The original version of this paper was presented as an invited plenary paper
at the IFAC Conference on System Structure and Control, Nantes, France , July 8–10, 1998.
This slightly revised version appeared in Annual Reviews in Control (IFAC, Elsevier), 23 (1), 207-219, 1999.

MAX-PLUS ALGEBRA AND SYSTEM THEORY: WHERE WE ARE AND WHERE TO GO NOW [1]

Guy Cohen [*,†] **Stéphane Gaubert** [†] **Jean-Pierre Quadrat** [†]

[*] *Centre Automatique et Systèmes, École des Mines de Paris, Fontainebleau, France,* cohen@cas.ensmp.fr
[†] *INRIA-Rocquencourt, Le Chesnay, France*

Abstract: More than sixteen years after the beginning of a linear theory for certain discrete event systems in which max-plus algebra and similar algebraic tools play a central role, this paper attempts to summarize some of the main achievements in an informal style based on examples. By comparison with classical linear system theory, there are areas which are practically untouched, mostly because the corresponding mathematical tools are yet to be fabricated. This is the case of the geometric approach of systems which is known, in the classical theory, to provide another important insight to system-theoretic and control-synthesis problems, beside the algebraic machinery. A preliminary discussion of geometric aspects in the max-plus algebra and their use for system theory is proposed in the last part of the paper.

Résumé: Plus de seize ans après le début d'une théorie linéaire de certains systèmes à événements discrets dans laquelle l'algèbre max-plus et autres outils algébriques assimilés jouent un rôle central, ce papier cherche à décrire quelques uns des principaux résultats obtenus de façon informelle, en s'appuyant sur des exemples. Par comparaison avec la théorie classique des systèmes linéaires, il existe des domaines pratiquement vierges, surtout en raison du fait que les outils mathématiques correspondants restent à forger. C'est en particulier le cas de l'approche géométrique des systèmes qui, dans la théorie classique, est connue pour apporter un autre regard important sur les questions de théorie des systèmes et de synthèse de lois de commandes à côté de la machinerie purement algébrique. Une discussion préliminaire sur les aspects géométriques de l'algèbre max-plus et leur utilité pour la théorie des systèmes est proposée dans la dernière partie du papier.

Keywords: Discrete event systems, max-plus algebra, dioids, algebraic system theory

1. INTRODUCTION

For what later became the Max-Plus working group at INRIA, the story about discrete event systems (DES) and max-plus algebra began in August 1981, that is more than sixteen and a half years ago, at the time this paper is written. Actually, speaking of 'discrete event systems' is somewhat anachronistic for that time when this terminology was not even in use. Sixteen years is not a short period of time compared with that it took for classical linear system theory to emerge as a solid piece of science. On the one hand, those who have been working in the field of max-plus linear systems have benefitted from the guidelines and concepts provided by that classical theory. On the other hand, the number of researchers involved in this new area of system theory for DES has remained rather small when compared with the hundreds of their colleagues who contributed to the classical theory. In addition, while this classical theory was based on relatively well established mathematical tools, and in particular linear algebra and vector spaces, the situation is quite different with max-plus algebra: this algebra, and similar other algebraic structures sometimes referred to as 'semirings' or 'dioids', were already studied by several researchers when we started to base our system-theoretic work upon such tools; yet, today, a very basic understanding of some fundamental mathematical issues in this area is still lacking, which certainly contribute to slow down the progress in system theory itself. This is why an account of the present situation in the field can hardly separate the system-theoretic issues from the purely mathematical questions.

[1] This work has been partially supported by a TMR contract No. ERB-FMRX-CT-96-0074 of the European Community in the framework of the ALAPEDES network.

Indeed, the models and equations involved are not restricted to DES: connections with other fields (optimization and optimal decision processes, asymptotics in probability theory, to quote but a few) have been established since then, and this has contributed to create a fruitful synergy in this area of mathematics. Yet, this paper will concentrate on DES applications. To be more specific, while classical system theory deals with systems which evolve in time according to various physical, chemical, biological ... phenomena which are described by ordinary or partial differential equations (or their discrete-time counterparts), DES refer to 'man-made' systems, the importance of which has been constantly increasing with the emergence of new technologies. Computers, computer networks, telecommunication networks, modern manufacturing systems and transportation systems are typical examples. Among the basic phenomena that characterize their dynamics, one may quote *synchronization* and *competition* in the use of common resources. Competition basically calls for *decisions* in order to solve the conflicts (whether at the design stage or on line, through priority and scheduling policies). Through 'classical' glasses, synchronization looks like a very nonlinear and nonsmooth phenomenon. This is probably why DES have been, for a long time, left apart by classical system and control theory; they were considered rather in the realm of operations research or computer science, although they are truly dynamical systems.

Linear models are the simplest abstraction (or ideal model) upon which a large part of classical system and control theory have been based until the late sixties. To handle more complex models, say, with smooth nonlinearities, it was necessary to adapt the mathematical tools while keeping most of the concepts provided by earlier developments: differential geometry, power series in noncommutative variables, differential algebra have been used to develop such models for which essential questions such as controllability and observability, stabilization and feedback synthesis, etc., have been revisited. Max-plus, min-plus and other idempotent semiring structures turn out to be the right mathematical tools to bring back linearity, in the best case, or at least a certain suitability with the nature of phenomena to be described, in this field of DES.

The purpose of this paper is twofold. On the one hand, it tries to summarize some of the most basic achievements in the last sixteen years in this new area of system theory turned towards DES performance related issues (as opposed to logical aspects considered in the theory of Ramadge and Wonham (1989)). Because of the space limitation, we will mostly proceed by way of examples and the treatment will be necessarily sketchy. We will rely upon several surveys already devoted to the subject (Cohen *et al.*, 1989*a*; Cohen, 1994; Quadrat and Max Plus, 1995; Gaubert and Max Plus, 1997) in addition to the book (Baccelli *et al.*, 1992*b*). On the other hand, the paper tries to suggest new directions of developments. This essentially concerns the understanding of *geometric* aspects of system theory in the max-plus algebra. Investigations are currently undertaken in this area, so we will just sketch the kind of questions we try to address by discussing examples.

2. LINEAR EQUATIONS OF TEG

2.1 *State space equations*

A common tool to describe discrete event systems is the Petri net formalism of which a basic knowledge is expected from the reader (see e.g. (Murata, 1989)). Since we are interested in performance related issues, we consider *timed* Petri nets. The subclass of *timed event graphs* (TEG) is the class in which all places have a single transition upstream and a single one downstream [2]. A single downstream transition for each place practically means that all potential conflicts in using tokens in places have been already arbitrated by some predefined policy. A single upstream transition means that there is a single source of token supply for each place (hence there is no competition in either consumption or supply of tokens in TEG). These limitations are certainly restrictive for most applications, and they can generally be satisfied by making some design and scheduling decisions at an upper hierarchical level (the purpose may then be to evaluate these decisions and to try to improve them). But this is the price to pay for dealing with *linear* systems. Attempts to deal with more general Petri nets can be found e.g. in (Baccelli *et al.*, 1992*a*; Gaubert and Mairesse, 1997; Cohen *et al.*, 1998). Yet, there are many interesting real systems which can be fairly well described by TEG.

TEG correspond exactly to the class of timed Petri nets which are described by max-plus or min-plus linear equations. Consider for example the TEG depicted in Fig. 1. While dots represent tokens as usual, bars represent the holding times of places measured in a common time unit, that is, the minimum time a token must stay in a place before it can be used to fire the downstream transition (with no loss of generality, holding times can be put in places only, the firing of transitions being instantaneous).

The convention is that transitions have names (indicated in the figure) which are also the names of variables attached to them. The first variables considered are *daters*: $x_i(k)$ denotes the earliest time at which transition x_i can fire for the $(k+1)$-st time (because the first event is numbered 0 for some tricky reason). The following recursive equations can be established (Cohen *et al.*, 1985; Cohen *et al.*, 1989*a*; Baccelli *et al.*, 1992*b*):

[2] Hence, in event graphs, places can be considered as 'arcs' and transitons as 'nodes'.

$$x_1(k) = x_3(k-2) \oplus u(k)\ , \tag{1a}$$
$$x_2(k) = \big(1 \otimes x_1(k)\big) \oplus \big(1 \otimes x_3(k-2)\big)\ , \tag{1b}$$
$$x_3(k) = \big(3 \otimes x_1(k-1)\big) \oplus \big(1 \otimes x_2(k)\big)\ , \tag{1c}$$
$$y(k) = x_3(k)\ , \tag{1d}$$

where $\oplus$ stands for max and $\oplus$ for $+$. The occurrence of max is a direct consequence of synchronization: one must wait for the presence of at least one token in all upstream places of any transition, hence, for the *last* such condition to be satisfied before the transition firing can occur.

