iff multiplication is commutative.*

Notice In the sequel of this paper, the sign \otimes will be most of the time omitted (or replaced by a dot) as it is usual in conventional algebra.

Comments

- (i) Some of the results hereafter do not depend on Axiom 5 but this axiom seems to be necessary when moving from “scalar” to “matrix” dioids as we shall see later on.
- (ii) Axiom 6 prevents addition from being cancellative¹ unless the dioid reduces to ε ($a \oplus a = a \oplus \varepsilon$ would imply $a = \varepsilon$). Thus dioids cannot be embedded into *rings*. Following Gondran and Minoux [9], we leave the terminology “semiring” sometimes encountered in the literature for the cancellative case.
- (iii) Our definition of a dioid is somewhat less general than that of Gondran and Minoux who require the following weaker substitute for Axiom 6

$$a = b \oplus c \text{ and } b = a \oplus d \implies a = b \tag{2}$$

which would be sufficient for stating Theorem 1 hereafter. In fact, all dioids encountered in the second part of this paper satisfy Axiom 6. An example of a dioid satisfying (2) but not Axiom 6 is $(\mathbb{R}^+, +, \times)$. However this again corresponds to a cancellative addition and it is natural to embed this dioid in $(\mathbb{R}, +, \times)$, that is in our usual algebra.

- (iv) Helbig [10], who himself refers to Zimmermann (see references therein), defines an *extremal algebra* with axioms which are very close but stronger than ours on two points:

- product is commutative;

¹Cancellative means that $a \oplus b = a \oplus c \implies b = c$.

- Axiom 6 is replaced by the stronger one

$$x \oplus y = \text{either } x \text{ or } y.$$

With this new axiom, the order introduced in Theorem 1 hereafter becomes a *total* order. \square

Actually, dioids are algebraic structures somewhere between linear algebra and lattices. As a matter of fact, thanks to Axiom 6, it is possible to associate a partial order relation to addition as shown by Theorem 1 hereafter. With this order relation, a “sup-semilattice” is obtained. Modulo another completeness assumption which involves *infinite* sums, a complete lattice can be obtained by a classical construction which introduces a new inner operation. This is the topic of the next section.

2.2 Lattice properties of dioids

Theorem 1 *In a dioid, the following relation denoted by \geq is a partial order relation.*

$$a \geq b \iff a = a \oplus b \quad (\iff \exists c : a = b \oplus c)$$

Two elements a and b in \mathcal{D} always have an upper bound², namely $a \oplus b$ and ε is the minimal element of \mathcal{D} . Moreover, multiplication is isotone for this order, that is

$$a \geq b \implies \forall c, ac \geq bc$$

and the same for the left product.

With this order, the dioid is a sup-semilattice. A semilattice is complete iff every subset³ has an upper bound. We now give a similar definition of completeness in the context of dioids.

Definition 3 (Complete dioid) *A dioid is complete iff it is closed for infinite sums and Axiom 3 extends to infinite sums.*

With the former requirement, the upper bound of any subset is simply the sum of all its elements. The latter requirement in the above definition is natural and it may be viewed as a *continuity* property of multiplication. There is a maximal element of the dioid denoted by ∞ (the sum of all elements) which is absorbing for addition ($\infty \oplus a = \infty$). The question arises of whether

$$\forall a \in \mathcal{D}, a \neq \varepsilon, \infty \otimes a = a \otimes \infty = \infty \quad (3)$$

(of course $\infty \otimes \varepsilon = \varepsilon$). This can indeed be proved for *Archimedean* dioids.

²In lattice terminology, an upper bound is an element which is $\geq a$ and $\geq b$ and less than, or equal to, every other element sharing the same property. It is unique when it exists.

³finite or infinite

Definition 4 (Archimedean dioid) A dioid is Archimedean iff the following holds true

$$\forall a, b \in \mathcal{D}, \exists c \text{ and } d \in \mathcal{D} : ac \geq b \text{ and } da \geq b$$

Theorem 2 In a complete Archimedean dioid, (3) holds true.

Proof We give the proof for right multiplication by ∞ . From Definition 4, given a , for all b , there exists c_b such that $ac_b \geq b$. One has that

$$a\infty = a\left(\bigoplus_{x \in \mathcal{D}} x\right) \geq a\left(\bigoplus_{b \in \mathcal{D}} c_b\right) = \bigoplus_{b \in \mathcal{D}} ac_b \geq \bigoplus_{b \in \mathcal{D}} b = \infty$$

□

A complete dioid being a complete sup-semilattice, and since there is also a minimal element ε , a new inner operation denoted by \wedge , the lower bound⁴, can be constructed so that the semilattice becomes a complete lattice. This is a classical construction [8, pp. 175–176]. We briefly recall it. For every subset \mathcal{C} of \mathcal{D} , we consider the set

$$\mathcal{E} := \{x \in \mathcal{D} \mid x \leq y, \forall y \in \mathcal{C}\}$$

which is nonempty since it contains at least ε . The lower bound of \mathcal{C} , say z , is defined as the upper bound of \mathcal{E} which exists by assumption. Note that z belongs to \mathcal{E} . As a matter of fact, \mathcal{E} is bounded from above by all $y \in \mathcal{C}$. Since z is less than, or equal to, every element greater than \mathcal{E} (by definition), $z \leq y, \forall y \in \mathcal{C}$ q.e.d.

If $\mathcal{C} = \{a, b, c, \dots\}$, z is denoted by $a \wedge b \wedge c \wedge \dots$. In general, we use the notation $\bigwedge_{x \in \mathcal{C}} x$. This operation \wedge is also associative, commutative, idempotent and has ∞ as “null” element ($\infty \wedge a = a, \forall a$). The following property, called *absorption law*, holds true [8, p. 184]

$$\forall a, b \in \mathcal{D}, a \wedge (a \oplus b) = a \oplus (a \wedge b) = a \quad (4)$$

Remark 1 One has the following equivalences

$$a \geq b \iff a = a \oplus b \iff b = a \wedge b$$

However, this symmetry involves the lattice properties only. From the dioid point of view, one should remember that multiplication is isotone in particular because of Axiom 3. But distributivity will not hold true in general for multiplication over \wedge as one can check for the dioid \mathcal{M} introduced in the second part of this paper. This nonsymmetry is recalled by the notations: $+$ and \times are circled whereas \wedge is not. Nevertheless, since multiplication is isotone, it is easy to check that

$$(a \wedge b)c \leq (ac) \wedge (bc) \quad (5)$$

and the same for left multiplication. □

⁴obeying the dual definition of the upper bound — see footnote 2

Notice that \wedge does not necessarily distribute over \oplus or conversely. We can however state that

$$\forall a, b, c \in \mathcal{D}, \quad (a \wedge b) \oplus c \leq (a \oplus c) \wedge (b \oplus c) \quad (6)$$

$$(a \oplus b) \wedge c \geq (a \wedge c) \oplus (b \wedge c) \quad (7)$$

A lattice is distributive when equality holds true in (6)–(7). Indeed, equality in either (6) or (7) implies the other equality too [8, p. 188]. The following two theorems can also be found in [8, pp. 207 and 212 respectively].

Theorem 3 *A necessary and sufficient condition for a lattice to be distributive is that*

$$\left. \begin{array}{l} a \wedge c = b \wedge c \\ a \oplus c = b \oplus c \end{array} \right\} \implies a = b$$

Theorem 4 *Every multiplicative group \mathcal{G} supplied with an order relation such that multiplication is isotone and \mathcal{G} is a sup-semilattice is a reticulated group and a distributive lattice.*

“Reticulated group” means in particular that multiplication distributes over both the upper and the lower bounds (in our case, over \oplus and \wedge). Hence, in this situation, all desirable properties hold true simultaneously. However, the assumption that a dioid is a multiplicative *group* is rather strong and it will not hold true for the dioid \mathcal{M} considered in the second part.

Definition 5 (Distributive dioid) *A dioid \mathcal{D} is distributive iff it is complete and*

$$\forall \mathcal{C} \subset \mathcal{D}, \quad \forall a \in \mathcal{D}, \quad \left(\bigwedge_{c \in \mathcal{C}} c \right) \oplus a = \bigwedge_{c \in \mathcal{C}} (c \oplus a) \quad (8)$$

$$\left(\bigoplus_{c \in \mathcal{C}} c \right) \wedge a = \bigoplus_{c \in \mathcal{C}} (c \wedge a) \quad (9)$$

Notice that here distributivity is required to extend to infinite subsets too. Moreover, both properties should be required now since one does not imply the other [8, p. 189].

2.3 Matrix dioids

Starting from a “scalar” dioid \mathcal{D} , consider square $n \times n$ matrices with entries in \mathcal{D} . Sum and product of matrices are defined conventionally after the sum and product of scalars in \mathcal{D} . The set of $n \times n$ matrices supplied with these two operations is also a dioid which is denoted by $\mathcal{D}^{n \times n}$. The only point that deserves some attention is the existence of an identity element. Thanks to Axiom 5, the usual identity matrix with entries equal to e on the diagonal and to ε elsewhere is the identity element of $\mathcal{D}^{n \times n}$. This identity matrix will also be denoted by e

and it will be clear from the context which e is meant. In the same way, the null $n \times n$ matrix is denoted by ε .

Notice that if \mathcal{D} is a commutative dioid, this will not be the case for $\mathcal{D}^{n \times n}$ in general. In the same way, if \mathcal{D} is Archimedean, $\mathcal{D}^{n \times n}$ will not be so in general. But since addition of matrices simply involves the addition of similar entries, it is clear that $\mathcal{D}^{n \times n}$ is complete (respectively distributive) whenever \mathcal{D} is so. Also $A \geq B \iff \{a_{ij} \geq b_{ij}, i = 1, \dots, n, j = 1, \dots, n\}$ and $(A \wedge B)_{ij} = a_{ij} \wedge b_{ij}$.

It is sometimes useful to think of $A \in \mathcal{D}^{n \times n}$ as representative of a graph with n nodes and with a directed arc from node i to node j weighted by A_{ji} if this entry of A is different from ε , and with no arc otherwise. The product (in \mathcal{D}) of the weights of arcs composing a path or a circuit is called *weight* of this path or circuit. The number of arcs is called the *length* of the path or circuit. The entry (j, i) of some power A^p of A indicates the maximal weight over all paths of length equal to p and going from i to j . “Maximal” of course refers to the upper bound of the weights of these paths for the order of \mathcal{D} . It is worth noticing that this interpretation holds true because non existing arcs received a weight equal to ε , so that paths that would make use of this non existing arcs are also weighted by ε thanks to Axiom 5, and thus they cannot be maximal.

3 Linear equations in complete dioids

The most general linear equation in a dioid is

$$ax \oplus b = cx \oplus d \tag{10}$$

where $a, b, c, d \in \mathcal{D}$ or $\mathcal{D}^{n \times n}$ and x is the unknown in the same dioid. Throughout this section, \mathcal{D} is supposed to be a complete dioid. Hereafter, we only consider particular subclasses of this general equation which will prove useful in the second part of this paper.

3.1 The equation $ax \oplus b = c$ and residuation

In \mathcal{D} (or $\mathcal{D}^{n \times n}$), we first consider the equation in x

$$ax \oplus b = c \tag{11}$$

This equation does not always admit a solution but it may also have several or an infinity of solutions. This may as well happen in usual linear algebra when, for example, a is a non invertible matrix. Pseudo-inverses are then considered. Here, a first obvious necessary condition for existence is that $b \leq c$. It is met in particular if $b = \varepsilon$. Even in this case, existence and uniqueness are not guaranteed. In order to recover existence and uniqueness when $b \leq c$, we shall modify the concept of “solution” to (11) and we shall retain that of “greatest subsolution”.

Definition 6 A subsolution of (11) is an x such that $ax \oplus b \leq c$.

Theorem 5 The set of subsolutions of (11), say \mathcal{S} , is not empty iff $b \leq c$. Then the upper bound of \mathcal{S} belongs to \mathcal{S} and it is also the greatest subsolution of

$$ax = c \quad (12)$$

Proof The condition $b \leq c$ is an immediate consequence of (11) and Theorem 1. Conversely, if it is satisfied, then \mathcal{S} contains at least ε thanks to Axiom 5. Let z be the upper bound of \mathcal{S} . The following proves that z belongs to \mathcal{S}

$$az = a\left(\bigoplus_{y \in \mathcal{S}} y\right) = \bigoplus_{y \in \mathcal{S}} ay \leq c$$

Finally, we prove that \mathcal{S} is identical to \mathcal{S}' which is defined as the set of subsolutions of (12). On the one hand, $y \in \mathcal{S} \Rightarrow ay \oplus b \leq c \Rightarrow ay \leq c \Rightarrow y \in \mathcal{S}'$. On the other hand, $y \in \mathcal{S}' \Rightarrow ay \leq c \Rightarrow ay \oplus b \leq b \oplus c = c$ since $\{b \leq c \Leftrightarrow c = c \oplus b\}$. Hence $y \in \mathcal{S}$. \square

From now on, we thus limit ourselves to the study of equation (12).

Definition 7 (Residuation) The (left) residue of c by a , denoted by $a \setminus c$, is defined as the greatest subsolution of (12).

The terminology comes from lattice-ordered semigroups (see [8, pp. 220]).

Remark 2 Since we do not assume that \mathcal{D} is commutative, it is possible to consider equations of the form $xa = c$ and the corresponding right pseudo-inverse solution would be denoted by c/a .

Remark 3 It should be kept in mind that one cannot in general associate a particular element of \mathcal{D} (supposedly denoted by a^b) to the mapping $c \mapsto a \setminus c$, in such a way that $a \setminus c = a^b \otimes c$, for the reason that that mapping is not “linear” in \mathcal{D} as we are going to see it in the study of the new operation \setminus .

Theorem 6 If (12) has a true solution, then $a \setminus c$ is also a solution and it is the greatest one.

The proof is straightforward. As a corollary, if a^{-1} exists, then the solution is unique and $a \setminus c = a^{-1}c$.

Theorem 7 The expression $a \setminus b$ is nonincreasing as a function of a and nondecreasing as a function of b . Moreover the following inequalities and equalities hold true

$$a(a \setminus b) \leq b \quad (13)$$

$$(a \setminus a) \geq e \quad (14)$$

$$a(a \setminus a) = a \quad (15)$$

$$e \setminus a = a \quad (16)$$

$$\varepsilon \setminus a = \infty \quad (17)$$

$$(a \setminus b)c \leq a \setminus (bc) \quad (18)$$

$$a \setminus (b \setminus c) = (ba) \setminus c \quad (19)$$

$$(a \setminus b) \oplus (a \setminus c) \leq a \setminus (b \oplus c) \quad (20)$$

$$(a \setminus b) \wedge (c \setminus b) = (a \oplus c) \setminus b \quad (21)$$

$$(a \setminus b) \wedge (a \setminus c) = a \setminus (b \wedge c) \quad (22)$$

$$(a \setminus b) \oplus (c \setminus b) \leq (a \wedge c) \setminus b \quad (23)$$

Proof Most of the inequalities and of some equalities are readily derived from the very definition of residuation. We only give a proof for the non obvious equalities. For (15), two opposite inequalities are derived from (13) and (14).

