UNIVERSIDAD DE LOS ANDES
FACULTAD DE CIENCIAS

NOTAS DE MATEMATICA

COURSE NOTES: MAX ALGEBRA APPROACH TO DISCRETE EVENT SYSTEMS

POR

GEERT JAN OLSDER

DEPARTAMENTO DE MATEMATICA

Universidad de los Andes

Facultad de Ciencias

Departamento de Matemática

Course Notes: Max Algebra Approach to Discrete Event Systems

Geert Jan Olsder

Notas de Matemática

Serie: Pre-Print

No. 191

Mérida - Venezuela

1999

Notas de Matemática, No. 191
Mérida, March 1999.

Course Notes: Max Algebra Approach to Discrete Event Systems

Geert Jan Olsder
E-mail: g.j.olsder@math.tudelft.nl

Contents

1 Motivation: an Example

The theory of Discrete Event Systems (DES) is a research area of current vitality. The development of this theory is largely stimulated by discovering

general principles which are (or are hoped to be) useful to a wide range of application domains. In particular, technological and/or 'man-made' manufacturing systems, communication networks, transportation systems, and logistic systems, all fall within the class of DES. One of the key features that characterize these systems is that . their dynamics are *event-driven* as opposed to *time-driven*, i.e., the behavior of a DES is governed only by occurrences of different types of events over time rather than by ticks of a clock.

In the literature, one can distinguish three main fields of activities, all claiming and using the phrase 'discrete event (dynamic) system'. In this paper only one of the three fields will be discussed. For the other two, the reader is referred to [5].

We will consider a closed network of railway connections and study questions related to the design of time tables according to which the trains run within this network. Specifically, we will study the intercity network of the Netherlands. This network, given in Figure 1, consists of 11 lines (routes along which the trains run in both directions). An example of such a line is the line with number 10 which starts (respectively terminates) at Amsterdam and terminates (respectively starts) at Vlissingen. There are no circular lines, but inclusion of circular lines into the model would not constitute any difficulties. We will not consider international connections: it is assumed in our model that trains arriving at the border return direction and continue in opposite direction (this is what actually happens with some trains at some border stations). Hence the name 'closed' network. On each line, both directions, there is a fixed number of trains which return at both endpoints in the opposite direction. Thus each line can be considered as a closed circuit (or as a circular line). The travelling times between stations is given and assumed to be deterministic.

Suppose there is no time table. A very simple policy for trains to run would be to drop and pick up passengers at stations and then directly continue. This is not very realistic since one would like trains to remain at a station for some time such as to wait for other arriving trains and allow change overs. Now suppose that realistic constraints are included which state which trains have to wait for each other, at each station, such as to allow change overs before the trains can continue on their respective lines.

Without an apriori time table, and with the constraints of trains waiting for each other, how fast can the system operate? The actual time table is based on a half hour schedule and hence from this fact of real life we know that without time table the system must be 'operable' in a time of