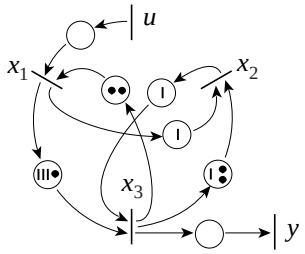

Fig. 1. A TEG

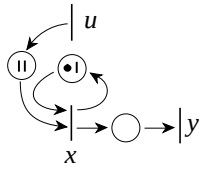

Fig. 2. Its reduced form

2.2 *Idempotent semirings ('dioids'): a few line digest*

The *max-plus semiring* is the set $\mathbb{R}$ of real numbers (plus $-\infty$), endowed with max as 'addition' and $+$ as 'multiplication'. It is an *idempotent semiring*, also called *dioid*, i.e. a set equipped with a commutative, associative and idempotent sum ($a \oplus a = a$), a 'zero' denoted ε and equal to $-\infty$, an associative product, a 'unit' element denoted e and equal to 0, in which product distributes over sum (guess what would happen if we interchange the roles of max and $+$). Of course, the product is also commutative, but this is a feature which will be lost, for example, when considering square matrices instead of scalars, with the natural matrix addition and multiplication derived from scalar operations. An element $x \neq \varepsilon$ of the max-plus dioid has an inverse for $\otimes$, namely $-x$, but the existence of a multiplicative inverse is not part of the minimal set of axioms used to define 'dioids' in general, although it provides useful additional properties when it holds true.

Remark 1. By the loose expression 'max-plus algebra', we generally mean the max-plus dioid as defined above, or the similar structure with $\mathbb{Z}$ instead of $\mathbb{R}$. In the max-plus algebra, the 'unit' element e (equal to 0) should not be confused with 1; $1 \otimes a$ is *not* equal to a and $1 \otimes 1 = 2$). As usual, the multiplication sign $\otimes$ is often omitted and $\otimes$ has priority over $\oplus$.

Due to the idempotent character of addition, a dioid cannot be embedded in a ring. But thanks to idempotency, it can be equipped with the natural order relation $a \succeq b$ iff $a = a \oplus b$. Then, $a \oplus b$ coincides with the least upper bound of $\{a, b\}$, which is usually denoted $a \vee b$. Hence, a dioid is in particular a sup-semilattice (this is sometimes the most important structure to consider, which is obviously extended to 'vectors'). If, in addition, the sup-semilattice is *complete* (i.e. infinite sets have a least upper bound for the natural order, and multiplication is left and right distributive with respect to least upper bounds — this is the case in particular for the max-plus semiring, completed with $+\infty$), then the greatest lower bound of two elements (denoted $a \wedge b$) automatically exists.

2.3 *Canonical equations*

Equations (1) can be written in *matrix* form ('missing' entries are set to $\varepsilon = -\infty$). Generally speaking, for any timed event graph, one obtains the following kind of equations:

$$x(k) = \bigoplus_{i=0}^{M} \big(A_i x(k-i) \oplus B_i u(k-i)\big)\ , \tag{2a}$$
$$y(k) = \bigoplus_{i=0}^{M} C_i x(k-i)\ , \tag{2b}$$

where x, u, y are vectors of dimensions equal to the numbers of internal, input and output transitions [3], resp., A_i, B_i, C_i are matrices of appropriate dimensions with entries in the max-plus algebra, and M is the maximal number of tokens in the initial marking.

In transforming these equations towards a canonical form, the first stage aims at removing the implicit part $A_0 x(k)$ in (2a), if any. The nonzero entries of A_0 correspond to holding times of places with no tokens in the initial marking. In principle, in the corresponding subgraph, there should be no circuits; otherwise, all transitions in those circuits are frozen for ever since the numbers of tokens in circuits are preserved during the event graph evolution. Consequently, there is a numbering of internal transitions such that A_0 can be written in strictly lower triangular form; hence, A_0^n becomes zero for a sufficient large n (not greater than the matrix dimension) and the so-called 'Kleene star', that is, the infinite sum

$$A_0^* = \bigoplus_{n \in \mathbb{N}} A_0^n \tag{3}$$

is well defined. Generally speaking, in the max-plus algebra (and in a more general framework indeed) a^*b is the *least* solution of the implicit equation $x = ax \oplus b$ whenever a^* can be given a meaning.

These considerations help removing the implicit part of (2a) considered from $k = 0$ to $+\infty$ as an implicit equation in the state trajectory $x(\cdot)$. Picking the least

[3] Internal transitions are those having both upstream and downstream transitions, input transitions have only downstream transitions, and output transitions have only upstream transitions. If there are arcs directly connecting input to output transitions (through places of course), then there are additional terms of the form $D_i u(k-i)$ in (2b), which does not fundamentally change the rest of manipulations to come.

solution in this implicit equation subsumes that transition firings occur as soon as they become possible, but also that the 'initial conditions' $\{x(k)\}_{k<0}$ are the least ones, that is, ε. This amounts to assuming that tokens of the initial marking are immediately available at the beginning of the game. Other nonzero initial conditions can be enforced at the price of controlling the arrival of tokens of the initial marking by additional auxiliary input transitions (see (Baccelli *et al.*, 1992*b*, §5.4.4.2)).

The next stage in equation manipulation aims at obtaining a unit delay in the first term of the right-hand side of (2a) and a zero delay in the second term therein, together with a zero delay in the right-hand side of (2b). This is obtained by increasing the 'state vector' dimension which must incorporate delayed versions of the x_i and u_i variables. This stage is classical in system theory and need not be described in details here.

Finally, the canonical form of (2) is (without introducing a new notation for the possibly augmented state vector)

$$x(k) = Ax(k-1) \oplus Bu(k)\ ;\quad y(k) = Cx(k)\ . \quad (4)$$

The implicit part can be eliminated by successive substitutions of scalar variable, rather than by a naive matrix star computation (there should be an appropriate order for these substitutions for the same reason why A_0 can be written in a strictly lower triangular form). For example, considering (1) again, one would first substitute the right-hand side of (1a) for $x_1(k)$ in the right-hand side of (1b), then use this new equation to eliminate $x_2(k)$ in the right-hand side of (1c). After the implicit part has been so eliminated, it is realized that x_2 no longer appears at the right-hand side of the dynamics (including the observation (1d)). Consequently, x_2 is not part of the state vector. On the other hand, a new state variable must be introduced to account for the second-order delay in x_3: let us set $x_4(k) = x_3(k-1)$ (the reader may imagine the corresponding manipulation in the event graph). Finally, from (1), one derives the canonical form (4) with the following state vector and matrices:

$$x = \begin{pmatrix} x_1 \\ x_3 \\ x_4 \end{pmatrix};\quad A = \begin{pmatrix} \varepsilon & \varepsilon & e \\ 3 & \varepsilon & 2 \\ \varepsilon & e & \varepsilon \end{pmatrix};\quad B = \begin{pmatrix} e \\ 2 \\ \varepsilon \end{pmatrix};$$
$$C = \begin{pmatrix} \varepsilon & e & \varepsilon \end{pmatrix}. \quad (5)$$

Remark 2. There is another representation of TEG in terms of 'counters' instead of daters: let $x_i^\flat(t)$ denotes the number of the first firing to occur at transition x_i at or after time t (we assume time is discrete to preserve the symmetry to be explained later on between daters and counters). In mathematical terms, $x_i^\flat(t) = \inf_{x_i(k)\geq t} k$. Using either the definition directly or the theory of residuation (note that $t \mapsto x_i^\flat(t)$ is a possible definition for the inverse of $k \mapsto x_i(k)$), one can show that counters obey min-plus linear equations. There is an alternative definition of counters as $x_i^\sharp(t) = \sup_{x_i(k)\leq t} k$ and one can prove that $x_i^\flat(t) = x_i^\sharp(t-1)+1$. Indeed, these two definitions pertain to the notions of *dual residuation* and of *residuation* of the dater function, resp. (see §4.2). For some tricky reason, the former definition is preferable to the latter.