For (19), let $x = b \setminus c, y = a \setminus x, z = (ba) \setminus c$. On the one hand, $\{bx \leq c \text{ and } ay \leq x\} \Rightarrow \{bay \leq bx \leq c\} \Rightarrow \{y \leq z\}$. On the other hand, $\{baz \leq c\} \Rightarrow \{az \leq x\} \Rightarrow \{z \leq y\}$.

For (21), let $x = (a \oplus c) \setminus b, y = a \setminus b, z = c \setminus b$. On the one hand, $\{(a \oplus c)x \leq b\} \Rightarrow \{ax \leq b \text{ and } cx \leq b\} \Rightarrow \{x \leq y \text{ and } x \leq z\} \Rightarrow \{x \leq y \wedge z\}$. On the other hand, $\{ay \leq b \text{ and } cz \leq b\} \Rightarrow \{a(y \wedge z) \leq b \text{ and } c(y \wedge z) \leq b\} \Rightarrow \{(a \oplus c)(y \wedge z) \leq b\} \Rightarrow \{y \wedge z \leq x\}$.

For (22), let $x = a \setminus b, y = a \setminus c, z = a \setminus (b \wedge c)$. On the one hand, $\{ax \leq b \text{ and } ay \leq c\} \Rightarrow \{a(x \wedge y) \leq ax \wedge ay \leq b \wedge c\} \Rightarrow \{x \wedge y \leq z\}$. On the other hand, $\{az \leq b \wedge c\} \Rightarrow \{az \leq b \text{ and } az \leq c\} \Rightarrow \{z \leq x \text{ and } z \leq y\} \Rightarrow \{z \leq x \wedge y\}$. \square

Comments (15) is a familiar equality for pseudo-inverses; (16) says that e is a neutral element for residuation; (18) and (20) shows that $a \setminus b$ is not a linear function of b ; (22) means that \setminus is distributive over \wedge . When a^{-1} and b^{-1} exist, an interesting consequence of (21) is the formula⁵

$$(a \oplus b) \setminus e = a^{-1} \wedge b^{-1} \quad (24)$$

In a (complete) commutative dioid, again with (21), one has that

$$\begin{aligned} (ab) / (a \oplus b) = (a \oplus b) \setminus (ab) &= (a \setminus (ab)) \wedge (b \setminus (ba)) \\ &\geq ((a \setminus a)b) \wedge ((b \setminus b)a) && \text{from (18)} \\ &\geq b \wedge a && \text{from (14)} \end{aligned}$$

but obviously equalities hold throughout when a^{-1} and b^{-1} exist. \square

⁵Recall that $-\max(a, b) = \min(-a, -b)$.

3.2 Matrix residuation

Till now, all that have been said about residuation applies to a complete “scalar” dioid \mathcal{D} or to a “matrix” dioid $\mathcal{D}^{n \times n}$. The following theorem relates the residue of $B \in \mathcal{D}^{n \times n}$ by $A \in \mathcal{D}^{n \times n}$ (also denoted by $A \setminus B$) to the scalar residuation.

Theorem 8 *Let $A, B \in \mathcal{D}^{n \times n}$, then*

$$A \setminus B = A^T \odot B \quad (25)$$

where A^T denotes the transpose of A and \odot is a new matrix product where the operations \oplus and \otimes of \mathcal{D} are replaced respectively by \wedge and \setminus of \mathcal{D} (recall that \setminus distributes over \wedge from (22)).

Proof We consider $n = 2$ without loss of generality. Moreover, it is obvious that each pair of corresponding columns of $X = A \setminus B$ and B can be considered independently of other pairs. For example, the first column $X_{\cdot 1}$ must satisfy

$$A_{11}X_{11} \oplus A_{12}X_{21} \leq B_{11} \quad (26)$$

$$A_{21}X_{11} \oplus A_{22}X_{21} \leq B_{21} \quad (27)$$

which implies

$$A_{11}X_{11} \leq B_{11} \quad \text{and} \quad A_{21}X_{11} \leq B_{21}$$

$$A_{12}X_{21} \leq B_{11} \quad \text{and} \quad A_{22}X_{21} \leq B_{21}$$

or

$$X_{11} \leq Y_{11} := (A_{11} \setminus B_{11}) \wedge (A_{21} \setminus B_{21})$$

$$X_{21} \leq Y_{21} := (A_{12} \setminus B_{11}) \wedge (A_{22} \setminus B_{21})$$

On the other hand, Y_{11} and Y_{21} also verify (26)-(27), which means that indeed $Y_{\cdot 1} = X_{\cdot 1}$ by definition of X . Let us check this for (26) for example

$$\begin{aligned} A_{11}[(A_{11} \setminus B_{11}) \wedge (A_{21} \setminus B_{21})] \oplus A_{12}[(A_{12} \setminus B_{11}) \wedge (A_{22} \setminus B_{21})] &\leq A_{11}(A_{11} \setminus B_{11}) \oplus A_{12}(A_{12} \setminus B_{11}) \\ &\leq B_{11} \quad \text{from (13)} \end{aligned}$$

This completes the proof. \square

This theorem extends a result provided by Cuninghame-Green [7] in the context of the dioid $(\mathbb{R} \cup \{-\infty, +\infty\}, \max, +)$ (where $\wedge = \min$).

3.3 A particular transformation

We now come back to an arbitrary (complete) dioid \mathcal{D} and we show an interesting transformation which, in some situation, reduces residuation to multiplication. An application of this will be encountered in the second part of this

paper. Let φ be an arbitrary but fixed element in \mathcal{D} and let us consider the transformation

$$\forall a \in \mathcal{D}, \tilde{a} := a \setminus \varphi \quad (28)$$

Notice that $\widetilde{a \oplus b} = \tilde{a} \wedge \tilde{b}$ from (21) which shows that this transformation is an homomorphism from the sup-semilattice (\mathcal{D}, \oplus) to the inf-semilattice (\mathcal{D}, \wedge) . As a direct consequence of (19) and (28), we also have that

$$\forall a, b \in \mathcal{D}, \tilde{ba} = a \setminus \tilde{b} \quad (29)$$

Let us interpret \tilde{x} as a way of “coding” x . We see that, if we have to perform the *left* residuation of the “coded” b by a , we can rather multiply b itself by a to the *right* and then code the result. Of course, the process of coding itself involves residuation, but it may be quite simple for elements (at the “denominator”) with can be written as a sum of invertible elements thanks to (21) (this is what happen in the second part of this paper). However, even in this case, this does not mean that we can *always* replace residuation by product since there may be a problem with “decoding”: the transformation is not a bijection in general.

Everything above applies as well to a matrix dioid $\mathcal{D}^{n \times n}$. Let us however discuss the relationship between scalar and matrix coding. If a scalar coding has been defined by some φ , we define the $n \times n$ matrix

$$\Phi = \begin{pmatrix} \varphi & \infty & \dots & \infty \\ \infty & \varphi & & \vdots \\ \vdots & & \ddots & \vdots \\ \infty & \dots & \dots & \varphi \end{pmatrix} \quad (30)$$

and we define \tilde{X} as $X \setminus \Phi$. Using (25), it is easy to check that (say, for $n = 2$)

$$X = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \implies \tilde{X} = \begin{pmatrix} \tilde{a} & \tilde{c} \\ \tilde{b} & \tilde{d} \end{pmatrix}$$

where of course *e.g.* \tilde{b} corresponds to scalar coding of b with φ . Notice the transposition in the formula above: coding a column vector amounts to coding its entries and transposing, the result being a row vector.

Finally, as a consequence of (25) and (29), we have that

$$\widetilde{BA} = A \setminus \tilde{B} = A^T \odot \tilde{B}$$

3.4 The equation $x = ax \oplus b$ and the “star” operation

We consider the following implicit equation in x

$$x = ax \oplus b \quad (31)$$

Let

$$a^* := e \oplus a \oplus a^2 \oplus \dots \quad (32)$$

Interesting, and easy to check, properties of this *star* operation are

$$(a^*)^p = a^*, \forall p \in \mathbb{N} \quad \text{and} \quad (a^*)^* = a^* \quad (33)$$

Let also $a^+ := a \otimes a^*$ and notice that

$$a^* = e \oplus a^+ \quad \text{and} \quad a^* \geq a^+ \quad (34)$$

Theorem 9

- (i) a^*b is the least solution of (31).
- (ii) For all solution x , one has that $x = a^*x$.

Proof

- (i) Clearly, if x is a solution, $x \geq b$ and $x \geq ax$. On the one hand, $a^p x \geq a^p b$. On the other hand, $x \geq ax \geq \dots \geq a^p x$. Hence, $x \geq a^p b$, thus $x \geq \bigoplus_{p=0}^{\infty} a^p b = a^*b$. Finally, it is straightforward to check that a^*b is also a solution.
- (ii) From (31), one gets that $x = a(ax \oplus b) \oplus b = \dots = a^p x \oplus (e \oplus \dots \oplus a^{p-1})b$ for all p . Summing up all these equalities for $p \in \mathbb{N}$ yields $x = a^*x \oplus a^*b$. But from (i) above, $a^*x \geq a^*b$ or $a^*x \oplus a^*b = a^*x$ which completes the proof. \square

Theorem 9 answers the issue of existence of solutions to (31) in complete dioids. About uniqueness, we observe that if the homogeneous equation

$$x = ax \quad (35)$$

has a solution $y \neq \varepsilon$, yz is also a solution of (35), $\forall z \in \mathcal{D}$, and obviously $a^*b \oplus yz$ is a solution of the nonhomogeneous equation (31). In usual linear algebra, it is well known that *all* solutions of the nonhomogeneous equation are obtained by adding all solutions of the homogeneous equation to a particular solution of the nonhomogeneous. Here, for such a result to hold true, the “particular solution” has to be the *least* one (namely a^*b) because “adding” also means “increasing”.

Theorem 10 *Let \mathcal{D} be a distributive dioid. Then, if x is a solution of (31), it can be written as $x = y \oplus a^*b$ where y is a solution of (35).*

Proof For any given solution x of (31), we consider the subset $\mathcal{C}_x := \{c \mid x = c \oplus a^*b\}$. If $c \in \mathcal{C}_x$, then $ac \in \mathcal{C}_x$ too. As a matter of fact, $x = c \oplus a^*b \Rightarrow ax = ac \oplus a^+b \Rightarrow x = ax \oplus b = ac \oplus a^*b$ from the equality in (34). Also, if $c, d \in \mathcal{C}_x$, then $c \oplus d \in \mathcal{C}_x$ since $x = x \oplus x = c \oplus d \oplus a^*b$. Therefore, if $c \in \mathcal{C}_x$, then $a^*c \in \mathcal{C}_x$ too.

Let now $z := \bigwedge_{c \in \mathcal{C}_x} c$. Since \mathcal{D} is distributive, $z \oplus a^*b = \bigwedge_{c \in \mathcal{C}_x} (c \oplus a^*b) = x$, which means that $z \in \mathcal{C}_x$. Hence, $az \in \mathcal{C}_x$ too. From the very definition of z , it follows that $z \leq az$ and thus $a^*z \leq a^+z$. But the inequality in (34) shows that indeed $y = ay$ if we set $y := a^*z$. To complete the proof, notice that $y \in \mathcal{C}_x$, as observed earlier. \square

3.5 The homogeneous equation $x = ax$

It is possible to characterize all solutions of (35) in the situations described by Theorems 11 and 12 hereafter. Let us first introduce a new notation for an *invertible* a , that is when there exists a^{-1} such that $a \otimes a^{-1} = a^{-1} \otimes a = e$.

$$a^\# := a^* \otimes (a^{-1})^* = (a \oplus a^{-1})^* = a^* \oplus (a^{-1})^* \quad (36)$$

These equalities can be checked easily as the following additional properties

$$a \otimes a^\# = a^* \otimes a^\# = a^\# \quad (37)$$

Theorem 11 (Scalar case) *Consider Equation (35) with $a, x \in \mathcal{D}$, where \mathcal{D} is a distributive and commutative dioid. Moreover, it is assumed that a can be written as a finite sum of invertible elements $\{a_k \mid k = 1, \dots, p\}$. Then, any solution x of (35) is of the form*

$$x = \bigoplus_{k=1}^p a_k^\# y_k a^* \quad (38)$$

where the y_k 's are arbitrary elements of \mathcal{D} .

Proof The proof is by induction on the integer p . For $p = 1$, if $x = ax$, then $x = a^*x$, but also $x = a^{-1}x$, thus $x = (a^{-1})^*x$, and finally $x = a^\#x$. Conversely, if $x = a^\#y$ for any y , from (37) it follows that $x = ax$. Notice that since $a^\# = a^\#a^*$, the theorem is proved for $p = 1$.

Assume that the statement of the theorem is true up to some value $p - 1$. If $x = (a_p \oplus b)x$ where $b = \bigoplus_{k=1}^{p-1} a_k$, from Theorem 9-(ii), one has that $x = a_p^*x$. Now, using Theorem 10 and this Theorem 11 for $p = 1$ (already proved), one has that $x = a^*bx \oplus a_p^\#y_p$ for some $y_p \in \mathcal{D}$. Using commutativity and the fact that $x = a_p^*x$, one gets that $x = bx \oplus a_p^\#y_p$. Again using Theorem 10 and the induction assumption, one concludes that $x = b^*a_p^\#y_p \oplus (\bigoplus_{k=1}^{p-1} a_k^\#y_k b^*)$ for some y_k 's. The desired formula (38) is obtained by multiplying this last equality by a_p^* , by remembering that $x = a_p^*x$, and by noticing that $a^* = a_p^*b^*$ from formula (49) established *independently* later on.

Conversely, if x is given by (38), we must show that $x = ax$. Actually, it suffices to prove this for any x corresponding to all y_k null, except one, say y_i taken equal to e , since then, by linearity, the statement will hold true for all linear combinations of such x 's. If $x = a_i^\#a^*$, then $ax = a_i^\#a^+ \leq x$ from (34).