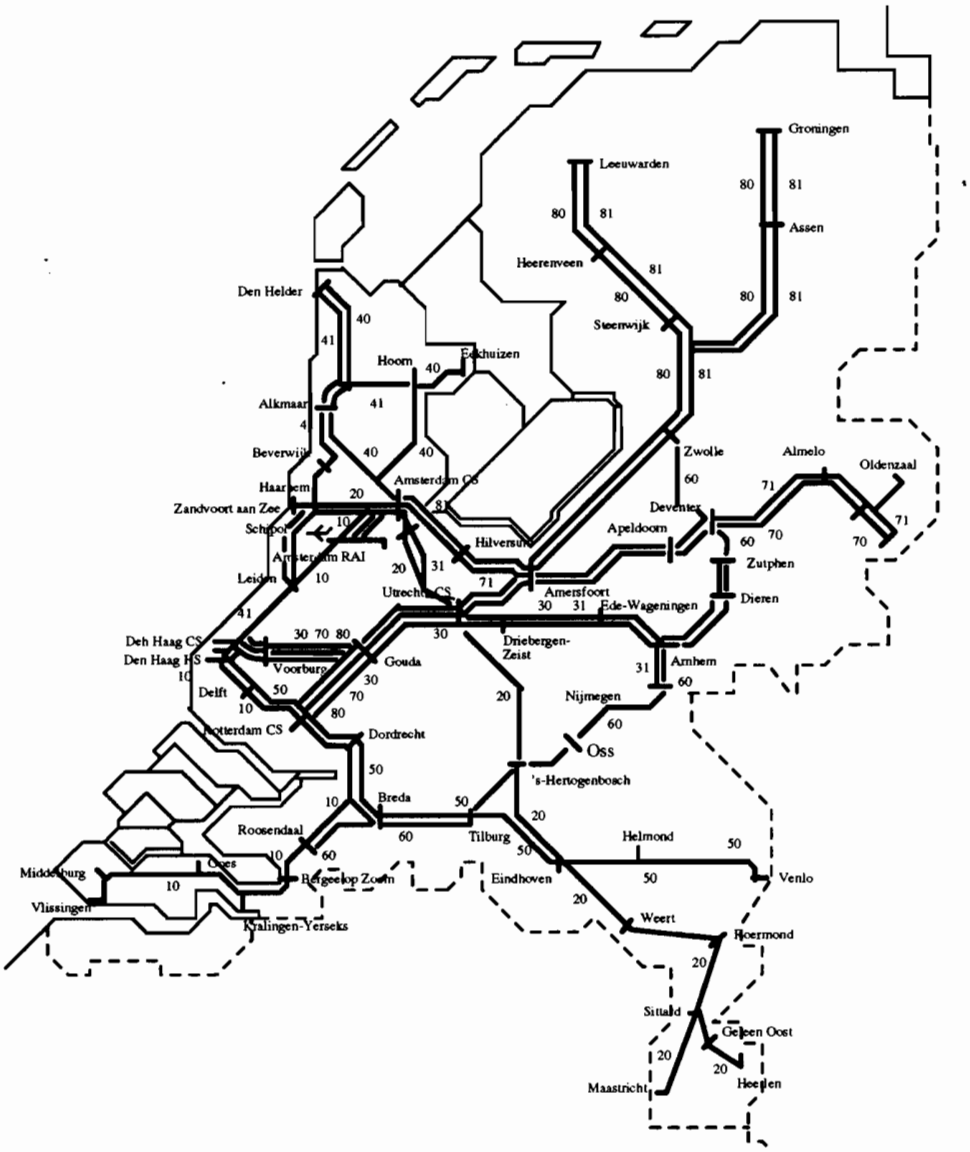

Figure 1: The Dutch intercity network

at most half an hour (this time will be indicated by λ which turns out to be equal to about 27 minutes). The difference between the actual half hour and the theoretical λ minutes is the flexibility in the system which causes propagation of possible delays to disappear in finite time. Various questions can be posed now, some of which are:

- How do perturbations progagate through the system and how long does it take before they have completely disappeared?
- If one would add an extra five minutes to all changeover times, is it still possible to design a time table based on a half hour service? (answer: yes).
- What are the crucial parts of the system which determine the minimum operation time of λ minutes? Suppose one could add some extra trains to the system, on which lines should they be set in such as to reduce this minimum operation time as much as possible?
- With the minimum operation time schedule, is it possible to have a 'regular' time table? With regular is meant here that if a train leaves station A in the direction of station B, the next train in the same direction will leave exactly λ minutes later.
- The lines along which the trains run were supposed to be given. Is it possible to design an 'optimal line structure'?

Answers to most of these questions will be given in section 2, in Example 2.2.

2 The Max-Plus Algebra Approach to DES

This section starts with the modelling of an example, in terms of an extremely simple railway network, in subsection 2.1, and subsequently this modelling is given a more abstract setting in subsections 2.2 and 2.3. In the latter subsection the intercity network example as discussed above , is revisited. Subsection 2.4 deals with the so-called γ-transform, which is the max-plus algebra equivalent of the z-transform in the conventional theory of linear difference equations. In subsection 3 various extensions are briefly considered.

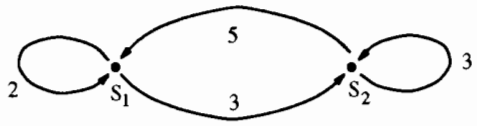

Figure 2: The two stations example

2.1 The Stage for the Max-Plus Approach

In a metropolitan area there are two railway stations, S_1 and S_2, which are interconnected by a railway system as indicated in Figure 2. This railway system consists of an innercircle and of two outer circles. The trains on these outer circles deliver and pick up passengers in the suburbs. The stations in the suburbs have not been drawn since they do not play any role in the model to be formulated.

Suppose there are four trains (two at each station) and they leave the stations at time 0, one along each track. They reach the other (or the same) station after a certain time which is indicated in the figure. The arriving trains at a station have to wait for each other such as to allow the passengers to change trains.

Figure 2 can easily be redrawn as an event graph. The two stations are transitions and in the four railway tracks one can put a place. If a train is running along a track, one puts a token in the place corresponding to this track.