2.4 *Transfer functions*

In classical system theory, the z-transform allows one to represent discrete-time trajectories by formal power series with positive and negative powers of the formal variable z. For dater trajectories $\{x(k)\}$, we introduce the γ-transform $X(\gamma) = \bigoplus_{k\in\mathbb{Z}} x(k)\gamma^k$, where γ is an indeterminate which may also be considered as the backward shift operator (formally, $\gamma x(k) = x(k-1)$). Starting either from the rough form (1) or from the canonical form (4) and applying the γ-transform yields implicit equations in $X_i(\gamma)$ (plus an equation for $Y(\gamma)$) which can be solved again by appealing to the Kleene star (now, of polynomials in γ with max-plus coefficients). With our example of Fig. 1, it is easy to eliminate all $X_i(\gamma)$ but $X_3(\gamma)$, which is also $Y(\gamma)$, and to obtain

$$Y(\gamma) = 2(2\gamma^2 \oplus 3\gamma^3)^*(e \oplus 1\gamma)U(\gamma)\ .$$

The next stage is to realize that $(2\gamma^2 \oplus 3\gamma^3)^*(e \oplus 1\gamma)$ coincides with $(1\gamma)^*$ (simply by expanding both expressions). Hence, we finally obtain

$$Y(\gamma) = 2(1\gamma)^*U(\gamma)\ . \quad (6)$$

This expression allows the calculation of the output trajectory corresponding to any input history; hence, it completely summarizes the input-output relationship. Generally speaking, from the canonical form (4), it follows that

$$Y(\gamma) = C(\gamma A)^*BU(\gamma)\ . \quad (7)$$

The expression $H(\gamma) = C(\gamma A)^*B$ is called the *transfer matrix* (or *function* in the single-input-single-output case).

The right-hand side of (7) is the product of two formal power series, namely $H(\gamma)$ and $U(\gamma)$. Back in the event domain (that of index k), $y(\cdot)$ is a 'convolution' of the sequences $h(\cdot)$ and $u(\cdot)$, of which $H(\gamma)$ and $U(\gamma)$ are the γ-transforms: indeed, 'convolution' means 'sup-convolution' in the max-plus algebra. As Laplace transform converts convolutions into products in classical system theory, γ-transform converts sup-convolutions into products here. When we restrict inputs $u_j(\cdot)$ to be nondecreasing control histories, we can also limit ourselves to consider nondecreasing functions $h_{ij}(\cdot)$. Such a trajectory $h_{ij}(\cdot)$ is the *impulse response* of system (4) when looking at output i and input j; more precisely, it is the trajectory $y_i(\cdot)$ caused by an infinity of tokens placed at transition u_j at time 0, whereas at all other input transitions, it is assumed that unlimited numbers of tokens are available since $-\infty$; the reader may check that in terms of

γ-transforms this indeed corresponds to $U_j(\gamma) = \gamma^*$ and $U_l(\gamma) = \varepsilon$ for $l \neq j$. In fact, for γ-transforms of *nondecreasing* dater trajectories, γ^* behaves as the unit element e. The story about nondecreasing sequences is longer than what we can tell here and is at the heart of the two-dimensional representation and the $\mathcal{M}_{\text{in}}^{\text{ax}}[\![\gamma, \delta]\!]$ algebra alluded to at Rem. 3 hereafter.

Back to our example, it should not be difficult to check that (6) is also the transfer function of the TEG represented in Fig. 2 which is described by the one-state variable system

$$y(k) = 1y(k-1) \oplus 2u(k) \ . \tag{8}$$

By comparing Fig. 2 with Fig. 1, or (1)–(5) with (8), the reader should convince himself that relatively simple algebraic calculations bring simplifications of a given (and already relatively simple) system which can hardly be obtained by other means. These simplifications would be more spectacular if we had started from a more complex system. Of course, then, the help of some software (e.g., that of S. Gaubert named 'MAX' and based on Maple) would be desirable to achieve the calculations. To convince the reader of the interest of transfer function calculation, we invite him to reconsider the slight variation of Fig. 1 in which the arc from transition u goes to transition x_2 instead of x_1. The corresponding system admits

$$H(\gamma) = 1 \oplus 3\gamma^2(1\gamma)^* \tag{9}$$

as its transfer function which cannot be realized with less than 2 state variables (this is, by the way, a good exercise to try out!). These changes seem rather unpredictable without appealing to algebra.

Remark 3. Since a representation with counters can also be used (see Rem. 2), there is an associated transfer function using the δ-transform, where δ is the backward shift operator in the time domain rather than in the event domain as γ (formally, $\delta x(t) = x(t-1)$). Instead of (6), we would get $Y(\delta) = \delta^2(1\delta)^*U(\delta)$. Note that, because of the double delay represented by the factor δ^2, 3 state variables are now necessary to realize this transfer function in the canonical form with counters, whereas only 1 was required with daters for the same system. This is not surprising since delays are related to the initial marking in the dater representation, whereas they are related to holding times in the dater representation. This remark shows however that the notion of *minimal realization* needs some careful elaboration.

In (Baccelli *et al.*, 1992*b*, Chap. 5), a two-dimensional representation of input-output maps with γ and δ as commutative formal variables of power series with boolean coefficients is explained, and its advantages over the one-dimensional representations in γ (with coefficients in max-plus) or in δ (with coefficients in min-plus) are enumerated. There is no room to develop the corresponding theory and to introduce the so-called $\mathcal{M}_{\text{in}}^{\text{ax}}[\![\gamma, \delta]\!]$ algebra here. In this new representation, the transfer function of our example reads $H(\gamma, \delta) = \delta^2(\gamma\delta)^*$.

3. A QUICK REVIEW OF SYSTEM-THEORETIC RESULTS FOR TEG

3.1 *Asymptotic behavior and eigenvalues*

Conventional linear systems have 'modes' which are reached asymptotically when systems are stable; these modes are related to their eigenstructures. Similar notions exist for *autonomous* TEG obeying equations of the form $x(k) = Ax(k-1)$. As usual, an eigenvalue is a (rational) number λ (possibly equal to ε) such that there exists a nontrivial eigenvector x (that is, $x \neq \varepsilon$) satisfying $Ax = \lambda x$. In the max-plus algebra, λx means that the same scalar value λ is added to all coordinates of x. Hence, if $x(0)$ is equal to such an eigenvector, at every stage (that is, every time the event counter k is incremented by 1), the same time amount λ elapses at all transitions. Algebraically, $x(k) = \lambda^k x(0)$. Essential questions are whether such an eigenpair exists in general, and whether all initial conditions are eventually absorbed in a similar 'periodic' regime.

When A is *irreducible* (that is, the corresponding TEG is strongly connected, or otherwise stated, all transition firings are dependent on each other in the long term), the answer is relatively easy: there exists a unique eigenvalue but possibly several eigenvectors. The eigenvalue is given by the formula

$$\lambda = \bigoplus_{j=1}^{n} \left(\text{trace}(A^j)\right)^{1/j} \ , \tag{10}$$

where n is the dimension of the square matrix A and all operations are in the max-plus algebra. In a less cryptic way, λ is the *largest average circuit weight* of the directed graph canonically associated with A, or, equivalently, the largest average weight of a directed circuit in the original TEG. When there is exactly one token in each internal place, the average weight of such a circuit is defined as the number of bars divided by the number of arcs or places along the circuit. More generally, when the TEG is not in the 'canonical' form in which a place (an arc) between internal nodes (transitions) corresponds exactly to one token of the initial marking, the average weight is the ratio of the number of bars by the number of tokens along the circuit. For the TEG of Fig. 1, this ratio is equal to 1 for all circuits, and this is also the case for that of Fig. 2.

The structure of the 'eigenspace' is related to the structure of the *critical graph*, which is the subgraph such that all nodes and arcs belong to at least a critical circuit (that is, a circuit for which the extremal average weight λ is reached). More precisely, let x_i be a transition belonging to a critical circuit and consider the 'normalized' matrix $A_\lambda = \lambda^{-1}A$ (which means subtracting λ, assumed finite here, from ev-

ery entry of A). An eigenvector is obtained as the i-th column of $(A_\lambda)^+ = A_\lambda(A_\lambda)^*$. All columns of this matrix corresponding to transitions in the same strongly connected component of the critical graph provide proportional eigenvectors. In particular, if the critical graph is strongly connected (which is the case with the TEG in Fig. 1), there is a *unique* eigenvector (up to a multiplicative constant). We refer the reader to (Baccelli *et al.*, 1992*b*; Gaubert and Max Plus, 1997) and references therein for a complete treatment of these questions even in the case when A is not irreducible.