On the other hand, $ax \geq a_i x$ but $a_i x = x$ from (37), hence $ax \geq x$, and finally $ax = x$. \square

We are now going to give an extension of this theorem in the situation when a is a $n \times n$ matrix, denoted now by $A \in \mathcal{D}^{n \times n}$, and x is a $n \times 1$ column vector denoted by \vec{x} . Notice that we could as well consider the unknown $X \in \mathcal{D}^{n \times n}$ since, in equation $X = AX$, there is no interaction between the columns of X which may thus be considered separately as in

$$\vec{x} = A\vec{x} \quad (39)$$

Notice also that $\mathcal{D}^{n \times n}$ is not commutative in general, and therefore Theorem 11 cannot be applied directly.

Theorem 12 (Vector case) *Consider the vector equation (39) on a dioid \mathcal{D} which is subject to the same assumptions as in the previous theorem. Moreover, it is assumed that each entry of A which is not equal to ε , say A_{ij} , can be written as a finite sum of invertible elements $\bigoplus_{k=1}^{p_{ij}} A_{ijk}$. A directed arc from node j to node i is associated to each such A_{ijk} , with this A_{ijk} as the weight. This arc is in parallel with the other arcs from j to i . Let $\Gamma(A)$ (or simply Γ) be the set of elementary circuits of this graph. For all circuit $\kappa \in \Gamma$, the weight w_κ is invertible. Then, any solution \vec{x} of (39) is of the form*

$$\vec{x} = \bigoplus_{\kappa \in \Gamma} w_\kappa^\# y_\kappa (A^*)_{.i_\kappa} \quad (40)$$

where the y_κ 's are arbitrary elements of \mathcal{D} , $(A^*)_{.i_\kappa}$ is the i_κ^{th} column of A^* and i_κ is the number of any node belonging to the circuit κ .

The proof of this theorem is, in its present form, rather lengthy and it will be skipped here. Let us just explain why an arbitrary node of the circuit κ may be picked up.

Lemma 1 *Let i and j be two nodes belonging to the same circuit of weight w (assumed invertible). Then*

$$w^\# A_{.j}^* = w^\# A_{ij}^* A_{.i}^* \quad \text{and} \quad w^\# A_{.i}^* = w^\# A_{.i}^+$$

which means in particular that two columns of $w^\# A^*$ (or of $w^\# A^+$) corresponding to nodes on the same circuit of weight w are proportional.

Proof From $A^* = A^* A^*$, we have that $(A^*)_{ki} \geq (A^*)_{kj} (A^*)_{ji}$ for all i, j, k . Hence, letting k vary

$$(A^*)_{.i} \geq (A^*)_{ji} (A^*)_{.j} \quad (41)$$

On the other hand, if i and j are on the same circuit of weight w , we certainly have that $(A^*)_{ji} (A^*)_{ij} \geq w$, hence from (37)

$$w^\# (A^*)_{ji} (A^*)_{ij} \geq w^\# \quad (42)$$

Then

$$\begin{aligned}
w^\#(A^*)_{ij}(A^*)_{.i} &\leq w^\#(A^*)_{.j} && \text{from (41),} \\
&&& \text{inverting the roles of } i \text{ and } j \\
&\leq w^\#(A^*)_{ji}(A^*)_{ij}(A^*)_{.j} && \text{from (42)} \\
&\leq w^\#(A^*)_{ij}(A^*)_{.i} && \text{from (41)}
\end{aligned}$$

Therefore equality holds throughout, proving the former statement.

As for the latter, observe that A^* and A^+ may differ only on the diagonal. Obviously, if node i is on a circuit of weight w , then $(A^+)_{ii} \geq w$, hence $w^\#(A^+)_{ii} \geq w^\#$ or $w^\#(A^+)_{ii} = w^\# \oplus w^\#(A^+)_{ii} = w^\#(e \oplus (A^+)_{ii}) = w^\#(A^*)_{ii}$ which completes the proof. \square

4 Rational closure and rational representations

4.1 Rational closure and rational calculus

We consider a complete dioid \mathcal{D} and a subset \mathcal{E} of \mathcal{D} which contains ε and e . Considering Equation (31) with data a and b in \mathcal{E} , the least solution a^*b exists in \mathcal{D} but not necessarily in \mathcal{E} .

Definition 8 (Rational closure) *The rational closure of \mathcal{E} (denoted by \mathcal{E}^*) is the smallest subset of \mathcal{D} which contains \mathcal{E} and all finite sums, products and star operations over its elements. A subset $\mathcal{E} \subset \mathcal{D}$ (containing ε and e) is rationally closed iff $\mathcal{E}^* = \mathcal{E}$.*

The definition of \mathcal{E}^* implies that \mathcal{E}^* is a *subdioid* of \mathcal{D} since it is stable for addition and multiplication. Obviously, $a^*b \in \mathcal{E}^*$ when $a, b \in \mathcal{E}$. Observe also that $(\mathcal{E}^*)^*$ is the same as \mathcal{E}^* .

Let us go now from “scalars” to matrices. On the one hand, we can consider the subset $\mathcal{E}^{n \times n} \subset \mathcal{D}^{n \times n}$ of $n \times n$ matrices with entries in \mathcal{E} and its rational closure $(\mathcal{E}^{n \times n})^*$. This is a subdioid of $\mathcal{D}^{n \times n}$. On the other hand, we can consider the subdioid $(\mathcal{E}^*)^{n \times n} \subset \mathcal{D}^{n \times n}$ made of $n \times n$ matrices with entries in \mathcal{E}^* .

Theorem 13 *The subdioid $(\mathcal{E}^{n \times n})^*$ is included in the subdioid $(\mathcal{E}^*)^{n \times n}$.*

This theorem will be improved later on (the inclusion is in fact an equality), but it is stated here in this weaker form because this result will be needed soon. The proof is based on the following technical lemma.

Lemma 2 *For a matrix A partitioned into four blocks, one has that*

$$A^* = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^* = \begin{pmatrix} a^* \oplus a^*b(ca^*b \oplus d)^*ca^* & a^*b(ca^*b \oplus d)^* \\ (ca^*b \oplus d)^*ca^* & (ca^*b \oplus d)^* \end{pmatrix} \quad (43)$$

Outline of proof Remember that the *star* operation is related to the least solution of an implicit equation of type (31). It is possible to write down the block-system of equations corresponding to $X = AX \oplus e$ (remember that e is the identity matrix) and to solve it in a progressive manner, as in Gaussian elimination. Placing the partial solutions already obtained in the next block-equation preserves the property of least solution since all operations involved are isotone. Formula (43) corresponds to solving blocks (1,1) and (1,2) and then (2,1) and (2,2). \square

Notice that another path to solve the system could be (2,1) and (2,2) and then (1,1) and (1,2) which is equivalent to interchanging the role of a (resp. b) and d (resp. c). This observation leads to the identity

$$(ca^*b \oplus d)^* = d^* \oplus d^*c(bd^*c \oplus a)^*bd^* \quad (44)$$

Proof of Theorem 13 The sums and products of matrices in $\mathcal{E}^{n \times n}$ obviously belong to $(\mathcal{E}^*)^{n \times n}$. To prove that $(\mathcal{E}^{n \times n})^*$ is included in $(\mathcal{E}^*)^{n \times n}$, it remains to prove that we do not get outside of this latter set when performing *star* operations over elements of $\mathcal{E}^{n \times n}$. This is done by induction over the dimension n . The statement holds true for $n = 1$. Assuming it holds true up to some n , let us prove it for $n + 1$. It suffices to consider a partitioning of an element of $\mathcal{E}^{(n+1) \times (n+1)}$ into blocks such that a is in $\mathcal{E}^{n \times n}$. By inspection of (43), and using the induction assumption, the proof is easily completed. \square

4.2 Rational representations

We are now going to establish some results on representations of rational elements. For reasons that will become more apparent later on, we distinguish two particular subsets of \mathcal{E} , namely \mathcal{B} and \mathcal{C} . There is no special requirement about them except that we always assume that they both contain ε and e . In particular, we allow \mathcal{B} and \mathcal{C} to be overlapping and even identical.

Theorem 14 \mathcal{E}^* coincides with the set of elements x which can be written as

$$x = C_x A_x^* B_x \quad (45)$$

where B_x is an $n_x \times 1$ matrix with entries in \mathcal{B} , n_x being an arbitrary but finite integer number, C_x is a $1 \times n_x$ matrix with entries in \mathcal{C} and A_x is an $n_x \times n_x$ matrix with entries in \mathcal{E} .

For short, a representation of x like (45) will be called a $(\mathcal{B}, \mathcal{C})$ -representation.

Proof Let \mathcal{S} be the subset of all elements of \mathcal{D} having a $(\mathcal{B}, \mathcal{C})$ -representation. \mathcal{S} includes \mathcal{E} because of the following identity

$$x = \begin{pmatrix} e & \varepsilon \end{pmatrix} \begin{pmatrix} \varepsilon & x \\ \varepsilon & \varepsilon \end{pmatrix}^* \begin{pmatrix} \varepsilon \\ e \end{pmatrix} \quad (46)$$

Suppose that we have proved that \mathcal{S} is stable by addition, multiplication and star operation (what we postpone to the end of this proof), then \mathcal{S} is equal to its rational closure \mathcal{S}^* . Since \mathcal{S} includes \mathcal{E} , \mathcal{S}^* includes \mathcal{E}^* . On the other hand, from Theorem 13, we see that A_x^* has its entries in \mathcal{E}^* . From (45), it is thus clear that \mathcal{S} is included in \mathcal{E}^* . Finally, we conclude that $\mathcal{S} = \mathcal{S}^* = \mathcal{E}^*$.

For the proof to be complete, we have to prove that considering two elements of \mathcal{S} , say x and y , which, by definition, have an $(\mathcal{B}, \mathcal{C})$ -representation, $x \oplus y$, $x \otimes y$ and x^* also have a $(\mathcal{B}, \mathcal{C})$ -representation. This is a consequence of the following formulas

$$\begin{aligned} C_x A_x^* B_x \oplus C_y A_y^* B_y &= \begin{pmatrix} C_x & C_y \end{pmatrix} \begin{pmatrix} A_x & \varepsilon \\ \varepsilon & A_y \end{pmatrix}^* \begin{pmatrix} B_x \\ B_y \end{pmatrix} \\ C_x A_x^* B_x \otimes C_y A_y^* B_y &= \begin{pmatrix} C_x & \varepsilon & \varepsilon \end{pmatrix} \begin{pmatrix} A_x & B_x & \varepsilon \\ \varepsilon & \varepsilon & C_y \\ \varepsilon & \varepsilon & A_y \end{pmatrix}^* \begin{pmatrix} \varepsilon \\ \varepsilon \\ B_y \end{pmatrix} \\ (C_x A_x^* B_x)^* &= \begin{pmatrix} \varepsilon & e \end{pmatrix} \begin{pmatrix} A_x & B_x \\ C_x & \varepsilon \end{pmatrix}^* \begin{pmatrix} \varepsilon \\ e \end{pmatrix} \end{aligned}$$

These formulas are proved by making repeated uses of (43). \square

Theorem 15 *The subdoids $(\mathcal{E}^{n \times n})^*$ and $(\mathcal{E}^*)^{n \times n}$ are identical. Consequently, $(\mathcal{E}^*)^{n \times n}$ is rationally closed.*

Proof The inclusion in one direction has been stated in Theorem 13. Therefore, we need only to prove the reverse inclusion. Let $x \in (\mathcal{E}^*)^{n \times n}$ and assume that $n = 2$ for the sake and simplicity and without loss of generality. Then x can be written as

$$x = \begin{pmatrix} x_1 & x_2 \\ x_3 & x_4 \end{pmatrix}$$

with entries $x_i \in \mathcal{E}^*$. Each x_i has a $(\mathcal{B}, \mathcal{C})$ -representation with matrices A, B, C indexed by a subscript i , A_i being of dimension $n_i \times n_i$. We have that

$$\begin{aligned} x &= \begin{pmatrix} C_1 A_1^* B_1 & C_2 A_2^* B_2 \\ C_3 A_3^* B_3 & C_4 A_4^* B_4 \end{pmatrix} \\ &= \begin{pmatrix} C_1 & C_2 & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & C_3 & C_4 \end{pmatrix} \begin{pmatrix} A_1 & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & A_2 & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & A_3 & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon & A_4 \end{pmatrix}^* \begin{pmatrix} B_1 & \varepsilon \\ \varepsilon & B_2 \\ B_3 & \varepsilon \\ \varepsilon & B_4 \end{pmatrix} \end{aligned}$$

The inner dimension is $\sum_{i=1}^4 n_i$ which is augmented to the next multiple of 2 (and more generally of n) by adding enough rows and columns of ε 's in the matrices. Then, since the outer dimension is 2 and the inner dimension is now a multiple of 2, by appropriately partitioning these matrices into 2×2 blocks,

one can consider this representation as a $(\mathcal{B}, \mathcal{C})$ -representation, but with 2×2 -dimensional entries. This proves, applying Theorem 14 in the dioid $\mathcal{D}^{2 \times 2}$, that x belongs to $(\mathcal{E}^{2 \times 2})^*$. \square

Comments

- (i) The inner dimension n_x for any $x \in \mathcal{E}^*$ in a representation like (45) depends of course on the subsets \mathcal{B} and \mathcal{C} in which the matrices B_x and C_x are allowed to take their entries. Unfortunately, for fixed \mathcal{B} and \mathcal{C} , we did not solve the problem of the “*minimal*” representation. Therefore, we do not know how to give to this n_x an intrinsic meaning. But it may be expected — see hereafter — that the larger \mathcal{B} and \mathcal{C} are, the smaller n_x can be. As an extreme situation, let us consider $\mathcal{B} = \mathcal{C} = \{\varepsilon, e\}$. Indeed, this is the minimal subsets we may consider (recall that \mathcal{B} and \mathcal{C} must at least includes ε and e). “The” corresponding n_x will then be “maximal” and it could be a canonical measure of “complexity” of the rational element x
- (ii) More generally, let us consider a family of nested subsets $\mathcal{B}_1 \subset \dots \subset \mathcal{B}_i \subset \dots$ and $\mathcal{C}_1 \subset \dots \subset \mathcal{C}_i \subset \dots$. For a given $x \in \mathcal{E}^*$, we get a corresponding family of $(\mathcal{B}_i, \mathcal{C}_i)$ -representations (whose inner dimensions are denoted by n_i). Obviously, a $(\mathcal{B}_i, \mathcal{C}_i)$ -representation is also a $(\mathcal{B}_j, \mathcal{C}_j)$ -representation if $j > i$ (hence $n_i \geq n_j$, speaking of hypothetical minimal representations in all cases). On the other hand, it is interesting to see how one can pass from a $(\mathcal{B}_j, \mathcal{C}_j)$ - to a $(\mathcal{B}_i, \mathcal{C}_i)$ -representation. Let (A_j, B_j, C_j) be the former representation. Then, B_j (resp. C_j) can be split up into $B_i \oplus B'$ (resp. $C_i \oplus C'$), such that B_i (resp. C_i) has its entries in \mathcal{B}_i (resp. \mathcal{C}_i)⁶. The following formula can be checked using (43)

$$\begin{aligned} C_j A_j^* B_j &= (C_i \oplus C') A_j^* (B_i \oplus B') \\ &= \begin{pmatrix} C_i & \varepsilon & e \end{pmatrix} \begin{pmatrix} A_j & B' & \varepsilon \\ \varepsilon & \varepsilon & \varepsilon \\ C' & \varepsilon & \varepsilon \end{pmatrix}^* \begin{pmatrix} B_i \\ e \\ \varepsilon \end{pmatrix} \end{aligned}$$

Observe that the last expression is a $(\mathcal{B}_i, \mathcal{C}_i)$ -representation. \square

4.3 Yet other rational representations

So far, we have considered representations of elements of \mathcal{E}^* by triples of matrices (A, B, C) , such that the entries of A are taken in \mathcal{E} , whereas those of B (resp. C) are allowed to lie in an arbitrary subset \mathcal{B} (resp. \mathcal{C}) of \mathcal{E} containing at least $\{\varepsilon, e\}$. Recall that \mathcal{B} and \mathcal{C} need not be distinct nor even disjoint.