Suppose that there is no time table and that the trains leave directly after the change over of the passengers at the stations and that the time needed for change overs has been incorporated in the travelling time. This 'travelling time' was called 'holding time' in the theory of Petri nets (see subsection 3.1). If this process of departing and arriving trains is continued, the departure time $x_i(k+1)$ for the $k+1$-st departure at station S_i satisfies

$$\begin{array}{rcl} x_1(k+1) & = & \max(x_1(k)+2,\ x_2(k)+5), \\ x_2(k+1) & = & \max(x_1(k)+3,\ x_2(k)+3), \end{array} \tag{1}$$

for $k = 0, 1, 2, \ldots$.

With $x_1 = 0,\ x_2 = 0$, the evolution of equations (1) becomes

$$\underset{x(0)}{\begin{pmatrix} 0 \\ 0 \end{pmatrix}} \to \underset{x(1)}{\begin{pmatrix} 5 \\ 3 \end{pmatrix}} \to \underset{x(2)}{\begin{pmatrix} 8 \\ 8 \end{pmatrix}} \to \underset{x(3)}{\begin{pmatrix} 13 \\ 11 \end{pmatrix}} \to \underset{x(4)}{\begin{pmatrix} 16 \\ 16 \end{pmatrix}} \to \cdots$$

This pattern of departure times shows a periodic solution superimposed on a linear drift, the 'period' equals 2 and the average time between two subsequent departures is 4. From a time table point of view (time tables must be as 'regular' as possible), it is better to start with the initial departures $x_1 = 1$, $x_2 = 0$, since then the evolution becomes

$$\underset{x(0)}{\begin{pmatrix} 1 \\ 0 \end{pmatrix}} \rightarrow \underset{x(1)}{\begin{pmatrix} 5 \\ 4 \end{pmatrix}} \rightarrow \underset{x(2)}{\begin{pmatrix} 9 \\ 8 \end{pmatrix}} \rightarrow \underset{x(3)}{\begin{pmatrix} 13 \\ 12 \end{pmatrix}} \rightarrow \cdots$$

where the interdeparture time is now exactly 4 at each station and thus the departure times are very regular (they have 'period' 1). By trial and error it turns out that, whatever the initial condition, after possibly a short transient period of time, a periodic behavior of either period 1 or 2 is obtained with (average) interdeparture times 4. A solution with an (average) departure time smaller than 4 is not possible, since for a train to go around in the innercircle costs $3+5 = 8$ time units. There are two trains on the innercircle and therefore the (average) interdeparture time is limited from below by $8/2 = 4$.

With the above sketched railway system and trains it is not possible to design a time table with interdeparture times smaller than 4. If one wants a faster time table, one must change the problem. To this end, let us add a train on the innercircle, such that three trains will run along this circle all the time. Suppose that initially this extra train is situated at station S_1. The equations for the departure times now become

$$x_1(k+1) = \max(x_1(k)+2,\ x_2(k)+5), \tag{2}$$
$$x_2(k+1) = \max(x_1(k-1)+3,\ x_2(k)+3), \tag{3}$$

which can be rewritten as a set of first order equations as

$$\begin{aligned} x_1(k+1) &= \max(x_1(k)+2,\ x_2(k)+5), \\ x_2(k+1) &= \max(x_3(k)+3,\ x_2(k)+3), \\ x_3(k+1) &= x_1(k). \end{aligned} \tag{4}$$

With the initial condition $x_1 = 0$, $x_2 = 0$, $x_3 = 0$ the evolution of the latter set of equations becomes

$$\underset{x(0)}{\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}} \rightarrow \underset{x(1)}{\begin{pmatrix} 5 \\ 3 \\ 0 \end{pmatrix}} \rightarrow \underset{x(2)}{\begin{pmatrix} 8 \\ 6 \\ 5 \end{pmatrix}} \rightarrow \underset{x(3)}{\begin{pmatrix} 11 \\ 9 \\ 8 \end{pmatrix}} \rightarrow \underset{x(4)}{\begin{pmatrix} 14 \\ 12 \\ 11 \end{pmatrix}} \rightarrow \cdots$$

which shows, after a transient part, a regular behavior of 'period' 1 with interdeparture times 3. This interdeparture time is caused by the outerloop at station S_2; on this loop there is one train which needs 3 time units to travel around. The innerloop is not the bottleneck anymore: this innerloop itself would lead to a lower limit of 8/3 (travelling time of the loop divided by the number of trains on this loop). In order to lower the interdeparture times even more, one should add the next extra train to the outerloop of S_2. If we do so, the equations for the departure times become

$$\begin{aligned} x_1(k+1) &= \max(x_1(k)+2,\ x_2(k)+5), \\ x_2(k+1) &= \max(x_3(k)+3,\ x_2(k-1)+3), \\ x_3(k+1) &= x_1(k), \end{aligned} \tag{5}$$

which can be rewritten as a set of first order equations as

$$\begin{aligned} x_1(k+1) &= \max(x_1(k)+2,\ x_2(k)+5), \\ x_2(k+1) &= \max(x_3(k)+3,\ x_4(k)+3), \\ x_3(k+1) &= x_1(k), \\ x_4(k+1) &= x_2(k). \end{aligned} \tag{6}$$