The critical graph also plays a role when considering the asymptotic behavior of the iterates $A^k x(0)$ of the autonomous system from any initial condition $x(0)$. Again, in the simplest case of irreducible matrices, it can be proved that

$$\exists c \geq 1, K \in \mathbb{N}: \ \forall k > K, \ \ A^{k+c} = \lambda^c A^k \ . \quad (11)$$

That is to say, K is the duration of a transient part beyond which, if $c = 1$, any initial condition has been absorbed in an eigenvector. If $c > 1$, the behavior is 'periodic' over c steps, with the same average time λ between two successive firings at all transitions. This c is called the *cyclicity* and an exact formula for it is: the `lcm` over all strongly connected components of the critical graph of the `gcd` of the 'lengths' (that is, token numbers) of all circuits in each strongly connected component of that graph. With the TEG of Fig. 1, all internal arcs and transitions belong to the critical graph which is strongly connected. The are two elementary circuits with 2 tokens and one with 3: the `gcd` of 2 and 3 is $c = 1$. By computing the successive powers of A in (5), it is discovered that $K = 5$, $c = 1$ and $\lambda = 1$. The length of the transient cannot be bounded after the dimension of A. An effective bound, which involves the numerical values of the entries of A, and in particular the average weight of the 'second critical circuit' of A, is implicit in the proof of (11).

3.2 *Stabilization, feedback synthesis and resource optimization*

A completely observable and controllable (conventional) linear system can be stabilized by dynamic output feedback. With TEG, all trajectories are nondecreasing, and stability must be given an adequate meaning: by 'stability', we essentially mean that tokens do not accumulate indefinitely inside the graph. A sufficient condition is that the whole system is synchronized, that is, it consists of a single strongly connected component. A TEG is *structurally controllable* (resp. *observable*) if every internal transition can be reached by a directed path from at least one input transition (resp. is the origin of at least one directed path to some output transition). Structurally controllable and observable TEG can be stabilized by output feedback in that the graph can be made strongly connected by adding appropriate arcs from output to input transitions.

However, since new circuits are created by closing the feedback loops, there is a risk that the eigenvalue of the closed-loop system gets larger than that of the open-loop system, which means a deterioration in performance (that is, of the throughput $1/\lambda$, with a classical interpretation of the inverse here). Therefore, an interesting question is how to enforce stability while preserving performance, or at least not lowering it too much (of course, the system cannot be speeded up by adding new circuits, hence new synchronization constraints). This problem can be viewed as the equivalent notion of *pole placement* or *loop shaping* in classical system theory. For TEG, this means that the new circuits created by feedback must have an average weight which remains below a given threshold. Since all such circuits traverse the feedback arcs, it suffices to put enough tokens in the initial marking of these arcs: this yields a *dynamic* feedback in that $u(k)$ is made dependent of some $y(k-m)$. Obviously, for m large enough, the ratio (nr. of bars/nr. of tokens) of such circuits ceases to be critical.

Nevertheless, from the practical point of view, increasing m means increasing the number of tokens permanently present in the system, and sometimes this even requires additional physical resources (parking or storage room, pallets to carry parts in a workshop, etc.). Hence, the next problem is to ensure the desired level of performance under 'budget' constraints. We are here in the realm of resource optimization (Gaubert, 1995), (Gaubert, 1992, Chap. 9). The principle of 'kanban' systems is also very akin to the previous considerations (Di Mascolo, 1990).

Recently, the problem of feedback synthesis have been reconsidered by Cottenceau *et al.* (1998) in the following form. Consider a system $Y = HU$ (say, here, $Y, U, H \in \mathcal{M}_{\text{in}}^{\text{ax}}[\![\gamma, \delta]\!]$) and the feedback law $U = FY \oplus V$, which yields the closed-loop system $Y = (HF)^* HV$. Instead of trying to preserve the open-loop system eigenvalue only, the idea is to find the *greatest causal* feedback law F which preserves the *whole open-loop transfer function* H. 'Causal' essentially means that F can be represented by a sum of monomials in (γ, δ) with nonnegative exponents only (this is a 'quick and dirty' definition). 'Greatest feedback law' means that inputs will be delayed as much as possible, which intuitively aims at minimizing the number of tokens present in the system. However, the authors did not prove that their design enforces stability (in the previous sense) for structurally controllable and observable systems in general. But they showed that their problem admits a simple analytic solution based on residuation theory (see §4.2 hereafter), namely F is the causal part (keep only monomials with positive exponents) of $H \backslash H / H$, where $\backslash$ and $/$ are the residuated operations of left, resp. right, multiplication of power series. The reader may consider the exercise of calculating this F for system (8), represented in Fig. 2, the transfer function of which (in $\mathcal{M}_{\text{in}}^{\text{ax}}[\![\gamma, \delta]\!]$) is, according to (6), $H = \delta^2(\gamma\delta)^*$. The

answer is $F = \gamma^2(\gamma\delta)^*$. An implementation of this feedback is represented in Fig. 3.

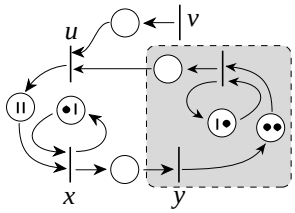

Fig. 3. Feedback law (in the grey box) preserving open-loop transfer

3.3 *Realizability, rationality and periodicity*

In conventional system theory, a necessary and sufficient condition for a transfer function to admit a finite dimensional time-invariant linear system realization is that it is *rational*. For $\mathcal{M}_{\text{in}}^{\text{ax}}[\![\gamma, \delta]\!]$ transfer matrices, an even stronger result holds true since the following *three* properties are equivalent:

(1) the transfer matrix can be realized by a TEG with constant (nonnegative) holding times;
(2) the transfer matrix is rational (and causal);
(3) the transfer matrix is periodic (and causal).

In Rem. 4 below, we discuss a more mathematical statement of the first property above. The second property means that each entry of the matrix belongs to the closure of $\{\varepsilon, e, \gamma, \delta\}$ by finitely many $\oplus$, $\otimes$ and * operations. The third property means that each entry can be written as an expression of the form $p \oplus qr^*$ in which p and q are polynomials in (γ, δ) which represent the transient behavior and the repeated pattern, resp., whereas r is a monomial $\gamma^k\delta^t$ which reproduces the pattern q along the 'slope' t/k. For TEG with strongly connected internal transitions, this slope is nothing but the unique eigenvalue (in the dater representation). Additional constraints can be put on the relative degrees and valuations of p, q and r. For example, the transient part p need not extend beyond the point where the periodic part starts, that is the degrees of p in (γ, δ) can be strictly less than the valuations of q.

For systems with very long transient parts (check for example $\delta^{20}(\gamma\delta)^* \oplus (\delta^{11}\gamma^{10})^*$), this representation may not be very clever. Consider now the transfer function (9) again (which may be written as $\delta \oplus \gamma^2\delta^3(\gamma\delta)^*$). Obviously, $p = \delta$, $q = \gamma^2\delta^3$ and $r = \gamma\delta$. The left-hand side of Fig. 4 depicts the TEG which is immediately suggested by this way of writing the transfer function, and which corresponds to a 3-dimensional state system in terms of daters. The right-hand side of the same figure represents a TEG with the same transfer function and which corresponds to a 2-dimensional state vector (as was announced earlier). Indeed, the corresponding way of writing the transfer function is $\delta\big(\gamma^2\delta^2(\gamma\delta)^*\big)^*$, that is, with two levels of stars. This example suggests that such a representation may be more appropriate in some cases.

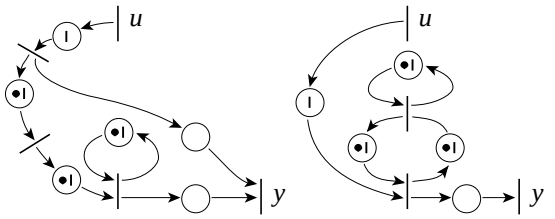

Fig. 4. Two TEG with the same transfer function (9)

This issue of 'canonical' representations of elements in $\mathcal{M}_{\text{in}}^{\text{ax}}[\![\gamma, \delta]\!]$ in a way which allows one to easily check the equality of two such elements in this algebra and which is, at the same time, easy to recover (after various manipulations), efficient in terms of storage, of simulation, and of calculation is mostly an open question; it is central for the design of algebraic computational software tools in $\mathcal{M}_{\text{in}}^{\text{ax}}[\![\gamma, \delta]\!]$.