For motivations that are to be found in the second part of this paper, we are going to consider other possibilities for the entries of A, B, C . Namely, in

⁶ B_i and C_i may as well be null matrices.

addition to \mathcal{B} and \mathcal{C} on which assumptions remain the same, we consider a “covering” $(\mathcal{F}, \mathcal{G})$ of \mathcal{E} (that is $\mathcal{E} = \mathcal{F} \cup \mathcal{G}$ but $\mathcal{F} \cap \mathcal{G}$ needs not be empty). We always assume that $\varepsilon, e \in \mathcal{F}$ if we want to consider \mathcal{F}^* . Generally speaking, considering two subsets \mathcal{S} and \mathcal{F} , we introduce the following notations

$$\mathcal{F}^* \otimes \mathcal{S} := \{x \mid x = \bigoplus_{i=1}^p a_i b_i \text{ for some finite } p \in \mathbb{N}, a_i \in \mathcal{F}^*, b_i \in \mathcal{S}\}$$

The notation $\mathcal{S} \otimes \mathcal{F}^*$ is similarly defined. Notice that ε belongs to the subsets so defined.

Theorem 16 \mathcal{E}^* coincides with the set of elements x which can be written as in (45) but with entries of A_x lying in $\mathcal{F}^* \otimes \mathcal{G}$, those of B_x in $\mathcal{F}^* \otimes \mathcal{B}$ and those of C_x in \mathcal{C} . We call this an “observer” representation.

Alternately, there exist other representations such that the entries of A_x are in $\mathcal{G} \otimes \mathcal{F}^*$, those of B_x are in \mathcal{B} and those of C_x are in $\mathcal{C} \otimes \mathcal{F}^*$. We call these “controller” representations.

Proof Only the former statement will be proved. The latter can be proved similarly. We first prove that if $x \in \mathcal{E}^*$, then x does have an observer representation. From Theorem 14, we know that x has a $(\mathcal{B}, \mathcal{C})$ -representation. Consider the matrix A of this $(\mathcal{B}, \mathcal{C})$ -representation⁷ which can be written $A^{\mathcal{F}} \oplus A^{\mathcal{G}}$ such that $A^{\mathcal{F}}$ contains only elements of \mathcal{F} and $A^{\mathcal{G}}$ only elements of \mathcal{G} ⁸. Therefore, we have $x = C(A^{\mathcal{F}} \oplus A^{\mathcal{G}})^* B$. An observer representation is derived by making use of the identity

$$(c \oplus d)^* = (d^* c)^* d^* \quad (47)$$

which is a direct consequence of (44) (let $a = \varepsilon$ and $b = e$ therein).

Conversely, if x has an observer representation (A, B, C) , then $x \in \mathcal{E}^*$. As a matter of fact, it is easy to realize that the entries of A, B, C lie in subsets of \mathcal{E}^* . Recall also that $(\mathcal{E}^*)^* = \mathcal{E}^*$. The conclusion follows. \square

Remark 4 Another form of (47) is obtained by letting $a = \varepsilon$ and $c = e$ in (44)

$$(b \oplus d)^* = d^* (b d^*)^* \quad (48)$$

5 Rational representations in commutative dioids

In the previous section, we obtained representations of elements of \mathcal{E}^* involving a single *star* operation, but on a square matrix of arbitrary dimension. In (complete) commutative dioids (Definition2), we are going to see that rational elements can be represented with a single level of *star* on “scalar” elements.

⁷We drop the subscript x .

⁸If $\mathcal{F} \cap \mathcal{G}$ is not empty, entries of A in the intersection may be arbitrarily put into either $A^{\mathcal{F}}$ or $A^{\mathcal{G}}$, or even in both matrices thanks to Axiom 6.

Lemma 3 *In a complete commutative dioid \mathcal{D} , one has that*

$$\forall a, b \in \mathcal{D}, \quad (a \oplus b)^* = a^*b^* \quad (49)$$

Certainly the simplest way of proving this formula is by direct calculation using the very definition (32) of *star* and commutativity. But we are going to indicate how to get it as a corollary of a result which is worth mentioning. This result is stated without proof (the proof is similar to that of Theorem 9 albeit a bit more involved).

Theorem 17 *In a complete (not necessarily commutative) dioid \mathcal{D} , consider the following implicit equation in the unknown x*

$$x = ax \oplus xb \oplus c \quad (50)$$

*Then, a^*cb^* is the smallest solution of (50).*

If commutativity is assumed, clearly formula (49) follows. With this formula at hand, identity (43) can be given a new useful form, at least when d is a scalar (i.e. a 1×1 block).

Lemma 4 *In a commutative dioid, for a matrix A partitioned into four blocks where $A_{22} = d$ is 1×1 , one has that*

$$A^* = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^* = \begin{pmatrix} a^* \oplus d^*a^*bc(a \oplus bc)^* & d^*(a \oplus bc)^*b \\ d^*c(a \oplus bc)^* & d^* \oplus d^*c(a \oplus bc)^*b \end{pmatrix} \quad (51)$$

Proof Since d and ca^*b are scalars, using (49), one gets that $(ca^*b \oplus d)^* = (ca^*b)^*d^*$. Moreover, from (44) with $d = \varepsilon$, it comes that $(ca^*b)^* = e \oplus c(a \oplus bc)^*b$. With these two equalities, the lower right-hand block of (51) can be derived from the corresponding block of (43).

Considering now the upper right-hand block of (43), $a^*b(ca^*b \oplus d)^* = d^*a^*b(e \oplus c(a \oplus bc)^*b)$, from an equality above and (49). Then, with (48), $(a \oplus bc)^* = a^*(bca^*)^*$. Hence

$$\begin{aligned} a^*b(ca^*b \oplus d)^* &= d^*a^*(e \oplus bca^*(bca^*)^*)b \\ &= d^*a^*(bca^*)^*b \\ &= d^*(a \oplus bc)^*b \end{aligned}$$

Similar calculations yield the other blocks of (51). □

Theorem 18 *Let $A \in \mathcal{D}^{n \times n}$ where \mathcal{D} is a (complete) commutative dioid. Then all entries of A^* are finite sums of the form $\bigoplus_i a_i(b_i)^*$, where a_i is a finite product of entries of A and b_i is a finite sum of weights of circuits of the graph associated to A (see Section 2.3).*

Proof The proof is by induction. The statement is true for $n = 1$. Suppose that it holds also true up to $n - 1$. Consider the partitioning (51) of A with d being 1×1 . Notice that to matrix bc is associated a graph whose all circuits are of length 2, starting from one of the first $n - 1$ nodes, going to the n^{th} node and coming back to the initial node. These of course are among the circuits of the graph associated to A . Considering the formula (51), the proof is easily completed by noticing that products of stars of “scalar” elements are converted to stars of sums of these elements using (49). \square

Comments The claim of this theorem is essentially that the stars contained in the entries of A^* are only those of weights of circuits of the associated graph. This result is intuitively appealing, given the interpretation of the entries of the powers A^p of A (see Section 2.3) and the fact that for $p \geq n$, the paths of length p necessarily contain circuits. \square

Theorem 19 *Let \mathcal{E} be a subdioid of the commutative dioid \mathcal{D} , that is a subset of \mathcal{D} containing ε and e and which is stable for finite sums and products of its elements. Then, \mathcal{E}^* coincides with the set of elements x which can be written as*

$$x = \bigoplus_{i=1}^{m_x} a_i(b_i)^* \quad (52)$$

where m_x is an arbitrary finite integer and $a_i, b_i \in \mathcal{E}$.

This is a straightforward consequence of Theorems 14 and 18.

6 Equivalence modulo z in commutative dioids

Considering a commutative dioid \mathcal{D} , we are going to consider the quotient of this dioid by an equivalence relation of a particular form which will appear to be of special interest in the second part of this paper.

Definition 9 *Let $a, b, z \in \mathcal{D}$, a commutative dioid. We say that a and b are equivalent (or equal) modulo z , which is denoted by $a \equiv b \pmod{z}$ iff $az^* = bz^*$.*

Theorem 20 *$a \equiv b \pmod{z}$ is an equivalence relation. Let $[a]_z$ denote the equivalence class of a . The equivalence relation is compatible with the dioid structure, that is $[a \oplus b]_z$ and $[a \otimes b]_z$ depend only on $[a]_z$ and $[b]_z$ (and not on the particular a and b in these classes). Then, we can set*

$$[a]_z \oplus [b]_z := [a \oplus b]_z \quad \text{and} \quad [a]_z \otimes [b]_z := [a \otimes b]_z$$

which defines a (commutative) dioid structure for the quotient denoted by \mathcal{D}/z . Finally, each equivalence class $[a]_z$ has a greatest element which is equal to az^* .

Proof The relation is obviously reflexive, symmetric and transitive. Because of Axiom 3, it is easy to see that $[a \oplus b]_z$ can be defined after any representatives of $[a]_z$ and $[b]_z$. The same holds true for multiplication because $(az^*)(bz^*) = (ab)z^*$ using associativity and commutativity of product and the fact that $(z^*)^2 = z^*$. It is easy to check that \mathcal{D}/z is a dioid with the above mentioned sum and product. Finally, since $z^* \geq e$, az^* , which does not depend on the particular representative a of $[a]_z$ by definition, is not less than every such a . \square

Comments

- (i) Because of the last statement of Theorem 9, \mathcal{D}/z can be identified to the dioid denoted by $z^*\mathcal{D}$ which consists of the set of elements obtained by multiplying all the elements of \mathcal{D} by z^* . Whenever $z \leq e$, then $z^* = e$ and $\mathcal{D}/z = \mathcal{D}$.
- (ii) More generally, thanks to (49), we have the commutative diagram of Figure 2. This implies in particular that $\mathcal{D}/(y \oplus z)$ is equal to \mathcal{D}/z whenever $z \geq y$. \square

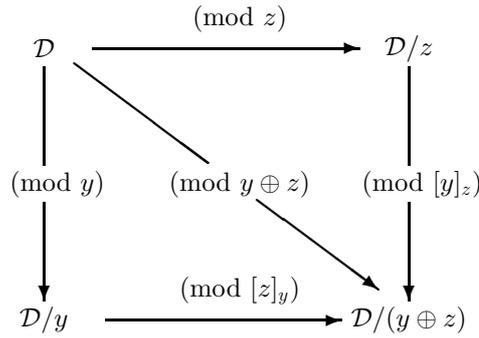


Figure 2: Commutative diagram

Remark 5 Since $\mathcal{D}^{n \times n}$ is not a commutative dioid in general, there is no point to consider $\mathcal{D}^{n \times n}/z$ with $z \in \mathcal{D}^{n \times n}$. But we may of course speak of the matrix dioid $(\mathcal{D}/z)^{n \times n}$ with $z \in \mathcal{D}$. \square

Part II

System theory

7 Event graphs, daters and counters

7.1 Timed event graphs

Event graphs are a particular class of Petri nets in which there is a single transition upstream, and a single transition downstream every place. See an example on Figure 1.

Notice In general, a place between transitions x (upstream) and y (downstream) will be labelled $y|x$ (observe that y is mentioned *before* x). This may be ambiguous if there are several arcs in parallel from x to y , in which case an additional index (e.g. numbering these parallel places from left to right) would be needed. \square

Situations as those of Figure 3-(a)-(b) are not allowed in event graphs. Case (b) corresponds to a non deterministic situation since it remains to decide whether tokens available in such a place are to be “consumed” by the left-hand **or** by the right-hand transition. Case (a) poses a kind of dual problem: tokens arriving in such a place may come from the right-hand **or** from the left-hand transition. In [5], it is shown that the resulting equations (analogous to those shown later on) involves a kind of “inf-convolution” over all past possibilities of arrivals of token at the place. This would yield *nonlinear* equations and an *infinite dimensional* state vector in the theory to be discussed hereafter.

To summarize, event graphs do not allow the modelling of “**or**”, and in particular of *competition* (case (b)). But they allow the modelling of “**and**” as reflected by “forks” and “joins” at transitions in Figure 1. For example, the join at transition y implies that tokens must be available in places $y|x_3$ **and** $y|x_2$ for transition y to be fired (this is a *synchronization* constraint). It is assumed that exactly one token is consumed in each upstream place when the transition is fired. Considering e.g. the fork at transition x_1 , tokens are created in places $x_2|x_1$ **and** $x_3|x_1$ whenever transition x_1 is fired (this may represent the splitting of a part into subparts in a workshop, or a message broadcast in a communication network for example). It is assumed that exactly one token is created in each downstream place by one firing of the transition.

An important property of event graphs is that *the number of tokens is constant in each circuit all along the life of the system*. In particular, if a circuit has no token in the initial marking (represented by dots in Figure 1), the transitions and places along this circuit should be removed since they will never be activated.