If we start again with zero initial conditions, the solution becomes

$$\cdots \rightarrow \underset{x(3)}{\begin{pmatrix} 10 \\ 8 \\ 8 \\ 3 \end{pmatrix}} \rightarrow \underset{x(4)}{\begin{pmatrix} 13 \\ 11 \\ 10 \\ 8 \end{pmatrix}} \rightarrow \underset{x(5)}{\begin{pmatrix} 16 \\ 13 \\ 13 \\ 11 \end{pmatrix}} \rightarrow \underset{x(6)}{\begin{pmatrix} 18 \\ 16 \\ 16 \\ 13 \end{pmatrix}} \rightarrow \underset{x(7)}{\begin{pmatrix} 21 \\ 19 \\ 18 \\ 16 \end{pmatrix}} \rightarrow \cdots \tag{7}$$

This solution has a 'period' 3 ($x_i(k+3) = x_i(k)+8$, $k \geq 4$) and the average interdeparture time is 8/3, which is caused by the innercircle. Another solution results, with the same interdeparture time but with 'period' 1, if one starts with the initial condition $x_1(0) = 5$, $x_2(0) = 8/3$, $x_3(0) = 7/3$, $x_4 = 0$. One then has $x_i(k+1) = x_i(k) + 8/3$, $i = 1, 2, 3, 4$, and $k = 0, 1, \ldots$.

2.2 Formalization

The basic form of the systems we will study is

$$\begin{aligned} x_i(k+1) &= \max(a_{i1} + x_1(k), a_{i2} + x_2(k), \ldots, a_{in} + x_n(k)) \\ &= \max_j(a_{ij} + x_j(k)), \quad i = 1, \ldots, n\ . \end{aligned} \tag{8}$$

It is common practice to change the notation somewhat. Addition $+$ will be written as $\otimes$ and max will be written as $\oplus$. This change of notation makes the resemblance with conventional linear difference systems visible:

$$x_i(k+1) = \bigoplus_j (a_{ij} \otimes x_j(k)), \quad i = 1, \ldots, n \ , \tag{9}$$

which in vector notation will be written as

$$x(k+1) = A \otimes x(k). \tag{10}$$

Of the latter equation one speaks as a linear (difference) equation in the max-plus algebra, this in clear analogy with linear difference equations in the conventional, 'plus-times', algebra. If it is clear from the context that the underlying algebra is the max-plus one, one even writes $x(k+1) = Ax(k)$ for (10). If the initial condition for (10) is $x(0) = x_0$, then

$$x(1) = A \otimes x_0 \ ,$$

$$x(2) = A \otimes x(1) = A \otimes (A \otimes x_0) = (A \otimes A) \otimes x_0 = A^2 \otimes x_0 \ .$$

It can be shown that indeed $A \otimes (A \otimes x_0) = (A \otimes A) \otimes x_0$. For the example given above it is easy to check this by hand. Instead of $A \otimes A$ we simply write A^2. We get, for the general case,

$$x(k) = \underbrace{(A \otimes A \otimes \cdots \otimes A)}_{\text{k times}} \otimes x_0 = A^k \otimes x_0 \ .$$

The matrices A^2, A^3,..., can be calculated directly. Let us consider the A-matrix of (1),

$$A = \begin{pmatrix} 2 & 5 \\ 3 & 3 \end{pmatrix}$$

then

$$A^2 = \begin{pmatrix} \max(2+2, 5+3) & \max(2+5, 5+3) \\ \max(3+2, 3+3) & \max(3+5, 3+3) \end{pmatrix} = \begin{pmatrix} 8 & 8 \\ 6 & 8 \end{pmatrix} \ .$$

In general

$$(A^2)_{ij} = \bigoplus_l a_{il} \otimes a_{lj} = \max_l (a_{il} + a_{lj}) \ . \tag{11}$$

In terms of the railway example, the quantity $(A^2)_{ij}$ can be interpreted as the maximum (with respect to l) of all connections from station S_j via station

S_l to station S_i. One speaks of paths of length two between the stations S_j and S_i. In graph-theory terminology, the stations are called nodes and the tracks between stations are called arcs. More generally, $(A^k)_{ij}$ denotes the maximum of all paths of length k, starting at node j and ending at node i.

In many networks such as a railway net there will not be an arc from each node to each other node. If there is no arc from node S_j to node S_i then the behaviour of node S_i is not directly influenced by that of node S_j. In such a situation it is useful to consider the element a_{ij} to be equal to $-\infty$. In (8) a term $-\infty + x_j(k)$ does not influence $x_i(k+1)$ as long as $x_j(k)$ is finite. The number $-\infty$ will occur frequently in the sequel and it will be indicated by ε.