Remark 4. Instead of speaking of realization of transfer matrices by TEG, one can state property (1) above as the fact that $H(\gamma, \delta)$ can be written as $C(\gamma A_1 \oplus \delta A_2)^* B$ (compare with (7)) for some *Boolean* matrices C, A_1, A_2, B of appropriate dimensions (that is, entries are solely equal to ε or e). Such a definition seems a good basis to tackle the problem of *minimal realization* which would be defined as the minimal inner dimension in this expression (that of A_1 and A_2). This way, neither the dater nor the counter representation is privileged and the amount of storage subsumed by the state vector dimension refers now to the storage of 'bits' of information (boolean values). For the transfer function (6), a possible realization is

$$A_1 = \begin{pmatrix} \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & e \\ \varepsilon & \varepsilon & \varepsilon \end{pmatrix}; \quad A_2 = \begin{pmatrix} \varepsilon & \varepsilon & \varepsilon \\ e & \varepsilon & \varepsilon \\ \varepsilon & e & \varepsilon \end{pmatrix}; \quad B = \begin{pmatrix} e \\ \varepsilon \\ \varepsilon \end{pmatrix};$$
$$C = \begin{pmatrix} \varepsilon & \varepsilon & e \end{pmatrix}.$$

At this moment, we have no non enumerative way to claim that this is a minimal realization. This problem of minimal realization remains a very challenging issue in the field: it is solved only for special subclasses of systems, generally in the framework of dater representations (see e.g. (Gaubert *et al.*, 1998) and references therein).

3.4 *Frequency responses*

In conventional linear system theory, sine functions of any frequency (and starting from time $-\infty$) are eigenfunctions of transfer functions $H(s)$, that is, the output is equal to the input up to amplification and phase shift. The amplification gain and the phase shift at the frequency ω are computed by replacing the formal operator s by the numerical value $j\omega$ in the expression of $H(s)$. For TEG, the analogues of sine functions are certain periodic inputs with any rational 'slope' in the

plane $\mathbb{Z}^2$ where the x-axis is the event domain and the y-axis is the time domain (these periodic inputs are in fact the best approximations from below, on the discrete $\mathbb{Z}^2$-grid, of continuous linear functions with corresponding slopes). The outputs caused by such inputs ('frequency responses') are identical to the inputs, up to the fact that they are shifted along the two axes. Shifts can be evaluated using the slope of the input as a numerical argument of the transfer function, in some way (see (Baccelli *et al.*, 1992*b*, §5.8) or (Cohen *et al.*, 1989*b*) for more detailed explanations). These shifts become infinite when the slope of the input gets strictly smaller than the asymptotic slope of the impulse response: indeed, *smaller* slope means *faster* input rate than what the system is able to process, and thus, tokens will accumulate indefinitely inside the system. In this case, the intrinsic (maximal) throughput of the system will show up instead at the output: this is a kind of 'low pass' effect.

In the evaluation of the event and time domain shifts at any frequency, it turns out that only the concave hull of the impulse response is important. For example, the transfer function in (9) has the same frequency response as the transfer $\delta(\gamma\delta)^*$ (when inputs are started from $-\infty$ in order to remove the transient part of the response).

3.5 *Costate equations and second-order theory*

In conventional optimal control, Pontryagin's minimum principle introduces a backward equation for a vector ξ called 'co-state' or 'adjoint state'. In the linear theory of TEG, a similar notion arises about the following problem: given an output (dater) trajectory $\{y(\cdot)\}$, find the *latest (greatest)* input trajectory $\{u(\cdot)\}$ which yields an output trajectory less (earlier) than the given one. This is again a typical problem in the theory of residuation which is discussed at §4.2: indeed, if $H(\gamma)$ is the transfer function, then the problem is to find the greatest $U(\gamma)$ such that $H(\gamma)U(\gamma) \preceq Y(\gamma)$. The solution of this problem is $U(\gamma) = H(\gamma)\backslash Y(\gamma)$ (recall that $\backslash$ denotes the residuation of multiplication to the left — call it 'left division'). It can be proved ((Baccelli *et al.*, 1992*b*, §5.6)) that, for the system (4), the solution can be explicitly computed by the backward recursive equations

$$\xi(k) = \big(A\backslash\xi(k+1)\big) \wedge \big(C\backslash y(k)\big)\ , \tag{12a}$$
$$u(k) = B\backslash\xi(k)\ , \tag{12b}$$

in which, e.g.,

$$(A\backslash b)_i = \min_j (b_j - A_{ji}) \tag{13}$$

(with a careful handling of infinite values, see (Baccelli *et al.*, 1992*b*, Example 4.65)). The 'costate' ξ does not follow the forward dynamics (4) because it corresponds to transition firing dates 'at the latest', rather than 'at the earliest' possible time, as it is the rule for the forward dynamics.

Consider the following scenario: a control history $\underline{u}(\cdot)$ is first used to produce an output trajectory $y(\cdot)$; this $y(\cdot)$ is then used in (12) to compute some $\xi(\cdot)$ and a new control input $u(\cdot)$ which is of course greater than, or equal to $\underline{u}(\cdot)$; finally, this new $u(\cdot)$, when used in (4), produces some new $x(\cdot)$, but the *same* output $y(\cdot)$ as $\underline{u}(\cdot)$ does. We get the following kind of state-costate equations:

$$x(k) = Ax(k-1) \oplus B\big(B\backslash\xi(k)\big)\ ; \tag{14a}$$
$$\xi(k) = A\backslash\xi(k-1) \wedge C\backslash\big(Cx(k)\big)\ . \tag{14b}$$

One can prove the intuitively appealing fact that $\xi_i(k) - x_i(k)$ is nonnegative: this is interpreted as the 'spare time' or the 'margin' which is available at transition x_i for the firing nr. k; in other words, an exogenous event may delay this event by this spare time without preventing the future deadlines to be met. Differences such as $\xi_i(k) - x_i(k)$ emerge as diagonal elements of the matrix $P(k) = \xi(k)\not/x(k)$. In conventional system theory, for linear-quadratic problems, the costate vector ξ is related to the state vector x by $\xi = Px$, where P is a matrix obeying a Riccati equation. For the time being, no recursive equation has been found for the ratio $\xi(k)\not/x(k)$. On this and similar topics related to what we consider as the analogue of a 'second order theory' (with 'correlation matrices' having to do with in-process stocks and times spent in the system), one may refer to (Baccelli *et al.*, 1992*b*, §6.6), (Max Plus, 1991; Cohen *et al.*, 1993).

4. TOWARDS GEOMETRIC SYSTEM THEORY

4.1 *From algebra to geometry*

Vectors and rectangular matrices have already showed up in the previous developments. While *square* matrices can be given a dioid structure with two *internal* operations called 'addition' and 'multiplication', vectors, for example, can be endowed with an internal addition, but the multiplication of interest is generally that of vectors by 'scalars' belonging to a dioid. are sometimes referred to as *moduloids* or *pseudomodules* or *semimodules* nowadays, and they have received (admittedly limited) attention. It is beyond the scope of this paper to discuss even the basic (multiple) notions of *linear independence* in such structures and the associated notions of *dimensions*. A few authors have initiated some work with the aim of understanding the geometry of moduloids (Wagneur, 1991). Compared with usual vector spaces, the situation is more involved, in that two moduloids with minimal generating sets with the same cardinality need not be isomorphic (Wagneur, 1996). Indeed, elements of minimal generating families play a role analogous to extremal rays of usual polyhedral cones.

In linear systems theory, the interest of the geometric point of view has been shown e.g. by Wonham (1979). The basic notions of controllability and observability (more general than those of *structural* controllability and observability referred to at §3.2) amounts to

surjectivity, resp. injectivity, of certain linear operators. Hence images and kernels as geometric objects (more than their representatives in terms of matrices) are central. The notion of decomposition of a 'space' into a 'direct sum of subspaces' is also important. An attempt to approach this problem in the context of moduloids can be found in (Wagneur, 1994). Another point of view has been initiated in (Cohen *et al.*, 1996; Cohen *et al.*, 1997). In this approach, *residuation* theory plays a central role. Hence a brief account of this theory is given in the next subsection.