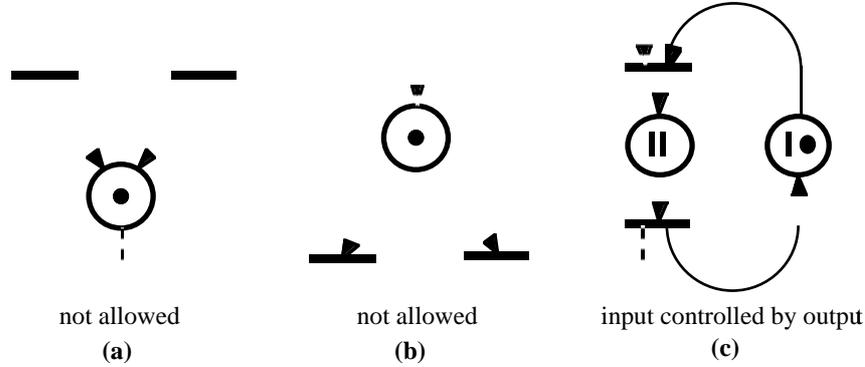


Figure 3: About event graphs

We are interested in *timed* event graphs. Without loss of generality, we may consider that transitions are immediate and that holding times (represented by bars in time units) apply only to places. As a matter of fact, if a transition has some duration, we may split it into two distinct transitions, the beginning and the end, separated by a place bearing the corresponding time (see Figure 3-(c)). Moreover, if it is meant that a new activation of the transition cannot occur before the end of the previous activation, it suffices to put a feedback arc with a place marked with one token. Moreover, this place can incorporate a holding time equal to some “reset time” if necessary.

The holding time at a place means that tokens must stay at least that time in the place before being available for downstream transition firing. This poses a problem concerning tokens of the initial marking: are they immediately available at the beginning of the game, or after which time will they be available⁹? For the time being we assume the following canonical initial informations¹⁰: *tokens of the initial marking are available immediately*. More general initial informations will be discussed later on. We assume that transitions are fired immediately when they can be fired (otherwise, the behaviour would not be deterministic).

Events of interest in a timed event graph are the transitions firings and the arrivals and departures of tokens at and from places. Events of a given type (firing of transition x , arrival at place $y|x$, etc ...) are numbered sequentially as they occur. To completely describe the behaviour of a timed event graph along time, it suffices to record the sequences of dates for all events of all types.

⁹In this latter case, the time after which they are available should be less than or equal to the holding time of the place.

¹⁰We say “initial informations” rather than “initial conditions” for reasons that will be more apparent later on.

Departure times of tokens from places are the same as firing times of downstream transitions. Also, arrival times of tokens at places are the same as firing times of upstream transitions (except for tokens in the initial marking for which initial informations should be given as already discussed). Hence, it suffices to record the sequences of transition firing dates only (and of initial informations if they are not canonical).

7.2 Daters in event-domain

From the above discussion, it is natural to introduce the following variables called “*daters*”. For a transition labelled x , we define x_n as the date at which transition x has been fired for the n^{th} time (indeed, n , as well as x_n , may be negative — see Remark 6 hereafter). We now show that these daters are related by linear equations in some dioid. Consider the example of Figure 1. We claim that

$$(x_1)_n = \max[(x_2)_{n-1}, (u_1)_n + 3]$$

As a matter of fact, for x_1 to be fired n times it is necessary that u_1 be also fired n times, but x_2 needs only be fired $n - 1$ times since there is one token already available in place $x_1|x_2$ at the beginning of the game. Moreover, if u_1 produces a token at some time t , this token will not be available before $t + 3$ for use by x_1 , whereas there is no such delay between x_2 and x_1 . The “max” reflects the behaviour of joins in Petri nets (synchronization). Finally, equality arises from the assumption that there is no extra delay for firing transitions whenever tokens are all available.

Applying the same simple rules (namely, one token in the initial marking induces a unit delay in index n , holding times are added to upstream daters, max is taken over all upstream transitions), we get, e.g. for x_3

$$(x_3)_n = \max[(x_1)_n, (x_2)_n + 1, (x_3)_{n-1} + 2] \quad (53)$$

We let the reader write down the equations for all other transitions of Figure 1.

Let us now consider that all variables $(u_i)_n, (x_i)_n, y_n \dots$ lie in the complete dioid $(\mathbb{Z} \cup \{-\infty\} \cup \{+\infty\}, \max, +)$, called \mathbb{Z}_{\max} , for short, in which $\oplus = \max, \otimes = +, \varepsilon = -\infty, e = 0$ and $+\infty$ is for completeness — observe that $\varepsilon \otimes (+\infty) = (-\infty) \otimes (+\infty) = \varepsilon = -\infty$. Then, (53) can be written as the following *linear* equation

$$(x_3)_n = (x_1)_n \oplus 1(x_2)_n \oplus 2(x_3)_{n-1} \quad (54)$$

which means that we have obtained a way of describing timed event graphs as *linear time-invariant dynamic systems* in discrete “time”. In fact, this is a state space representations in the *event domain*, rather than in the usual time domain, since the index n counts events instead of being the clock pulse.

7.3 Counters in time-domain

Actually all trajectories produced by such systems are monotone (nondecreasing) functions of n since we decided to number events in the order as they occur. Monotone functions are “essentially” invertible (there is a problem with horizontal pieces of the graph which requires some care) and we may consider the inverse functions. These are functions of time, denoted by t (hence the *time domain*), and values are interpreted as numbers of events (hence the name of “counters” for the corresponding variables).

Remark 6 *Actually, counters may be initialized at null or even negative values, and they are incremented by one at each occurrence. Therefore, when we speak of a “number n ” of events, we do not mean that n events occurred, but rather that the value of the corresponding counter is n . To avoid this confusion, we prefer to speak of an “event of number n ” and n may be negative. In the same way, the origin of time, and hence t , may be negative. \square*

We refer the interested reader to Caspi and Halbwachs [1] who introduced this duality of points of view between time and event domains very formally, and to whom we borrowed the terminology of daters and counters. Here we provide a direct introduction to counters and to the corresponding equations without reference to the previous section. A careful connection between counter and dater equations should be established, but it will not be discussed here since we are going to abandon both points of view and to introduce a new two-dimensional domain later on.

For a transition x , let x_t denotes the value of the counter at time t . Notice that, for convenience, we keep on calling the variable after the name of the transition although its physical meaning has changed (its is no longer a dater), but the new context is recalled by the index which is t now, instead of n . We claim that

$$(x_1)_t = \min[(x_2)_t + 1, (u_1)_{t-3}]$$

As a matter of fact, as already discussed, an event of some number at transition x_1 has to be related to an event of the same number at u_1 , but to an event of that number minus one at x_2 (due to the presence of one token in the initial marking of place $x_1|x_2$). Also, a token consumed at t by x_1 has to have been produced at least at time $t - 3$ by u_1 , since it is held at least three time unit at place $x_1|u_1$. Finally, the “min” reflects the fact that one firing of x_1 consumes one token from $x_1|u_1$ and from $x_1|x_2$ and that no more firing occurs as long as one of these upstream places remains empty.

The rules to get these counter equations directly from the graph or to derive them from the dater equations are apparent. Because max must be changed into min, the proper dioid to consider now is $(\mathbb{Z} \cup \{+\infty\} \cup \{-\infty\}, \min, +)$ called \mathbb{Z}_{\min} (notice that now $\varepsilon \otimes (-\infty) = (+\infty) \otimes (-\infty) = \varepsilon = +\infty$). Counter equations are linear for this dioid algebra: for example, the counterpart of (54) is

$$(x_3)_t = (x_1)_t \oplus (x_2)_{t-1} \oplus 1(x_3)_{t-2} \tag{55}$$

but it should be kept in mind that the meaning of variables and of the sign \oplus has changed.

Remark 7 *The passage from max to min is a nonlinear transformation as shown by the formula $\min(x, y) = -\max(-x, -y)$ which can be translated with \mathbb{Z}_{\max} notations into $\min(x, y) = (x^{-1} \oplus y^{-1})^{-1}$. The passage from daters to counters (inverse functions) is also a nonlinear transformation. But, there is an algebra suited to each domain so that equations remain linear. \square*

8 A 2-D domain and the $\text{MinMax}\llbracket\gamma, \delta\rrbracket$ algebra

8.1 Informations about events

We have seen that the type of events one need to consider in a timed event graph may be reduced to transition firings. We thus label a type of event by the name of the corresponding transition. For a transition x , we consider “pieces of informations” about events as pairs of integers $(n, t)_x$. What should be the interpretation of such a piece of information? As we have already seen, the graph carries pieces of information from upstream to downstream transitions. At a given transition, we “add” informations in the sense of simply taking the union of informations available *a priori* or coming from upstream.

But then we face the following problem. It may happen that, at a join, the information coming from left is, say, $b = (3, 3)$ whereas that coming from right is $c = (4, 2)$ (see Figure 4). If we interpret the pair (n, t) as “*event of number n occurred exactly at time t* ”, there will be a contradiction since b and c are not on the graph of a nondecreasing function. But the contradiction is resolved if we rather interpret (n, t) as “*at time t , the number of the last event that occurred is at most n* ”, or equivalently “*the event of number n occurred at the earliest at time t* ”. If we do so, then the piece of information b strictly dominates c in the sense that $b \Rightarrow c$ (think about it!), that is $b \cup c = b$. In fact b dominates all informations represented by points in the south-east cone of which b is the vertex in the 2-D domain of Figure 4. However, in the same figure, a and b does not dominate each other and $a \cup b$ (represented by the shaded area) is strictly stronger than b or a alone.

Clearly, we are introducing an idempotent addition (the union) over pieces of information, to which a partial order can be associated as in Theorem 1. For this addition to be an inner operation, we have to define the set of informations as the set of finite union of pieces of informations (union of cones), and for this semilattice to be complete, we have to extend the definition to infinite unions. The order is simply inclusion. There is a null element which is the information dominated by any other one, namely $\varepsilon = (+\infty, -\infty)$, which may be phrased “*the event of number ∞ occurred at the earliest at the beginning of times*”, which indeed says nothing about the behaviour of the system.

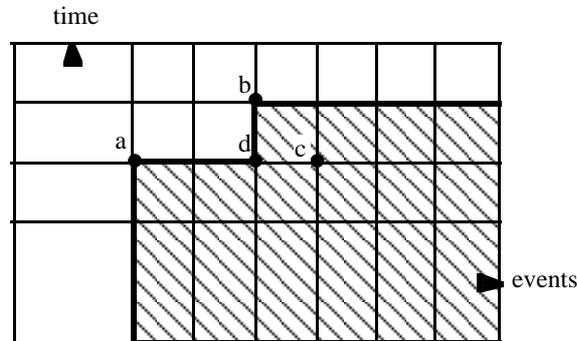


Figure 4: Cones of informations

Since we have a complete sup semilattice with a bottom element ε , we can get a lattice by the construction recalled in Section 2.2. Applying this construction, it is realized that the operation \wedge graphically amounts to the intersection of corresponding areas. For example, $a \cap b = d$ in Figure 4. It is then obvious that this lattice is distributive.

Remark 8 *By considering that a piece of information is equivalent to the collection of informations lying in the south-east cone of which that piece is the vertex, we have been able to resolve potential contradictions that would otherwise have arisen when having to consider $b \cup c$ for example. But one may argue that this is not the only way to avoid these contradictions and that setting $b \cup c = c$ rather than b would have done the same job. This corresponds to saying that a piece of information dominates all informations in the north-west, rather than south-east cone. Otherwise stated, this amounts to turning expressions as at most (resp. at the earliest) into at least (resp. at the latest) in sentences above. Indeed, this other way of performing “addition” of informations will prove to be useful in another context later on. The sum we have chosen is in fact compatible with the way join transitions gather informations coming from upstream in Petri nets (it is worth taking a while thinking of that).* \square

8.2 Shift operators

As it was apparent from earlier sections, the graph carries informations from upstream to downstream transitions. More precisely, in Figure 1, informations $(n, t)_{x_1}$ can be related to informations $(n', t')_{u_1}$ if the latter are shifted backwards by three units in time, and to informations $(n'', t'')_{x_2}$ if these are shifted backwards by one unit in the number of events.

Therefore, we associate a backward (elementary) operator $(m, d)_{y|x}$ to every arc or place $y|x$, where m takes the value of the number of tokens in the initial marking and d takes the value of the holding time. Then a piece of information $(n, t)_x$ yields a piece of information $(n', t')_y = (m, d)_{y|x} \cdot (n, t)_x := (n + m, t + d)$. The definition is extended in a straightforward manner to “sums” of pieces of informations by distributivity. It is also immediate to define the “sum” of two shift operators $f_{y|x}$ and $g_{y|x}$ in a usual way by summing up the informations produced by the two operators from the same information i_x . This corresponds to two parallel arcs between x and y in the graph. General (nonelementary) shift operators are defined as (finite or infinite) sums of elementary operators. Two arcs in series $y|x$ and $z|y$ provide a way of defining the “product” of operators as the usual composition of applications. Obviously, $(m', d')_{z|y} \circ (m, d)_{y|x} = (m + m', d + d')$. There is an identity operator represented by $(0, 0)$. The product of operators distributes over sum.

It is not hard to realize that we are just about constructing a dioid structure for shift operators. But it is possible to extend this structure to informations (observe that we already defined the sum of informations which behaves arithmetically as the sum of operators). In fact informations and operators need not be distinguished. As a matter of fact, we may conceptually consider that every information, say $(n, t)_x$, is produced from a canonical information $(0, 0)$ attached to a “source” transition s from which we draw an arc $x|s$ with weight (n, t) (if n and t are counted from negative origins, these origins are put at the source instead of $(0, 0)$). However this discussion is better pursued with more convenient algebraic notations and in a more abstract setting.

8.3 Coding informations and shift operators

A collection of informations $\{(n_i, t_i) \mid i \in I \subset \mathbb{N}\}$ about a type of event is a collection of points in the plane depicted on Figure 4. It may be seen as a pointwise measure in the plane (a measure in \mathbb{Z}^2). We may represent this “discrete event measure” by its characteristic function $\sum_{i \in I} \gamma^{n_i} \delta^{t_i}$ where the sum is a formal sum and γ and δ are formal variables. In fact, γ may be interpreted as the *backward shift operator on counting* and δ as the *backward shift operator on dating*. For example, by considering arcs such as $x_1|x_2$ (resp. $x_2|x_1$) in Figure 1, it can be shown that downstream discrete event measures can be derived from upstream ones by formal multiplication of the corresponding formal power series by γ (resp. δ).

Formal addition of power series corresponds to our previous sum or union of informations about a type of event. But we must reflect the fact that $a \cup d$ (resp. $b \cup d$) in Figure 4 is equal to a (resp. b). This imposes, in addition to the usual rules of formal sum and product of power series, the unusual additional rules

$$\gamma^n \delta^t \oplus \gamma^{n'} \delta^t = \gamma^{\min(n, n')} \delta^t \quad (56)$$

$$\gamma^n \delta^t \oplus \gamma^n \delta^{t'} = \gamma^n \delta^{\max(t,t')} \quad (57)$$

Notice that this is enough to get that $b \cup c = b$ in Figure 4 since $b \cup c = (b \cup d) \cup (d \cup c) = b \cup d = b$.