Linear systems in the max-plus algebra with inputs and outputs are given by

$$\left.\begin{array}{rcl} x(k+1) & = & Ax(k) \oplus Bu(k) \ , \\ y(k) & = & Cx(k) \ , \end{array}\right\} \tag{12}$$

which is short-hand notation for

$$\begin{array}{rcll} x_i(k+1) & = & \max(a_{i1} + x_1(k), \ldots, a_{in} + x_n(k), & \\ & & \quad b_{i1} + u_1(k), \ldots, b_{im} + u_m(k)), & i = 1, \ldots, n \ ; \\ y_i(k) & = & \max(c_{i1} + x_1(k), \ldots, c_{in} + x_n(k)), & i = 1, \ldots, p \ . \end{array}$$

A seeming generalization of (10) is

$$x(k+1) = A_0 x(k+1) \oplus A_1 x(k) \oplus \cdots \oplus A_{l+1} x(k-l), \tag{13}$$

which is implicit in $x(k+1)$ and which has extra delays. By repeated substitution of the whole right-hand side of (13) for the term $x(k+1)$ in this right-hand side, one gets

$$x(k+1) = A_0^* A_1 x(k) \oplus \cdots \oplus A_0^* A_{l+1} x(k-l) \tag{14}$$

where

$$A_0^* := I \oplus A_0 \oplus A_0^2 \oplus A_0^3 \oplus \cdots$$

The notation I refers to the identity matrix in the max-plus algebra: it has zeros on the main diagonal and ε's elsewhere. Equation (14) only makes sense if A_0^* is well defined (its elements are finite or ε). This is for instance the case if the precedence graph of A, see subsection 2.3, does not contain circuits, because then $A^k = \varepsilon$ for $k \geq n$. Equation (14) can be rewritten as a first order difference equation by augmenting the state space. This is a standard trick in system theory and has already been used in subsection 2.1.

2.3 Periodic Behaviour

Given a square matrix A, we consider the problem of existence of eigenvalues and eigenvectors in the max-plus algebra, that is, the existence of λ and $v \neq \varepsilon$ such that:

$$Av = \lambda v \ . \tag{15}$$

This equation has to be interpreted in the max-plus algebra sense; the expression λv means that one adds λ to each component of v. We already have seen examples of eigenvalues and eigenvectors in subsection 2.1; v corresponds to an initial state resulting in a solution with 'period' 1 and λ is the interdeparture time.

Before formulating a theorem about eigenvalues, some graph theory must be recapitulated. In the following definition the starting point is a square matrix, the entries of which may again assume the 'value' ε.

Definition 2.1 (Precedence graph). *The precedence graph of an $n \times n$ matrix A is a weighted digraph with n nodes and an arc (j, i) if $a_{ij} \neq \varepsilon$, in which case the weight of this arc receives the numerical value of a_{ij}. The precedence graph is denoted $\mathcal{G}(A)$.*

It is not difficult to see that any weighted digraph $\mathcal{G} := (\mathcal{V}, \mathcal{E})$, with $\mathcal{V}$ being the set of nodes and $\mathcal{E}$ being the set of arcs, is the precedence graph of an appropriately defined square matrix. The weight a_{ij} of the arc from node j to node i is defined as the ij-th entry of a matrix A. If an arc does not exist, the corresponding entry of A becomes ε. The matrix A thus defined has $\mathcal{G}$ as its precedence graph.

As we have seen before, the element (i, j) of $A^k = A \otimes \cdots \otimes A$, considered within the max-plus algebra denotes the maximum weight with respect to all paths of length k which go from node j to node i. If no such path exists, then $(A^k)_{ij} = \varepsilon$. The weight of a path ρ is denoted $|\rho|_{\mathrm{w}}$ and its length is denoted $|\rho|_{\mathrm{l}}$.

Definition 2.2. *The* mean weight of a path *is defined as the sum of the weights of the individual arcs of this path, divided by the length of this path. If the path is denoted by ρ, then the mean weight equals $|\rho|_{\mathrm{w}}/|\rho|_{\mathrm{l}}$. If such a path is a circuit one talks about the mean weight of the circuit, or simply the* cycle mean.