4.2 *Residuation theory in a few words*

The main purpose of residuation theory is to provide an answer to the problem of 'solving' equations in x of the form $f(x) = b$, where f is an *isotone* (i.e. order-preserving) mapping between two lattice-ordered sets which are *complete* (i.e. infinite subsets admit a *least upper bound* —`lub`, denoted $\vee$— and a *greater lower bound* —`glb`, denoted $\wedge$— which of course need not belong to the subset). The idea is to weaken the notion of 'solution' to that of 'subsolution' satisfying $f(x) \preceq b$ or to that of supersolution satisfying $f(x) \succeq b$ and to select the `lub` of these subsolutions, resp. the `glb` of these supersolutions. Which approach is adopted depends upon a 'continuity' property of f: the former approach is appropriate when f is lower-semicontinuous (l.s.c.), that is, $f\big(\bigoplus_{x\in X} x\big) = \bigoplus_{x\in X} f(x)$, for any subset X, which implies that the `lub` of subsolutions is itself a subsolution; dually, the latter approach is appropriate if f is upper-semicontinuous (`usc` — guess the definition!).

Remark 5. It should be kept in mind that if there exists a 'true' solution to the problem with *equality* (possibly nonunique), then either approach will also provide a true solution (if of course the corresponding continuity assumption is satisfied by f).

The following theorem summarizes an essential part of the story of residuation.

Theorem 6. Let f be an isotone mapping between two complete lattices $\mathcal{X}$ and $\mathcal{Y}$. The following three statements are equivalent: (1) For every b, there exists a greatest subsolution of $f(x) = b$; (2) The mapping f is `lsc` and $f(\varepsilon) = \varepsilon$ (where ε denotes the bottom element in any complete lattice); (3) There exists an isotone mapping $f^\sharp$ from $\mathcal{Y}$ to $\mathcal{X}$ such that

$$f \circ f^\sharp \preceq I \quad \text{(identity in } \mathcal{X}\text{)}, \tag{15a}$$

$$f^\sharp \circ f \succeq I \quad \text{(identity in } \mathcal{Y}\text{)}. \tag{15b}$$

Then f is said *residuated* and $f^\sharp$, which is uniquely defined by (15), and which is `usc`, is called its *residual*. In addition,

$$f \circ f^\sharp \circ f = f \; ; \quad f^\sharp \circ f \circ f^\sharp = f^\sharp \, . \tag{16}$$

Of course, an analogous theorem about dually residuated (`usc`) mappings and least supersolutions can also be stated: the dual residual is denoted $f^\flat$ and $(f^\sharp)^\flat = f$ (when f is residuated). So far, we have considered the residuals of the mappings $x \mapsto a \otimes x$ and $x \mapsto x \otimes a$, denoted $y \mapsto a \backslash y$ and $y \mapsto y / a$, resp., including the case when a is a matrix (see (13)). Indeed, there is already a rich calculus associated with residuation (see (Baccelli *et al.*, 1992*b*, §4.4)) but much remains probably to be done in this matter, including software.

As a specialization of Rem. 5 to the case when f is a $(m \times n)$-dimensional matrix A, $Ax = b$ has a solution iff $A(A\backslash b) = b$. In particular, to build a minimal generating set from a given finite generating set of m columns vectors a_i of dimension n, we have to apply the previous test for each $i = 1, \ldots, m$, with $b = a_i$ and A composed of the rest of vectors (those different from a_i and which have not yet been eliminated) and to eliminate this a_i if the test is satisfied (see e.g. (Gaubert and Max Plus, 1997) and references therein on this topic of 'weak bases').

4.3 *Projection on image parallel to kernel*

With usual vector spaces $\mathcal{U}, \mathcal{X}, \mathcal{Y}$, let $B : \mathcal{U} \to \mathcal{X}$ and $C : \mathcal{X} \to \mathcal{Y}$ be two linear operators. The projector Π_B^C onto $\operatorname{im} B$ parallel to $\ker C$ exists and is well defined iff $\mathcal{X}$ is the *direct sum* of $\operatorname{im} B$ and $\ker C$ (that is, $\mathcal{X} = \operatorname{im} B + \ker C$ and $\operatorname{im} B \cap \ker C = \{0\}$); moreover, if B is injective and C is surjective [4], then

$$\Pi_B^C = B(CB)^{-1}C \, . \tag{17}$$

With semimodules, keeping the definition $\ker C = \{x \mid C(x) = \varepsilon\}$ does not seem to provide a very interesting notion. This motivates the following set-theoretic definition.

Definition 7. (Kernel). Let $C : \mathcal{X} \to \mathcal{Y}$ denote any mapping between moduloids. We call *kernel* of C (denoted $\ker C$), the *equivalence relation* over $\mathcal{X}$ defined as:

$$x \overset{\ker C}{\sim} x' \Leftrightarrow C(x) = C(x') \Leftrightarrow x \in C^{-1}\big(C(x')\big) \, . \tag{18}$$

Definition 8. (Projection). Let $C : \mathcal{X} \to \mathcal{Y}$ and $B : \mathcal{U} \to \mathcal{X}$ denote any mappings between moduloids. For any $x \in \mathcal{X}$, we call *projection of x onto* $\operatorname{im} B$ *parallel to* $\ker C$ any $\xi \in \operatorname{im} B$ such that $\xi \overset{\ker C}{\sim} x$.

The questions of existence and uniqueness of the projection for given operators B and C is studied in (Cohen *et al.*, 1996) for residuated (or dually residuated) operators and in (Cohen *et al.*, 1997) for linear operators, together with explicit expressions for the

[4] The subspaces $\operatorname{im} B$ and $\ker C$ are important, not the operators B and C for which a certain flexibility exists.

projection. A brief informal summary is given hereafter. Let first assume that B and C are residuated and introduce

$$\Pi_B^C = B \circ (C \circ B)^\sharp \circ C \tag{19}$$

(to be compared with (17)).

- *Existence* of projections for all x is equivalent to the condition $C = C \circ \Pi_B^C$ (saying that $\xi = \Pi_B^C(x)$ is in the same class as $x \mod \ker C$), and also to the condition $\operatorname{im} C = \operatorname{im}(C \circ B)$.
- *Uniqueness* is equivalent to the condition $B = \Pi_B^C \circ B$ (saying that any $x \in \operatorname{im} B$ remains invariant by Π_B^C), and also to the condition $\ker B = \ker(C \circ B)$.

With matrices over, say, the max-plus algebra (they are also residuated operators), when existence and uniqueness are granted, the expression (19) of the projector (which is easily proved to be linear in this situation) becomes:

$$\Pi_B^C = \big(B \not/ (CB)\big) C = B\big((CB) \backslash\!\!\not\; C\big) \; . \tag{20}$$

Note that e.g. $B \not/ (CB)$ is, by definition (residuation in the matrix algebra), a matrix, and the above expression is understood as a product of matrices (which themselves arise from residuation of multiplication in sets of matrices).

Examples are easy to figure out in two-dimensional max-plus semimodules but some more general phenomena require at least dimension 3 to show up. In making drawings for homogeneous residuated operators (in particular linear operators), one must keep in mind a few facts.

- The image of an operator B such that $B(\alpha x) = \alpha B(x)$ for all vectors x and scalars α is invariant by translation along the first diagonal, since αx means adding (in the conventional sense) the same constant α to all coordinates.
- Also, for C with the same property, if $x \overset{\ker C}{\sim} x'$, then $\alpha x \overset{\ker C}{\sim} \alpha x'$, that is, equivalence classes can be derived from each other by translations along the first diagonal.
- Finally, C is injective over $\operatorname{im} C^\sharp$ (this is a consequence of (15b)), that is, equivalence classes intersect $\operatorname{im} C^\sharp$ at a single point.

Consider Fig. 5 in which three situations with (2×2)-dimensional matrices are represented. The grey area

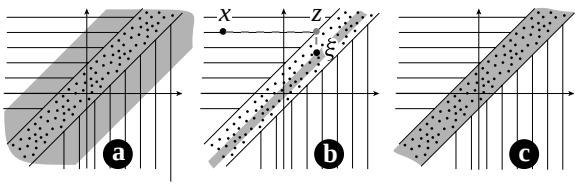

Fig. 5. Existence and uniqueness of projection

is that of $\operatorname{im} B$, the dotted area is that of the 'interior' of $\operatorname{im} C^\sharp$ in which equivalence classes of $\ker C$ are singletons, and the horizontal and vertical half-lines represent other equivalence classes in the rest of the plane. Hence not all equivalence classes have the same topology. Part a of the figure displays a case with existence of projection but no uniqueness everywhere (some classes crosses the grey area in more than one point): part b represents the case with uniqueness but no existence everywhere (some classes do not reach the grey area); part c is the case with existence and uniqueness everywhere [5]. One may consider the last situation as that of 'direct sum of $\ker C$ and $\operatorname{im} B$', but in an unusual sense (also different from Wagneur's (1994) meaning): in (Cohen *et al.*, 1997), the terminology 'direct factors' for $\operatorname{im} B$ and $\ker C$ is used.