We are now ready to introduce a new dioid more formally in order to provide a theoretical basis to all these calculations.

8.4 The dioid $\text{MinMax}\langle\langle\gamma, \delta\rangle\rangle$

We first introduce the set of formal power series $\mathcal{B}\langle\langle\gamma, \delta\rangle\rangle$ in two variables (γ, δ) with *boolean* coefficients and exponents in \mathbb{Z} . This set will be designated by \mathcal{L} in the sequel. Observe that this is a dioid. In particular the sum is idempotent since coefficients are boolean, ε corresponds to the series with coefficients all equal to zero and $e = \gamma^0 \delta^0$. Graphically, an element of \mathcal{L} is represented by a collection of points in \mathbb{Z}^2 : a monomial $\gamma^n \delta^t$ with boolean coefficient 1 yields a point with coordinates (n, t) ; null boolean coefficients yield no point (hence ε corresponds to the empty set and e to the origin of the plane). Addition corresponds to union and multiplication of two elements to the vector sum of the associated collections of points. Order is inclusion. \wedge is intersection.

As previously observed, this dioid is not adequate for our purpose since (56)–(57) are not part of the usual calculation rules with formal power series. In fact, we wish that a point $a = (n, t)$ in the 2-D domain, coded by the monomial $\gamma^n \delta^t$, be “equal” to the whole south-east cone of vertex a . But this cone is coded by $\gamma^n \delta^t \gamma^* (\delta^{-1})^*$ (recall the definition (32) of the *star*). Observe that $\gamma^* (\delta^{-1})^* = (\gamma \oplus \delta^{-1})^*$ from (49). We are thus led to consider the dioid \mathcal{L} modulo $\gamma \oplus \delta^{-1}$ based on Definition 9 and Theorem 20.

Definition 10 *The dioid $\text{MinMax}\langle\langle\gamma, \delta\rangle\rangle$, designated hereafter by \mathcal{M} for short, is the dioid $\mathcal{L}/(\gamma \oplus \delta^{-1})$.*

Comments

- (i) This dioid is commutative, complete and distributive. All these properties can be checked easily. The graphical interpretations of \oplus , \wedge and \otimes as \cup , \cap and vector sum respectively are still valid, but they now apply to unions of south-east cones.
- (ii) Starting from the formal definition of \mathcal{M} , it is easy to check that all calculations can be performed practically using the usual the rules of formal power series and the additional rules (56)–(57). It should be kept in mind that an element of \mathcal{M} is an equivalence class. Therefore it has several formal representations. For example, $e = \gamma^* (\delta^{-1})^* = \gamma^* = (\delta^{-1})^* = \gamma^* \oplus (\delta^{-1})^* = \gamma^0 = \delta^0 = \gamma^0 \delta^0$ in \mathcal{M} but not in \mathcal{L} ! Indeed, in each equivalence class, there is a minimal and a maximal representatives, “minimal” and “maximal” referring to the order in \mathcal{L} . The minimal representative is obtained by coding only the extremal north-west points of the shaded

area in Figure 4. The maximal one is obtained by coding all the points in the shaded area. This maximal representative is obtained algebraically by multiplication of any representative by $\gamma^*(\delta^{-1})^*$ (see last statement of Theorem 20).

- (iii) Let us consider how to calculate $a \wedge b$ in \mathcal{M} . First of all, since the dioid is distributive, it suffices to know how to calculate that expression when a and b are monomials. Then it can be checked that

$$\gamma^n \delta^t \wedge \gamma^{n'} \delta^{t'} = \gamma^{\max(n, n')} \delta^{\min(t, t')} \quad (58)$$

which should be compared with (56)–(57). Notice that \otimes does not distribute over \wedge in general, although this is true for multiplication by monomials.

- (iv) In a general *complete* dioid, let us introduce the following notation (compare with (32))

$$a^\wedge := e \wedge a \wedge a^2 \wedge \dots \quad (59)$$

Then, coming back to \mathcal{M} , consider the following table

$$\begin{aligned} \gamma^{+\infty} &\sim \gamma^\wedge \\ \gamma^{-\infty} &\sim (\gamma^{-1})^* \\ \delta^{+\infty} &\sim \delta^* \\ \delta^{-\infty} &\sim (\delta^{-1})^\wedge \\ \gamma^{+\infty} \delta^{-\infty} &\sim \varepsilon \\ \gamma^{-\infty} \delta^{+\infty} &\sim \infty \end{aligned}$$

It is not difficult to realize that elements in the left-hand side formally behave as their corresponding elements in the right-hand side, the latter being well-defined in \mathcal{M} . Therefore, we may take the right-hand expressions as definitions for the left-hand notations. The only additional rules to keep in mind are the following

$$\gamma^{-\infty} \gamma^{+\infty} = \gamma^{+\infty} \quad (60)$$

$$\delta^{-\infty} \delta^{+\infty} = \delta^{-\infty} \quad (61)$$

These rules can easily be proved from the definitions.

- (v) \mathcal{M} is not Archimedean. A counterexample is the following

$$\gamma^n \delta^{-\infty} \otimes \infty = \gamma^{-\infty} \delta^{-\infty} \neq \infty$$

which contradicts Theorem 2. □

9 Input-output representation in the 2-D domain

9.1 System equations and transfer matrices

Consider an event graph as in Figure 1. To each transition, we associate an element of \mathcal{M} which codes the whole set of informations available about the (generally infinite) sequence of events related to this transition. This element will be called after the name of the transition (*e.g.* x_1) but it should be kept in mind that it is different from a dater or a counter, and that it is a power series in two variables with positive (and possibly negative) exponents (indeed an equivalence class). If necessary, to recall this, we may sometimes say $x_1(\gamma, \delta)$ instead of x_1 .

From all that have been said until now, it should be clear that the following equations can be obtained for the case of Figure 1

$$\left. \begin{aligned} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} &= \begin{pmatrix} \varepsilon & \gamma & \varepsilon \\ \delta & \varepsilon & \varepsilon \\ e & \delta & \gamma\delta^2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \oplus \begin{pmatrix} \delta^3 & \varepsilon \\ \varepsilon & \gamma\delta \\ \varepsilon & \varepsilon \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \\ y &= \begin{pmatrix} \varepsilon & \gamma & \delta^3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \end{aligned} \right\} \quad (62)$$

In general, for any event graph, we get the following equations

$$x = Ax \oplus Bu \quad y = Cx \quad (63)$$

where x, y (resp. A, B, C) are column vectors (resp. matrices) of appropriate dimensions with entries in \mathcal{M} .

The former equation is an implicit equation of the form (31). We know that it has no unique solution in general. However, we are interested in the smallest solution since this corresponds to the *earliest possible dates* for events (recall that we assumed that transitions are fired as soon as they can). Using Theorem 9-(i), we conclude that $x = A^*Bu$ and

$$y = CA^*Bu \quad (64)$$

This completely specifies the input-output behaviour of the system: for given informations about the input transitions (they must be activated from outside), informations about the earliest possible outputs are calculated. The matrix $\mathcal{H} := CA^*B$ is called the “*transfer matrix*” of the system. Its expression has been given in (1) for the case of Figure 1.

In general, \mathcal{H}_{ij} corresponds to the “trajectory” of y_i for all inputs $u_k = \varepsilon$ except for $u_j = e$. In conventional system theory, it is called the “impulse

response” from j to i . Here, an input $\varepsilon = \gamma^{+\infty}\delta^{-\infty}$ can be interpreted as an infinity of tokens available in an imaginary place upstream the corresponding input transition since the beginning of times. This is the least constraining input one can imagine. On the other hand, an input $e = \gamma^*$ may be called an impulse at time 0. Before time 0, the input transition is frozen, and then an infinity of tokens are made available. Notice that tokens (or transition firings) are numbered from 0, not from 1, since $e = \gamma^0\delta^0 \oplus \gamma^1\delta^0 \oplus \dots$

9.2 About initial informations

In Section 7.1, we raised the problem of tokens of the initial marking on which we defined canonical informations as “*these tokens are immediately available*”. Notice that this corresponds to the “best” initial “conditions” in that they give the earliest possible outputs (that is the smallest ones in the \mathcal{M} order). Consider the very simple graph of Figure 5-(a): the transfer function is $y = \gamma^2\delta^2u$. For an impulse input $u = e$ (an infinity of tokens available at the input transition at time 0), the output is $\gamma^2\delta^2 \oplus \gamma^3\delta^2 \oplus \dots$ which means that “*the third and next tokens are out at time 2*” (recall that the third token is labelled 2, the first being labelled 0). Nothing is said about the first two tokens, those which belong to the initial marking. More exactly, at the earliest, we may suppose that these two tokens were out since $-\infty$. Indeed, there is no explanation why these two tokens did not leave the place at any time before 0. Therefore, our theory appears to be coherent. But suppose that one wants to mean that one of the tokens of the

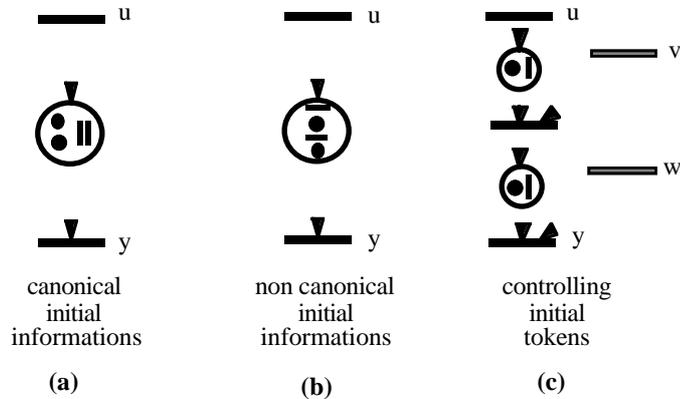


Figure 5: Canonical and non canonical initial informations

initial marking (the lower in Figure 5-(b)) is available only at time 0, whereas the other one (the upper) is available only at time 1 (that is, it spent only one

time unit in the place at the beginning of the game). This may be represented as shown in (b). To come back to canonical initial informations, we can draw the expanded graph of Figure 5-(c) where it is necessary to add the grey input transitions v and w . Without these controls, the two tokens would again be able to leave the system at any time before 0 (the transfer function would again be $\gamma^2\delta^2$). Now it is $y = w \oplus \gamma\delta(v \oplus \gamma\delta u)$. To release the tokens only at time 0, we set $v = w = e$ and we get $y = e \oplus \gamma\delta \oplus \gamma^2\delta^2 u$. The input-output relationship is the same as in the previous case but we have the additional informations that the first token leaves at time 0 and the second at time 1 (this is more informations than previously and indeed this new y is larger than the previous one in the \mathcal{M} order). We invite the reader to consider the same kind of non canonical initial informations but for tokens in a loop. Notice that we permanently used the expression “initial informations” rather than “initial conditions” because variables are attached to transitions whereas these informations are attached to tokens of the initial marking.

9.3 Circulation of tokens

Consider *e.g.* the event graph of Figure 1. Suppose that we bring only one token at the input transition u_1 at time 0. The corresponding inputs are $u_1 = e \oplus \gamma\delta^{+\infty}$ and $u_2 = \delta^{+\infty}$ and the output calculated with (1) is $y = \delta^8 \oplus \gamma\delta^{+\infty}$, indicating that a single token has been released at time 8. Then the system is stabilized, but with a new marking: since all transitions, but u_2 , have been activated once, all places have the same marking as initially except for $x_2|u_2$ where the token has disappeared. If we calculate the new transfer matrix for this new “initial marking”, we find that \mathcal{H}_1 is unchanged, but that \mathcal{H}_2 is divided by γ with respect to the previous one appearing in (1). The explanation is that changing the origins of dates and of counters has no effect on the transfer matrix provided that the same change be applied to all input and output transitions. This is what occurred in our example for the pair (u_1, y) since both have been fired once. But the rule is not fulfilled for (u_2, y) , hence the change in \mathcal{H}_2 .

Consider again the experiment of bringing one token at u_1 at time 0 from the initial situation of Figure 1. At time 2, the token is still in place $x_1|u_1$ since it must stay 3 time units there, and no other transition has been fired so far. From what has been said above, it can be predicted that if we write down the equations for this new situation, considering canonical initial informations (which anyway do not affect the transfer matrix), we shall find the same \mathcal{H}_2 , but a \mathcal{H}_1 which is multiplied by γ . Let us consider this new transfer matrix, say $\bar{\mathcal{H}}$, as the reference for what follows. Just after time 3, transition x_1 will have been fired, taking off one token from places $x_1|u_1$ and $x_1|x_2$ and adding one token in places $x_3|x_1$ and $x_2|x_1$. This new transition firing which has affected only an inner transition should not modify $\bar{\mathcal{H}}$. This can be checked by the reader. Of course, this will not be the case if transitions are fired by adding tokens in upstream places directly from outside, that is without passing through input

transitions. This may cause a dramatic change of the transfer matrix, especially if it affects the number of tokens in each circuit which is invariant for any event graph.

The fact that regular firings of inner transitions which move tokens around do not affect the transfer matrix is reminiscent of the well-known fact of conventional system theory that a linear change of basis in the state space does not modify the input-output relationship. Although we have not considered transfer matrices (in a single variable) corresponding to either dater or counter “state space” equations yet (this will come later on), we invite the reader to check whether those transfer matrices are not affected by inner circulations of tokens, and possibly to find out the corresponding state space basis change. It turns out that the answer is yes only with the counter approach. Notice that tokens of the initial marking cause delays in dater equations whereas they appear as coefficients in counter equations. Coming back to our 2-D domain, we let the reader check that, if transition x_i is regularly fired once, the new matrices, say $\bar{A}, \bar{B}, \bar{C}$, obtained by writing down equations after this firing are related to the previous A, B, C by the relations

$$\bar{A} = M^{-1}AM \quad \bar{B} = M^{-1}B \quad \bar{C} = CM$$

where M is a diagonal matrix with $M_{jj} = e$ for $j \neq i$ and $M_{ii} = \gamma$. As a matter of fact, firing x_i produces one token more (multiplication by γ) — resp. one token less (multiplication by γ^{-1}) — in all downstream — resp. upstream — transitions (indicated by the nonnull entries of column A_i — resp. of row A_i). The new matrices contain no entries with negative exponents iff the transition firing is regular.