We are interested in the maximum of these cycle means, where the maximum is taken over all circuits in the graph. This number will be called the

maximum cycle mean. If the cycle mean of a circuit equals the maximum cycle mean, then the circuit is called critical. The graph consisting of all critical circuits (if there happen to be more than one) is called the *critical graph* and denoted by $\mathcal{G}^c$. In the following theorem the notion 'strongly connected' (di-)graph is used. A graph is called strongly connected if there exists a path from any node to any other node. The matrix corresponding to a strongly connected graph is called *irreducible.*

Theorem 2.1. *We are given a square matrix A. If $\mathcal{G}(A)$ is strongly connected, then there exists one and only one eigenvalue and at least one eigenvector. The eigenvalue is equal to the maximum cycle mean of the graph:*

$$\lambda = \max_{\zeta} \frac{|\zeta|_w}{|\zeta|_l} \ ,$$

where ζ ranges over the set of circuits of $\mathcal{G}(A)$.

Definition 2.3 (Cyclicity of a graph). *Given a strongly connected graph, its cyclicity equals the greatest common divisor of the lengths of all its circuits. The cyclicity of an arbitrary graph (which may consist of several strongly connected subgraphs) equals the least common multiple of the cyclicities of all its maximal strongly connected subgraphs.*

Definition 2.4 (Cyclicity of a matrix). *A matrix A is said to be cyclic if there exist scalars M, λ and d such that $\forall m \geq M$, $A^{m+d} = \lambda^d A^m$. The least such d is called the cyclicity of A. The quantity λ equals the maximum cycle mean of A.*

The expression $A^{m+d} = \lambda^d A^m$ in the definition above must be interpreted in the max-plus algebra sense of course. Thus λ^d in the max-plus algebra means $d\lambda$ in the conventional algebra and $\lambda^d A^m$ refers to the addition of λ^d to each element of A^m.

Theorem 2.2. *Any irreducible matrix is cyclic. The cyclicity of the irreducible matrix A equals the cyclicity of $\mathcal{G}^c(A)$, being the critical graph corresponding to matrix A.*

Example 2.1. Consider the A-matrix of (6);

$$A = \begin{pmatrix} 2 & 5 & \varepsilon & \varepsilon \\ \varepsilon & \varepsilon & 3 & 3 \\ 0 & \varepsilon & \varepsilon & \varepsilon \\ \varepsilon & 0 & \varepsilon & \varepsilon \end{pmatrix}.$$

The corresponding precedence graph has three circuits, viz. from node 1 to node 1 with cycle mean $2/1 = 2$; from node 1 to 3 to 2 to 1 with cycle mean $(0+3+5)/3 = 8/3$; from node 2 to 4 to 2 with cycle mean $(0+3)/2 = 3/2$. (It is tacitly assumed here that node i corresponds to x_i.) The maximum cycle mean equals 8/3. There is only one critical circuit. The cyclicity of the critical graph (which equals the critical circuit) equals 3. The quantities of definition 2.4 are $M = 5$, $\lambda = 8/3$ and $d = 3$;

$$A^8 = \begin{pmatrix} 20 & 23 & 24 & 24 \\ 19 & 20 & 21 & 21 \\ 18 & 21 & 20 & 20 \\ 15 & 18 & 19 & 19 \end{pmatrix} = (\frac{8}{3})^3 A^5 = 8 \begin{pmatrix} 12 & 15 & 16 & 16 \\ 11 & 12 & 13 & 13 \\ 10 & 13 & 12 & 12 \\ 7 & 10 & 11 & 11 \end{pmatrix}.$$

□

Example 2.2 (Intercity railway network (continued)). The problem description has been given in Section 1. With the theory just given, answers to the various questions posed can be given now. The model which follows directly from the problem description is of the form (13), with the vector x being 53-dimensional. (It has been described explicitly in [3].) After rewriting it in the form of (10), a model with state vector of dimension 79 resulted. The matrix A will not be given here: it is made up of the different travelling times and standard change over times as set by the railway company. The answer to question 4 turns out to be $\lambda = 27\frac{1}{12}$. The critical circuit (this is not to be confused with a line of the net) turns out to be from Venlo to Eindhoven to Utrecht to Amsterdam to Zandvoort aan Zee and back. Hence if one wants a faster time table, one should add extra trains to a line (or lines) which forms part of this critical circuit (i.e. the lines numbered 20 and 50). If one would add $5\frac{2}{3}$ minutes to all change over times, to be incorporated into the model by adjusting the A matrix, then it turns out that eigenvalue $\lambda = 30$ minutes. Of course this eigenvalue is monotone with respect to the duration of the change over times. The critical circuit in this case with maximum change over time which allows a half hour time table is Venlo to Eindhoven to Utrecht to Amsterdam to Haarlem to Den Haag HS to Breda to Eindhoven to Venlo. If one would consider a very unfriendly way of handling passengers by not waiting for them, i.e. trains would simply stop at stations, deliver passengers and pick up passengers which happen to be there and depart again, then we can design a faster time table. It turns out

that in this case $\lambda = 26\frac{2}{11}$. The critical circuit is now Venlo to Eindhoven to Breda to Den Haag CS and back.