In the same paper, it is shown that the image or kernel associated with a matrix B need not admit a direct factor (unlike in classical linear vector spaces), and that a necessary and sufficient condition for this to hold true is that B is *regular*, meaning that there exists a *g-inverse* $B^\dagger$ which satisfies, by definition, $BB^\dagger B = B$. However, dimension 3 at least is required to show nonregular matrices.

More generally, for residuated mappings, even out of the case of existence and uniqueness, Π_B^C as given by (19) has a precise meaning: when applied to x, it provides the greatest element ξ in $\operatorname{im} B$ which is 'subequivalent' to $x \mod \ker C$, that is, such that $C\xi \preceq Cx$. The projector Π_B^C can be decomposed in two moves (see Fig. 5b) once written as

$$\Pi_B^C = B \circ B^\sharp \circ C^\sharp \circ C \; . \tag{21}$$

First, $z = C^\sharp \circ C(x)$ is the greatest element in the equivalence class of $x \mod \ker C$; then, $\xi = B \circ B^\sharp(z)$ is the greatest element in $\operatorname{im} B$ which is less than z. Notice that if x is already in $\operatorname{im} B$, then ξ is truly equivalent to $x \mod \ker C$ (for those x, existence is granted).

When B and C are matrices, it is an open problem to give necessary and sufficient conditions for Π_B^C to be linear: a priori, this operator involves a mix of max, min and $+$ operations; the case when $\operatorname{im} B$ and $\ker C$ are direct factors has already been identified as a case when this projector is linear, but it is not the only situation when linearity is preserved. Obviously, this issue is important for system theory since the notion of system aggregation and of reduced —not to say, minimal— state space representation basically involves such projectors: starting from a linear system, it is desirable to get a reduced system which is still linear in the same algebra.

[5] One can even show a fourth situation when neither existence nor uniqueness is ensured everywhere: this is the case when $\operatorname{im} B$ and $\operatorname{im} C^\sharp$ are not included in each other.

4.4 *Applications in system theory*

We return to systems described by (4). The state values which are reachable from the canonical initial condition ε are of course those in the image of the *reachability* (or *controllability*) *matrix* [6]:

$$\mathcal{R} = \begin{pmatrix} B & AB & A^2B & \dots \end{pmatrix} . \tag{22}$$

On the other hand, two state values which are equivalent modulo $\ker \mathcal{O}$, where $\mathcal{O}$ is the *observability matrix*

$$\mathcal{O} = \begin{pmatrix} C^\top & A^\top C^\top & (A^\top)^2 C^\top & \dots \end{pmatrix}^\top , \tag{23}$$

($^\top$ stands for transposition) can be merged from the input-output point of view. According to (Eilenberg, 1974, Prop. 5.2 and Th. 5.6) [7], from the module (in fact essentially set-theoretic) point of view, a minimal state 'space' is

$$\Xi = \operatorname{im} \mathcal{R} / \ker \mathcal{O} , \tag{24}$$

that is, the quotient of $\operatorname{im} \mathcal{R}$ (which is a semimodule) by the compatible equivalence relation (or congruence) $\ker \mathcal{O}$ which preserves the semimodule structure. By comparison with realization theory over fields, the difficulty is that the 'minimal' moduloid $\Xi = \operatorname{im} \mathcal{R} / \ker \mathcal{O}$, which is isomorphic to the image of the Hankel matrix of the system (Fliess, 1975), is in general not free. The following questions must then be addressed regarding this abstract construction which, by construction, retains a completely reachable and observable state 'space'.

- Can the abstract semimodule Ξ be given a more concrete representation (or, otherwise stated, what is the state vector corresponding to this minimal 'set-theoretic' representation)?
- When does minimality from the 'set-theoretic' point of view imply minimality from the computational point of view (that is, for the number of coordinates of some state vector which allows one to write down an internal representation of the form (4))?
- Is there a way to relate this minimal dimensionality with that suggested by the transfer function computation (although this problem of minimal realization of transfers is itself an open problem), that is, to relate the geometric and the algebraic points of view?

At this stage, there is, to the best of our knowledge, no definite answers to those questions. Observations made on examples suggest that the situation is not as simple as in classical linear system theory, but perhaps not so hopeless. We use the rest of available space to give a few unpublished results (without proof) and discuss some further examples.

[6] Unlike in classical linear algebra, it may be necessary to keep all powers of A up to infinity to get the whole image of this matrix. The same remark applies to the kernel of $\mathcal{O}$ to come.

[7] The treatment of (Eilenberg, 1974), which is in the case of rings and modules, can be readily extended to semirings and semimodules.

In order to find a concrete representation of elements of Ξ, we consider the (canonical) greatest representative ξ in each equivalence class of x (which we suppose to be a reachable state, i.e $x \in \operatorname{im} \mathcal{R}$): it is given by $\xi = \Pi^{\mathcal{O}}_{\mathcal{R}}(x)$.

Theorem 9. If the trajectory $x(\cdot)$ follows the dynamics (4) and is issued from the initial condition ε (or from any reachable initial state), then $\xi(k) = \Pi^{\mathcal{O}}_{\mathcal{R}}\big(x(k)\big)$ follows the (a priori nonlinear) dynamics

$$\xi(k) = \Pi^{\mathcal{O}}_{\mathcal{R}}\big(A\xi(k-1) \oplus Bu(k)\big) , \tag{25a}$$
$$y(k) = C\xi(k) , \tag{25b}$$

and it produces exactly the same output trajectory $y(\cdot)$ as $x(\cdot)$. Hence, it is another realization of the input-output transfer matrix.

The proof of this theorem will appear in a forthcoming paper. The advantage of this result is that the state ξ lives in a minimal set in terms of set inclusion. Its a priori drawback is that the dynamics is potentially nonlinear (unless $\Pi^{\mathcal{O}}_{\mathcal{R}}$ is linear, at least over reachable states) and it is unclear that the dynamics can be written in a smaller dimensional semimodule (for the time being, ξ has the same dimension as x). Examples show that it may happen that ξ lives in a set with many 'extremal points', which is no good sign for minimizing the dimension of the representation. Nevertheless, for all examples worked out, it seems that this set intuitively provides an indication of the minimal dimension needed to realize the transfer (in that, a surface, even with many ridges and corners, is a two-dimensional variety in $\mathbb{R}^3$, and a broken line is a one-dimensional one).

Before showing examples, observe that, again, $\Pi^{\mathcal{O}}_{\mathcal{R}}$ can be viewed as the composition of $\Pi^{\mathcal{O}} = \mathcal{O}^\sharp \circ \mathcal{O}$ and $\Pi_{\mathcal{R}} = \mathcal{R} \circ \mathcal{R}^\sharp$. These two projectors satisfy the following (kind of Lyapunov) implicit equations:

$$\Pi^{\mathcal{O}} = A^\sharp \circ \Pi^{\mathcal{O}} \circ A \wedge C^\sharp \circ C , \tag{26a}$$
$$\Pi_{\mathcal{R}} = A \circ \Pi_{\mathcal{R}} \circ A^\sharp \oplus B \circ B^\sharp . \tag{26b}$$

From these equations, an interpretation of the state $\xi(k)$ can be given: when applied to $x(k)$, $\Pi^{\mathcal{O}}$ first looks for the greatest state value at stage k which would generates future outputs not exceeding those contributed by $x(k)$ (independently of the contribution of future inputs which are yet unknown and whose effects will be superimposed by linearity); then, since this greatest compatible state value may not be a reachable state, $\Pi_{\mathcal{R}}$ finds its best approximation from below which is reachable. In light of this interpretation, we are not far from the computation of (12)–(14), except that we are here in a causal situation when future inputs are unknown.

Again, an important issue is: when is $\Pi^{\mathcal{O}}_{\mathcal{R}}$ a linear operator (with the consequence of the dynamics (25a) being then also linear)? Although some sufficient con-

ditions are known, we leave this subject as an open issue.