10 Output targets and the adjoint system

10.1 The problem of output targets as a residuation problem in \mathcal{M}

Instead of specifying input informations and calculating the earliest (smallest) resulting output, we may rather impose output targets and search for the latest (greatest) input that meets these specifications. More precisely, for a specified target z (a column vector the dimensionality of which is the number of output transitions, say p), we pose the following problem

$$\text{Find the greatest } v \text{ such that } CA^*Bv \leq z \quad (65)$$

where v has the dimensionality of the number of input transitions. This is reminiscent of Definition 7 of residuation and one has that $v = (CA^*B) \setminus z$.

There is however a point which deserves some attention. An elementary target specified by a pair (n, t) should be interpreted as “*the event of number n* ”

must occur **at the latest** at time t ” or “the number of the last event occurred up to time t must be **at least** n ”. A little thinking should make clear that (n, t) dominates (is stronger than) all other targets (n', t') such that $n' \leq n$ and $t' \geq t$, that is targets in the north-west cone of (n, t) . Considering several (or an infinity) of such elementary targets to be met all together amounts to “adding”, in the sense of taking the union of, the corresponding north-west cones. These unions of cones are elements of a dual dioid called $\mathcal{N} = \text{MaxMin} \llbracket \gamma, \delta \rrbracket := \mathcal{L}/(\gamma^{-1} \oplus \delta)$. The practical rules for calculating in \mathcal{N} (apart from the usual rules of power series) are obtained by inverting “max” and “min” in (56)–(57). However, z appearing in (65) must have entries that belong to \mathcal{M} and not to the dual dioid \mathcal{N} .

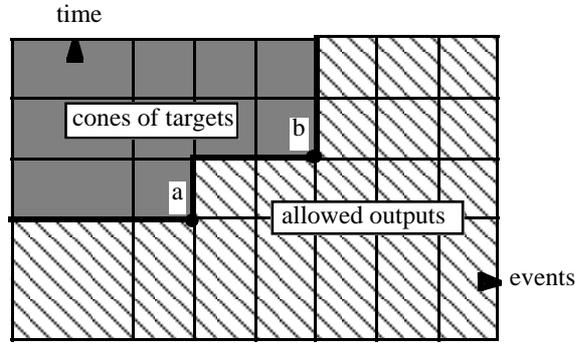


Figure 6: Cones of targets

In fact, a union of target cones (for example those of vertices a and b of Figure 6) defines an area of \mathbb{Z}^2 into which allowed output “trajectories” (represented by elements belonging to \mathcal{M}) must not penetrate. That is to say, we must code the shaded south-east area of Figure 6 to get the entries of $z \in \mathcal{M}^p$, and then we must require that an admissible trajectory be not greater than z (the border is allowed). The practical rule is the following: given an elementary target (n, t) for output transition j , we set $z_j = \gamma^{-\infty} \delta^t \oplus \gamma^n \delta^{+\infty} \in \mathcal{M}$ as the j^{th} coordinate of the right-hand side z of (65), and given a collection of targets $\{(n_i, t_i) \mid i \in I_j\}$, we set

$$z_j = \bigwedge_{i \in I_j} (\gamma^{-\infty} \delta^{t_i} \oplus \gamma^{n_i} \delta^{+\infty}) \quad (66)$$

Recall that residuation distributes over \wedge (see (22)).

10.2 Solving the target problem with the adjoint system

Notice that since the rules for “adding” a collection of elementary targets $\{(n_i, t_i) \mid i \in I_j\}$ (represented by an element of \mathcal{N}) are dual of those for adding a collection of pieces of informations, it would suffice to code this collection of targets as

$$\eta_j = \bigoplus_{i \in I_j} \gamma^{-n_i} \delta^{-t_i} \quad (67)$$

to come back into \mathcal{M} . The minus signs in the exponents have the desired effect of inverting “max” and “min” in calculations and this correspondence is obviously an isomorphism between dioids \mathcal{N} and \mathcal{M} . Then the question arises of the relationship of z given by (66) with η given by (67) which are two different ways of coding targets in \mathcal{M} .

Lemma 5 *Let $\varphi := \gamma^{-\infty} \delta^0 \oplus \gamma^0 \delta^{+\infty} \in \mathcal{M}$. Then the p -dimensional column vector with entries z_j given by (66) is related to the row vector with entries η_j given by (67) by $z_j = \tilde{\eta}_j = \eta_j \setminus \varphi$.*

The proof is straightforward using formula (21) and the fact that residuation by monomials (say $\gamma^{-n} \delta^{-t}$) amounts to multiplication by the inverse $\gamma^n \delta^t$. This is an example where residuation is easy to perform.

As a consequence of this lemma and the discussion around formula (29) in Section 3.3, the solution v of problem (65) can be obtained as follows (we use “tilde” on vectors with an obvious meaning — recall that column vectors yield row vectors by this transformation and conversely)

$$v = (CA^*B) \setminus z = (CA^*B) \setminus \tilde{\eta} = \tilde{\mu} \quad (68)$$

with

$$\mu = \eta CA^*B \implies \mu = \xi B \quad \xi = \xi A \oplus \eta C \quad (69)$$

which is the adjoint system of (64)–(63).

Remark 9 *Notice that, in (69), the implicit system in the right-hand side must be solved for the smallest solution if we want to replace \implies by \iff . This is coherent with the fact that “smallest” in the coding (67) of targets is equivalent to “greatest” (latest) in the coding (66), that is after the transformation “tilde”. \square*

10.3 Direct derivation of the adjoint system

Let us show how we could have obtained the solution of problem (65) through formulas (68)–(69) by reasoning directly on the graph as we did to obtain equations for the direct system. Consider the example of Figure 1 and the output transition y for which a collection of targets $\{(n_i, t_i) \mid i \in I_y\}$ are given. It is clear that these targets can be transferred back to the upstream transitions x_3

and x_2 by subtracting 3 units from t_i , respectively 1 unit from n_i . That is arcs with weights (shift operators) (m, d) operate back with $(-m, -d)$. Notice that joins in the direct graph are forks in the reverse graph. In the same way, forks become joins as for example at transition x_2 . Targets carried back from the downstream transitions y, x_3, x_1 must be “added” using rules which are completely dual of those for adding informations as already discussed. Finally, we see that we must operate with a dual algebra and with shift operators which have negative signs. Instead of that, we may decide that we are going to change the signs of all equations so that we come back to the previous algebra, shift operators get back their positive signs, but a collection of targets is now recorded as $\{(-n_i, -t_i)\}$. This corresponds exactly to the coding (67) and to equations (69), where the original algebra \mathcal{M} and the original operators A, B, C (not their “opposite”) are used, but where vectors are row vectors operating *on the left* of operators (because we must anyway scan the graph from output to input transitions).

11 Rationality, realizability and periodicity

11.1 A fundamental theorem

A timed event graph as that of Figure 1 can be represented as a finite-dimensional linear invariant system (63) which in turn can be represented by a transfer matrix (64).

Remark 10 *If there are direct arcs from input transitions to output transitions, we get an output equation of the form $y = Cx \oplus Du$ and a transfer matrix $CA^*B \oplus D$. However, by redefining the “state” vector as $\bar{x} = (x^T, u^T)^T$, it is possible to come back to the form $\bar{C}\bar{A}^*\bar{B}$ but at the price of increasing the “state” dimensionality. This has no importance in what follows since we are not going to deal with the problem of “minimal” realizations. \square*

Observe that for a realistic (“causal”) event graph, entries of A, B, C are *polynomials* of \mathcal{M} involving only *nonnegative* exponents. These polynomials belong to the “dioid closure” of the “generating subset” $\mathcal{E} = \{\varepsilon, e, \gamma, \delta\}$. This motivates the following definition of rationality.

Definition 11 (Rationality) *A matrix or a vector with entries in \mathcal{M} is said rational iff its entries lie in the subdioid $\mathcal{E}^* \subset \mathcal{M}$ (rational closure of $\mathcal{E} := \{\varepsilon, e, \gamma, \delta\}$ — see Definition 8).*

Definition 12 (Realizability) *A $p \times m$ -dimensional matrix \mathcal{H} in \mathcal{M} is said realizable iff it can be written*

$$\mathcal{H} = C(\gamma A_1 \oplus \delta A_2)^* B \tag{70}$$

where C and B are matrices of respective dimensions $p \times n$ and $n \times m$, n being an arbitrary but finite integer, A_1, A_2 are $n \times n$, and the entries of all matrices lie in $\{\varepsilon, e\}$.

Definition 13 (Periodicity) *An element h of \mathcal{M} is said periodic iff there exist nonnegative and finite integers ν, τ, r, s (r and s not both null) such that h can be written as*

$$h = P_{\nu-1, \tau-1} \oplus [\gamma^\nu \delta^\tau \otimes Q_{r-1, s-1} \otimes (\gamma^r \delta^s)^*] \quad (71)$$

where $Q_{r-1, s-1}$ stands for a polynomial in (γ, δ) with nonnegative exponents of degrees not greater than $(r-1, s-1)$, and $P_{\nu-1, \tau-1}$ has a similar meaning. A matrix \mathcal{H} is periodic iff all its entries are periodic.

The intuitive meaning of formula (71) considered as an impulse response is the following. There is a transient behaviour represented by P , followed by a periodic behaviour in which a pattern represented by Q is indefinitely reproduced around an asymptotic slope s/r . This ratio represents the asymptotic output rate of the response: in the average, r tokens get out each s time units.

If $s = 0$, the impulse response is polynomial since $(\gamma^p)^* = e$: this is the behaviour of a purely “moving average” system (no circuit in the graph). Notice that, here, thinking of a “polynomial impulse response” as a “finite impulse response”, as it is familiar in conventional system theory, may be misleading: for an “impulse” at the input, that is an infinity of tokens brought at time 0, there is a transient (some tokens get out as indicated by the polynomial), and then an infinity of tokens get out all together.

If $r = 0$, since $(\delta^s)^* = \delta^{+\infty}$, the response is “frozen” after a possible transient during which some tokens get out (this may rather be the kind of behaviour one would like to call a “finite response”). This happens when there is a circuit without token which is blocking the output transition.

Theorem 21 *For $\mathcal{H} \in \mathcal{M}^{p \times m}$, the following three statements are equivalent*

- (i) \mathcal{H} is realizable
- (ii) \mathcal{H} is rational
- (iii) \mathcal{H} is periodic

Proof The implication (i) \Rightarrow (ii) is straightforward. The converse (ii) \Rightarrow (i) follows, at least for the single-input single-output (SISO) case, from Theorem 14 with $\mathcal{B} = \mathcal{C} = \{\varepsilon, e\}$. It remains to split up matrix A_x which appears in the $(\mathcal{B}, \mathcal{C})$ -representation (45) into $\gamma A_1 \oplus \delta A_2$ which offers no difficulty. For the multiple-input multiple-output (MIMO) case, it suffices to realize the SISO transfer for each input-output pair individually, and then to concatenate these realizations, since we do not worry about minimality.

It remains to prove the equivalence (ii) \Leftrightarrow (iii). Since the definitions of periodicity and rationality refer to the entries of \mathcal{H} individually, it suffices to

deal with the SISO case. The implication (iii) \Rightarrow (ii) is obvious: if h can be written as (71), then clearly $h \in \mathcal{E}^*$. Conversely, if h is rational, since \mathcal{M} is a commutative dioid, we can use Theorem 19 (applied to the dioid closure of \mathcal{E}) to see that h can be written as

$$\begin{aligned} h &= \bigoplus_{i \in I} \gamma^{\alpha_i} \delta^{\beta_i} \left(\bigoplus_{j \in J_i} \gamma^{r_j} \delta^{s_j} \right)^* \\ &= \bigoplus_{i \in I} \gamma^{\alpha_i} \delta^{\beta_i} \bigotimes_{j \in J_i} (\gamma^{r_j} \delta^{s_j})^* \end{aligned} \quad (72)$$

where I and the J_i 's are finite sets, $\alpha_i, \beta_i, r_j, s_j$ are nonnegative integers and (72) follows from (49). The proof will be complete if we show that (72) can be reduced to (71): roughly speaking, this amounts to saying that some $(\gamma^{r_j} \delta^{s_j})^*$, essentially the one for which the ‘‘slope’’ s_j/r_j is the greatest, asymptotically ‘‘absorbs’’ all other terms of the same type to which it is added or which it multiplies. This is what is explained in the next technical lemma. \square

Lemma 6 *In \mathcal{M} , expressions of the form (72) can always be reduced to expressions of the form (71).*

Outline of proof This lemma can be proved in three steps. First, for given positive integers r, r', s, s'

$$[s/r > s'/r'] \implies [\forall \alpha, \alpha', \beta, \beta', \exists N : \forall n' \geq N, \exists n : \gamma^\alpha \delta^\beta (\gamma^r \delta^s)^n \geq \gamma^{\alpha'} \delta^{\beta'} (\gamma^{r'} \delta^{s'})^{n'} \text{ in } \mathcal{M}]$$

From this result, it is then proved that expressions of the form $\gamma^\alpha \delta^\beta (\gamma^r \delta^s)^* \oplus \gamma^{\alpha'} \delta^{\beta'} (\gamma^{r'} \delta^{s'})^*$ or $(\gamma^r \delta^s)^* (\gamma^{r'} \delta^{s'})^*$ can be reduced to expressions of the form $P(\gamma, \delta) \oplus Q(\gamma, \delta) (\gamma^r \delta^s)^*$, where $P(\gamma, \delta)$ and $Q(\gamma, \delta)$ are polynomials in (γ, δ) . Essentially the same thing can be proved when $s/r = s'/r'$ except that now the reduced expression is $P \oplus Q (\gamma^{r''} \delta^{s''})^*$ where r'' (resp. s'') is the least common multiple of r and r' (resp. s and s'). Finally, it remains to check that those expressions can be reduced to expressions like (71) for which additional conditions on the degrees of P and Q and on the valuation of Q must be met. \square

Comments

- (i) In the proof of Theorem 21, instead of proving the implication (ii) \Rightarrow (iii), we might have proved the implication (i) \Rightarrow (iii) using Theorem 18 and formula (51). This gives additional insight into how the unique term $(\gamma^{r_{ij}} \delta^{s_{ij}})^*$ appearing in the representation (71) of each \mathcal{H}_{ij} is related to the weights of circuits of matrix $A = \gamma A_1 \oplus \delta A_2$ appearing in (70). For an input-output pair (u_j, y_i) , we consider all (oriented) paths connecting these transitions and all circuits which have at least one node (i.e. transition) in common with those paths. These are the circuits of interest to determine the maximal ratio s_{ij}/r_{ij} . If there is no such circuit, this means that the

polynomial Q of (71) is null and $(\gamma^{r_{ij}} \delta^{s_{ij}})^*$ is irrelevant. As a consequence, if matrix A is strongly connected (which in particular precludes any direct path from the input to the output, given that we have not included a term Du — see Remark 10), then there is a unique $(\gamma^r \delta^s)^*$ for the whole transfer matrix \mathcal{H} .