The question of the design of an 'optimal line structure' cannot be answered directly by an application of the theory. One can of course compare different designs by calculating their critical circuits and minimum interdeparture times.

□

2.4 The γ-Transform

Conventional linear systems with inputs and outputs are of the form (12), though (12) itself has the max-plus algebra interpretation. This equation, now considered in the conventional way, is a representation of a linear system in the time domain. Its representation in the z-domain equals

$$Y(z) = C(zI - A)^{-1}BU(z) \ ,$$

where $Y(z), U(z)$ are defined by

$$Y(z) = \sum_{i=0}^{\infty} y(i)z^{-i}, \quad U(z) = \sum_{i=0}^{\infty} u(i)z^{-i} \ ,$$

where it is tacitly assumed that the system was at rest for $t \leq 0$ and where I refers to the unit matrix in the conventional algebra. The matrix $H(z) := C(zI - A)^{-1}B$ is called the transfer matrix of the system.

In the max-plus algebra context, the z-transform also exists, but here it is customary to refer to it as the γ-transform where γ operates as z^{-1} and is assumed to be real-valued. For instance, the γ-transform of u is defined as

$$U(\gamma) = \bigoplus_{i=0}^{\infty} u(i) \otimes \gamma^i \ , \tag{16}$$

and $Y(\gamma)$ and $X(\gamma)$ are defined likewise. Multiplication of (12) by γ^k yields

$$\left.\begin{array}{rcl} \gamma^{-1}x(k+1)\gamma^{k+1} & = & A \otimes x(k)\gamma^k \oplus B \otimes u(k)\gamma^k, \\ y(k)\gamma^k & = & C \otimes x(k)\gamma^k \ . \end{array}\right\} \tag{17}$$

If these equations are summed with respect to $k = 0, \ldots,$ and if we add $\gamma^{-1}x_0$ to both sides of the first equation thus obtained, then we obtain

$$\left.\begin{array}{rcl} \gamma^{-1}X(\gamma) & = & A \otimes X(\gamma) \oplus B \otimes U(\gamma) \oplus \gamma^{-1}x_0 \ , \\ Y(\gamma) & = & C \otimes X(\gamma) \ . \end{array}\right\} \tag{18}$$

The first of these equations can be solved by first multiplying (max-plus algebra), equivalently adding (conventional), left- and right-hand side by γ and then repeatedly substituting the right-hand side for $X(\gamma)$ within this right-hand side. This results in

$$X(\gamma) = (\gamma A)^*(\gamma B U(\gamma) \oplus x_0) \ .$$

Thus we obtain $Y(\gamma) = H(\gamma)U(\gamma)$, provided that $x_0 = \varepsilon$, and where the transfer matrix $H(\gamma)$ is defined by

$$H(\gamma) = C \otimes (\gamma A)^* \otimes \gamma \otimes B = \gamma CB \oplus \gamma^2 CAB \oplus \gamma^3 CA^2B \oplus \cdots \qquad (19)$$

The expression $Y(\gamma) = H(\gamma)U(\gamma)$ is the max-plus algebra equivalent of $Y(z) = H(z)U(z)$ in the conventional system theory. If one writes

$$H(z) = C(zI - A)^{-1}B = C(\frac{1}{\gamma}I - A)^{-1}B = \gamma C(I - \gamma A)^{-1}B =$$

$$\gamma C(I + \gamma A + \gamma^2 A^2 + \cdots)B,$$

one has obtained the equivalence of (19) in the conventional sense!

The transfer matrix (19) is defined by means of an infinite series and the convergence depends on the value of γ. If the series is convergent for $\gamma = \gamma'$, then it is also convergent for all γ's which are smaller than γ'. If the series does not converge, it still has a meaning as a formal series.

Exactly as in conventional system theory, the transfer matrix is especially useful when subsystems are combined to build larger systems, by means of parallel, series and feedback connections. For instance, the product of two transfer matrices (of which it is tacitly assumed that the sizes of these matrices are such that the multiplication is possible), is a new transfer matrix which refers to a system which consists of the original systems put into a series connection.

Suppose that $H(\gamma)$ is a scalar function, i.e. the system has one input and one output. The term $\gamma^k CA^{k-1}B$ in (19) can be written in conventional algebra as $c_{k-1} + k\gamma$ (where c_{k-1} represents the coefficient $CA^{k-1}B$) which is a straight line with slope k. The transfer function can be viewed as the maximum (of an infinite number) of such lines and hence is a continuous, piecewise linear and convex function of the variable γ.