4.5 *Working out an example*

Consider the matrices A, B, C given in (5). It turns out that the computation of (22) can be stopped at the power 1 of A (that is, the column rank of $\mathcal{R}$ — defined as the cardinality of a minimal generating set of the column space of $\mathcal{R}$ — is 2); the computation of (23) can be stopped at the power 2 (the row rank of $\mathcal{O}$ is 3). The calculations of $\Pi^{\mathcal{O}}$, $\Pi_{\mathcal{R}}$, and finally of $\Pi^{\mathcal{O}}_{\mathcal{R}} = \Pi_{\mathcal{R}} \circ \Pi^{\mathcal{O}}$ yield nonlinear expressions (however, it turns out that $\Pi^{\mathcal{O}}_{\mathcal{R}}$ is max-plus linear when restricted to $\operatorname{im}\mathcal{R}$). Explicitly,

$$\Pi^{\mathcal{O}}_{\mathcal{R}} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \big((2x_1 \oplus 1x_3) \wedge x_2\big) \begin{pmatrix} -2 \\ e \\ -1 \end{pmatrix} ,$$

which reveals that $\operatorname{im}\Pi^{\mathcal{O}}_{\mathcal{R}}$ is parametrized by a scalar! In addition, this image is the eigenspace of matrix A (but no general conclusion should be derived from this last observation which is certainly due to the dimension 1 of the minimal realization). The remarkable fact is that this geometrical dimension of $\operatorname{im}\Pi^{\mathcal{O}}_{\mathcal{R}}$ is the same as the order of the realization derived from the transfer function calculation and shown at Fig. 2.

The same calculations can be conducted with the variant of the TEG of Fig. 1 already used at §2.4, which led to the transfer function shown in (9), and to the two-dimensional realization shown at the right-hand side of Fig. 4. This variant differs only by matrix B which is equal to $\begin{pmatrix} \varepsilon & 1 & \varepsilon \end{pmatrix}^{\top}$. Hence, only $\mathcal{R}$ need be calculated again. Now, $\mathcal{R}$ has a column rank equal to 3, and $\Pi^{\mathcal{O}}_{\mathcal{R}}$ is nonlinear again (even on $\operatorname{im}\mathcal{R}$). Explicitly,

$$\Pi^{\mathcal{O}}_{\mathcal{R}} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} (-2)\alpha \\ \alpha \\ \beta \end{pmatrix}$$

with $\alpha = \big(2x_1 \oplus 1x_3\big) \wedge x_2$, $\beta = \big(1x_1 \wedge (-1)x_2\big) \oplus x_3$. Therefore, there exists a two-dimensional nonlinear parametrization of $\operatorname{im}\Pi^{\mathcal{O}}_{\mathcal{R}}$ in accordance with the minimal order found for the transfer function realization.

5. CONCLUSION

In the few lines left, let us insist on applications which did not receive enough attention in this paper (because of the lack of space) and also in the literature in general (with of course a few exceptions, see e.g. (Braker, 1993) in transportation or (Cohen *et al.*, 1985) in manufacturing), but which deserve more interest for themselves, and also for their potential to suggest new theoretical questions. On the theoretical side, identification end adaptive control, as initiated by Menguy (1997), are also promising directions of investigation.

6. REFERENCES

Baccelli, F., G. Cohen and B. Gaujal (1992*a*). Recursive equations and basic properties of timed Petri nets. *J. of Discrete Event Dynamic Systems* **1**(4), 415–439.

Baccelli, F., G. Cohen, G.J. Olsder and J.-P. Quadrat (1992*b*). *Synchronization and Linearity - An Algebra for Discrete Event Systems*. Wiley. New York.

Braker, H. (1993). Algorithms and applications in timed discrete event systems. Phd thesis. Delft University of Technology, the Netherlands.

Cohen, G. (1994). Dioids and discrete event systems. In: *Proc. 11th Int. Conf. on Anal. and Optim. of Systems, Sophia-Antipolis, France* (G. Cohen and J.-P. Quadrat, Eds.). Vol. 199 of *Lect. Notes in Contr. and Inform. Sc.*. Springer-Verlag. Berlin. pp. 223–236.

Cohen, G., D. Dubois, J.-P. Quadrat and M. Viot (1985). A linear system-theoretic view of discrete event processes and its use for performance evaluation in manufacturing. *IEEE Trans. on Aut. Cont.* **AC-30**(3), 210–220.

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

Cohen, G., S. Gaubert and J.-P. Quadrat (1993). From first to second-order theory of linear discrete event systems. In: *Proc. 12th IFAC World Congress, Sydney, Australia*.

Cohen, G., S. Gaubert and J.-P. Quadrat (1996). Kernels, images and projections in dioids. In: *Proc. Worksh. on Disc. Ev. Syst., Edinburgh, Scotland*.

Cohen, G., S. Gaubert and J.-P. Quadrat (1997). Linear projectors in the max-plus algebra. In: *Proc. 5th IEEE Med. Conf. on Cont. and Syst., Paphos, Cyprus*.

Cohen, G., S. Gaubert and J.-P. Quadrat (1998). Algebraic system analysis of timed Petri nets. In: *Idempotency* (J. Gunawardena, Ed.). pp. 145–170. Coll. of the Isaac Newton Inst.. Cambridge University Press. Cambridge, England.

Cohen, G., S. Gaubert, R. Nikoukhah and J.-P. Quadrat (1989*b*). Convex analysis and spectral analysis of timed event graphs. In: *Proc. 28th Conf. Dec. and Cont., Tampa, Florida*.

Cottenceau, B., L. Hardouin, J.-L. Boimond and J.-L. Ferrier (1998). Synthesis of greatest linear feedback for TEG in dioid. *IEEE Trans. on Aut. Cont.* to appear.

Di Mascolo, M. (1990). Modélisation et évaluation de performances de systèmes de production gérés en kanban. PhD thesis. INPG. Grenoble, France.

Eilenberg, S. (1974). *Automata, languages and machines*. Vol. A. Academic Press. New York.

Fliess, M. (1975). Matrices de Hankel. *J. Math. Pures. Appl.* **15**, 161–186.

Gaubert, S. (1992). Théorie des systèmes linéaires dans les dioïdes. Phd thesis. École des Mines de Paris, France.

Gaubert, S. (1995). Resource optimization and $(\min, +)$ spectral theory. *IEEE Trans. on Aut. Cont.*

Gaubert, S. and J. Mairesse (1997). Modelling and analysis of timed Petri nets using heaps of pieces. To appear in IEEE Trans. on Aut. Contr., abridged version in the Proceedings of the ECC'97, Bruxelles, 1997.

Gaubert, S. and Max Plus (1997). Methods and applications of (max,+) linear algebra. In: *14th Symp. on Theoretical Aspects of Computer Science, Lübeck, Germany, 27 Feb.-1 Mar. 1997* (R. Reischuk and M. Morvan, Eds.). Vol. 500 of *Lect. Notes in Comp. Sc.*. Springer-Verlag. Berlin. pp. 261–282.

Gaubert, S., P. Butkovič and R. Cuninghame-Green (1998). Minimal $(\max, +)$ realization of convex sequences. *SIAM J. Cont. Optim.* **36**(1), 137–147.

Max Plus (1991). Second order theory of min-linear systems and its application to discrete event systems. In: *Proc. 30th Conf. Dec. and Cont., Brighton, England*.

Menguy, É. (1997). Contribution à la commande des systèmes linéaires dans les dioïdes. Phd thesis. ISTIA, Université d'Angers, France.

Murata, T. (1989). Petri nets: properties, analysis and applications. *Proc. of the IEEE* **77**, 541–580.

Quadrat, J.-P. and Max Plus (1995). Max-plus algebra and applications to system theory and optimal control. In: *Int. Conf. of Mathematicians 1994, Zurich*. Birkaüser. Basel. pp. 1502–1511.

Ramadge, P.J.G. and W.M. Wonham (1989). The control of discrete event systems. *Proc. of the IEEE* **77**(1), 81–97.

Wagneur, É. (1991). Moduloids and pseudomodules. 1. dimension theory. *Discrete Math.* **98**, 57–73.

Wagneur, É. (1994). Subdirect sum decomposition of finite dimensional pseudomodules. In: *Proc. 11th Int. Conf. on Anal. and Optim. of Systems, Sophia-Antipolis, France* (G. Cohen and J.-P. Quadrat, Eds.). Vol. 199 of *Lect. Notes in Contr. and Inform. Sc.*. Springer-Verlag. Berlin. pp. 322–328.

Wagneur, É. (1996). Torsion matrices in the max-algebra. In: *Proc. Worksh. on Disc. Ev. Syst., Edinburgh, Scotland*. pp. 165–168.

Wonham, W.M. (1979). *Linear multivariable control: a geometric approach*. Springer-Verlag. Berlin. 2nd ed.