- (ii) In the definition of realizability, we might have required the weaker condition that $\mathcal{H} = CA^*B$ with entries of A, B, C be simply *polynomials* with nonnegative exponents (this would be called a polynomial representation). The “inner” dimension would then be smaller than above in general for a given \mathcal{H} . This corresponds to choosing \mathcal{E} as the dioid closure of $\{\varepsilon, e, \gamma, \delta\}$ and to taking \mathcal{B} and \mathcal{C} identical to \mathcal{E} . In fact, there is a lot of other intermediate possibilities (see the comments at the end of Section 4.2). \square

11.2 The eigenvalue problem in the 2-D domain

We have seen that the impulse response of a finite-dimensional linear invariant system typically consists of a transient followed by a periodic behaviour which is characteristic of the system. Usually, for an autonomous linear system, characteristic behaviours are associated with eigenvalues, and transients are avoided by starting with initial conditions corresponding to eigenvectors.

Consider the autonomous system $x_{n+1} = Ax_n$ in the event domain (recall that the algebra is \mathbb{Z}_{\max}). The eigenvalue problem $Ax = \lambda x$ has been extensively studied in [2, 3]. Essentially, for a matrix A representing a strongly connected graph, there exists a unique eigenvalue λ which is equal to the maximal ratio s/r over all circuits of the graph (s is again the number of time units and r is the number of tokens of the initial marking along a circuit). For non strongly connected graphs, say for a graph with two strongly connected components, it is easy to figure out that there are two eigenvalues if the critical circuit belongs to the upstream component (the two eigenvalues are those of the two components considered separately) whereas there is only one eigenvalue (the largest one) if the critical circuit is downstream. The structure of the “eigenspace” has also been extensively studied.

An eigenvalue $\lambda = s/r$ means that, in the average, each component of x_n (the dater) is incremented by s time units every time n (the counter) is increased by r . In the 2-D domain, we have to say something similar. Recall that now x represents the whole trajectory. If a trajectory extending from $-\infty$ to $+\infty$ can “slide” along itself by translations (r, s) or $(-r, -s)$ without changing, it is invariant when multiplied by any positive or negative power of the monomial $\gamma^r \delta^s$. An elegant way to say this is to appeal once again to the “calculus modulo” introduced in Section 6. More precisely, we have to check that $[x]_{\gamma^r \delta^s \oplus \gamma^{-r} \delta^{-s}}$ is polynomial, each polynomial entry representing a pattern which is repeated indefinitely. This justifies the following definition.

Definition 14 *We say that a pair of positive integers (r, s) is an eigenvalue of*

the square matrix A with rational — see Definition 11 — and polynomial entries iff there exists a vector x with the same kind of entries, which is not identically null and such that

$$x = Ax \text{ in } \mathcal{M}/(\gamma^r \delta^s \oplus \gamma^{-r} \delta^{-s}) \quad (73)$$

When calculating in $\mathcal{M}/(\gamma^r \delta^s \oplus \gamma^{-r} \delta^{-s})$, in addition to the rules (56)–(57), one is allowed to replace every positive or negative power of $\gamma^r \delta^s$ by e since they all fall in the same equivalence class (thanks to (36)–(37)). Also, because of the commutative diagram of Figure 2, all these simplification rules can be used in any order.

Theorem 22 *Let A be a square matrix with rational and polynomial entries and assume that A is represented by a strongly connected graph. Then, there exists an eigenvalue (r, s) , and for this eigenvalue the ratio s/r is maximal over all circuits of the graph (in this sense, the eigenvalue is unique).*

Outline of proof This statement is a direct consequence of formula (40). This formula gives the general form of an x verifying (39). To prove existence of an eigenvector, we set $y_\kappa = 0$ for all circuits except that or those achieving the maximal ratio s/r for which we set, say, $y_\kappa = e$. We then use Lemma 6 and we calculate modulo $\gamma^r \delta^s \oplus \gamma^{-r} \delta^{-s}$ if the critical circuit is unique and has weight (r, s) , or we appeal to the least common multiples of all critical pairs (r, s) otherwise (see the explanations following Lemma 6). The resulting x is polynomial and rational and it verifies (73).

Conversely, suppose that x verifies (73) for some pair (r, s) . This means that $\vec{x} := (\gamma^r \delta^s)^\# x$ verifies (39) and thus, from Theorem 12, it can be identified with an expression of the form (40). A contradiction arises when passing to the equality of those two expressions of \vec{x} modulo $\gamma^r \delta^s \oplus \gamma^{-r} \delta^{-s}$ if one assumes that s/r is not the maximal ratio over all circuits. This contradiction is based on the observation that $(\gamma^{r'} \delta^{s'})^* \pmod{\gamma^r \delta^s \oplus \gamma^{-r} \delta^{-s}}$ cannot be polynomial if $s'/r' > s/r$. \square

11.3 Transfer matrices associated to dater and counter equations

The following notations refer to those introduced in Section 4.3. Let us keep $\mathcal{B} = \mathcal{C} = \{\varepsilon, e\}$ and let $\mathcal{F} = \{\varepsilon, e, \delta\}, \mathcal{G} = \{\gamma\}$. It is obvious that the dioid \mathcal{F}^* is isomorphic to the (complete) dioid $(\mathbb{N} \cup \{-\infty\} \cup \{+\infty\}, \max, +)$ denoted by \mathbb{N}_{\max} for short. Moreover, $\mathcal{F}^* \otimes \mathcal{B}$ and $\mathcal{C} \otimes \mathcal{F}^*$ are identical to \mathcal{F}^* whereas $\mathcal{F}^* \otimes \mathcal{G}$ reduces to the set of elements of the form $a\gamma$ where $a \in \mathbb{N}_{\max}$.

Remark 11 *We might have started with $\mathcal{G} = \{e, \gamma, \gamma^2, \dots\}$, in which case $\mathcal{F}^* \otimes \mathcal{G}$ would amount to the set of polynomials in γ with coefficients in \mathbb{N}_{\max} . The rest of the discussion could be carried on, with some adaptations, in this situation.* \square

An observer (resp. controller) representation (see Theorem 16) is of the form $C(\gamma A)^* B$ where we may consider that the entries of A and B (resp. C) are in \mathbb{N}_{\max} , those of C (resp. B) are in $\{\varepsilon, e\}$. This can be viewed as the transfer matrix corresponding to the following *dater* equations (the algebra is \mathbb{N}_{\max})

$$x_{n+1} = Ax_n \oplus Bu_{n+1} \qquad y_n = Cx_n$$

and γ appears as the unit backward shift operator in the event domain.

Let $\mathbb{Z}_{\max}\langle\langle\gamma\rangle\rangle$ be the set of formal power series in γ with coefficients in \mathbb{Z}_{\max} . Consider the correspondence $\gamma^n \delta^t \mapsto t\gamma^n$ which associates to a monomial in (γ, δ) (with boolean coefficient 1) a monomial in $\mathbb{Z}_{\max}\langle\langle\gamma\rangle\rangle$. This is extended in an obvious manner to formal power series. Recall the calculations rules (56)–(57) in \mathcal{M} and notice that calculating in $\mathbb{Z}_{\max}\langle\langle\gamma\rangle\rangle$ amounts to taking the latter into account but not the former. That is to say, we must introduce the following additional rule in $\mathbb{Z}_{\max}\langle\langle\gamma\rangle\rangle$

$$a\gamma^n \oplus a\gamma^{n'} = a\gamma^{\min(n,n')}$$

This of course can be done more formally by introducing an equivalence relation. Coming back to state space equations, it is realized that an equation like $x_m = ax_{m-n} \oplus ax_{m-n'}$ can be simplified into $x_m = ax_{m-\min(n,n')}$, which is easy to understand directly.

All these considerations can dually be made with *counter* equations (*time domain*) by inverting the role of γ and δ when defining \mathcal{F} and \mathcal{G} , by changing \mathbb{N}_{\max} into \mathbb{N}_{\min} (with an obvious definition of the latter) and $\mathbb{Z}_{\max}\langle\langle\gamma\rangle\rangle$ into $\mathbb{Z}_{\min}\langle\langle\delta\rangle\rangle$, by introducing dual calculations rules, etc... In our opinion, the 2-D domain offers a complete symmetry between counting and dating which is somewhat lost when projecting in either one-dimensional domain.

12 Conclusion

In this paper, we have attempted to provide a unified introduction to unusual, albeit not new, algebraic tools, and to show how they can be used in the construction of a new “linear” system theory. This theory encompasses systems that can otherwise be described as event graphs, a special class of Petri nets. Essentially, in this class, logical “and” can be modelled easily, but not logical “or”, unlike in Automata theory. However, time and events can be manipulated simultaneously. Anyway, event graphs appear as the class of Petri nets that are “linear” when using dioid algebras. We have shown several ways of modelling them (in the event domain, in the time domain and in a 2-D domain) and a different algebra has to be used in each approach, but all fall into the category of dioids.

We have tried to give a sufficiently self-contained presentation, and some algebraic results appear here for the first time. However, there was no attempt

to be exhaustive. On the algebraic side, we did not discuss such issues as “linear dependence” in “vector spaces” on dioids (sets \mathcal{D}^n with an internal addition \oplus and an external multiplication \otimes by scalars in \mathcal{D}). This issue is basic for the system-theoretic issue of “minimal realizations”, a question which is still open (see however [14]). There are several possible definitions of linear dependence [13] and one of them can be characterized by the condition of some “determinant” being null [9]. We did not either discuss the Cayley-Hamilton theorem (see [12, 16]): it holds true in dioids simply because of the proof of Straubing [18] which is based only on combinatorial properties that are common to both the conventional and the dioid algebras. However, the formulation of Cayley-Hamilton theorem in dioids requires some care because of the lack of “minus sign” in these algebras. This theorem should play an important role in realization theory.

On the system-theoretic side, apart from the basic question of minimal realization, we believe that control and optimization issues should be more developed for this theory to prove practically useful. In this paper, we did not review such notions as “stability” and structural controllability and observability previously introduced in the framework of dater equations, nor the associated results concerning stabilizability by dynamic output feedback [4]. The practical significance of these results in the framework of flexible manufacturing systems was discussed in earlier papers [2, 3].

Anyway, considering the amount of efforts that has been devoted to the development of conventional linear system theory during at least twenty years, it must be considered that this new system theory is just in its infancy. However, it is interesting to see how continuous systems and discrete event systems can be essentially considered from the same points of view and concepts, provided that the algebra be adapted. This requires a specific effort.

Acknowledgement Mr. Ramine Nikoukhah who recently joined our group (on leave from MIT) helped us in clarifying some points such as the issue of initial conditions. His contribution is gratefully acknowledged.

References

- [1] P. Caspi and N. Halbwachs, “A functional model for describing and reasoning about time behaviour of computing systems ”, *Acta Informatica*, Vol. 22, pp. 595–627, 1986.
- [2] G. Cohen, D. Dubois, J.P. Quadrat and M. Viot, “Analyse du comportement périodique des systèmes de production par la théorie des dioides”, *INRIA Report No. 191*, Le Chesnay, France, 1983.
- [3] 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 Trans. on Automatic Control*, Vol. AC-30, pp. 210–220, 1985.
- [4] G. Cohen, P. Moller, J.P. Quadrat and M. Viot, “Linear system theory for discrete-event systems”, *23rd IEEE Conf. on Decision and Control*, Las Vegas, Nevada, 1984.
- [5] G. Cohen, P. Moller, J.P. Quadrat and M. Viot, “Une théorie linéaire des systèmes à évènements discrets”, *INRIA Report No. 362*, Le Chesnay, France, 1985.
- [6] G. Cohen, P. Moller, J.P. Quadrat and M. Viot, “Dating and counting events in discrete event systems”, *25th IEEE Conf. on Decision and Control*, Athens, Greece, 1986.
- [7] R.A. Cuninghame-Green, “*Minimax Algebra*”, Springer-Verlag, Berlin, 1979.
- [8] P. Dubreil and M.L. Dubreil-Jacotin, “*Leçons d’Algèbre Moderne*”, Dunod, Paris, 2nd ed., 1964.
- [9] M. Gondran and M. Minoux, “Linear algebra in dioids: a survey of recent results”, *Annals of Discrete Mathematics*, Vol. 19, pp. 147–164, 1984.
- [10] S. Helbig, “Optimization problems on extremal algebras: necessary and sufficient conditions for optimal points”, In: B. Borowski and F. Deutsch, Eds., “*Parametric optimization and approximation*”, Birkhäuser-Verlag, Basel, Switzerland, pp. 166-184, 1985.
- [11] F.P Kelly, “*Reversibility and Stochastic Networks*”, Wiley, New-York, New-York, 1979.
- [12] P. Moller, “Théorème de Cayley-Hamilton dans les dioides et application à l’étude des systèmes à évènements discrets”, *7th INRIA International Conference on Analysis and Optimization of Systems*, Antibes, France, (Proc. Springer-Verlag), 1986.
- [13] P. Moller, “Notions de rang dans les dioides vectoriels”, *CNRS/CNET/INRIA Seminar “Algèbres Exotiques et Systèmes à Evènements Discrets”*, Issy-les-Moulineaux, France, 1987.
- [14] G.J. Olsder, “Some results on the minimal realization of discrete event systems”, *25th IEEE Conf. on Decision and Control*, Athens, Greece, 1986.
- [15] G.J. Olsder, J.A.C. Resing, R.E. de Vries, M.S. Keane and G. Hooghiemstra, “Discrete event systems with stochastic processing times”, submitted to *27th IEEE Conf. on Decision and Control*, Austin, Texas, 1988.

- [16] G.J. Olsder, C. Roos, “Cramer and Cayley-Hamilton in the max-algebra”, to appear in *Linear Algebra and its Applications*, 1988.
- [17] C.V. Ramamoorthy and G.S. Ho, “Performance evaluation of asynchronous concurrent systems using Petri nets”, *IEEE Trans. on Software Engineering*, Vol. SE-6, pp. 440–449, 1980.
- [18] H. Straubing, “A combinatorial proof of the Cayley-Hamilton theorem”, *Discrete Mathematics*, Vol. 43, pp. 273-279, 1983.
- [19] R.P. Wiley and R.R. Tenney, “Performance evaluation of decision-free stochastic timed Petri nets”, *24th IEEE Conf. on Decision and Control*, Ft. Lauderdale, Florida, 1985.