3 Some Extensions and Recent Literature

In this subsection we will briefly mention the following related fields, specialisations and/or extensions; Petri nets and event graphs, Axiomatic foundations, Minimal realizations, Stochastic DES, Min-max-plus systems and nonexpansive mappings, Numerical procedures, 'Continuous' DES and the Fenchel transform.

3.1 Petri Nets and Event Graphs

It is assumed that the reader is familiar with the basic properties of Petri nets, see [15] or [21]. It will be shown that the max-plus algebra is extremely suitable in describing the timed behaviour of tokens in so-called event graphs, which form a subclass of Petri nets. In order to set the notation and the stage, we do start with some formal definitions.

Definition 3.1 (Petri net). *A Petri net is a pair $(\mathcal{G}, b)$, where $\mathcal{G} = (\mathcal{E}, \mathcal{V})$ is a bipartite graph with a finite number of nodes (the set $\mathcal{V}$) which are partitioned into the disjoint sets $\mathcal{P}$ and $\mathcal{Q}$; $\mathcal{E}$ consists of pairs of the form (p_i, q_j) and (q_j, p_i) with $p_i \in \mathcal{P}$ and $q_j \in \mathcal{Q}$. The initial marking b is an m-vector, with m being the number of elements in $\mathcal{P}$, of nonnegative integers. The elements of $\mathcal{P}$ are called places, those of $\mathcal{Q}$ are called transitions. The number of elements in these sets are m and n respectively. The elements of the vector b denote the number of tokens in the respective places. One talks about a timed Petri net if time durations are associated with places and transitions.*

Definition 3.2 (Event graph). *A (timed) Petri net is called a (timed) event graph if each place has exactly one upstream and one downstream transition.*

Definition 3.3 (State graph). *A (timed) Petri net is called a (timed) state graph if each transition has exactly one upstream and one downstream place.*

Note that the definitions of event graph and state graph are dual one to the other. In these definitions it was tacitly assumed that the networks are 'closed', i.e. all places (transitions) do have an upstream and a downstream transition (place). The definitions can be extended in the obvious way to

include input transitions (places), so-called *sources*, which do not have upstream places (transitions) and output transitions (places), so-called *sinks*, which do not have downstream places (transitions).

A transition can fire (or: can start firing if there is a positive firing time, see definition 3.4) if all its (directly) upstream places contain at least one token (which must be 'enabled' – see definition 3.5). After the firing these tokens are removed and one token is added to each of the (directly) downstream places.

Definition 3.4 (Firing time). *The firing time of a transition is the time that elapses between the starting and the completion of the firing of the transition.*

Definition 3.5 (Holding time). *The holding time of a place is the time a token must spend in the place before it can contribute to the enabling of the downstream transitions.*

Theorem 3.1. *An event graph with both firing times and holding times is equivalent to an event graph with only holding times (i.e. the firing times are zero). This equivalence means that the time instants at which the transitions fire are the same in both event graphs.*

Proof. See [2]. □

Theorem 3.2. *The number of tokens in any circuit of an event graph is constant.*

From now on we will only consider event graphs with firing times which are zero. Each place connects precisely one transition with precisely one (possibly different) transition. One says in such a situation that the upstream transition, say q_j, is a predecessor of the downstream transition, say q_i. Equivalently one can say that q_i is a successor of q_j. One writes in such a case $j \in \pi^-(i)$ and $i \in \pi^+(j)$.

We make the explicit assumption that if a place connects two transitions such as just has been described, there is no other place with does exactly the same. In general event graphs there can be more 'parallel' places in between two transitions of which one is the successor of the other. The reason for this restriction is purely a notational issue. The theory to be given can handle

the more general situation routinely. We also make the assumption that the underlying network is strongly connected.

If a place exists between the transitions q_j and q_i and q_j is upstream with regard to this place and q_i downstream, then the holding time of this place is indicated by a_{ij}. The holding times are nonnegative real numbers. The number of tokens in this place is indicated by b_{ij}. In the course of time, b_{ij} may change of course, but what is meant here, and also in the formulas to come, (20) and (21), is b_{ij} at the initial time. For some subtle issues with respect to initial conditions, see [2].

If $\tau_i(\chi)$ denotes the earliest time instant at which transition q_i has fired χ times, then

$$\tau_i(\chi) = \max_{j \in \pi^-(i)} a_{ij} + \tau_j(\chi - b_{ij}), i = 1, \ldots, n. \tag{20}$$

Rather than having used the conventional notation x for the state, we now used the symbol τ. This is to distinguish (20), sometimes referred to as the dater equations, from the so-called counter equations defined as

$$x_i(t) = \min_{j \in \pi^-(i)} b_{ij} + x_j(t - a_{ij}), \ i = 1, \ldots, n, \tag{21}$$

where $x_i(t)$ denotes the number of firings of transition q_i, $i = 1, \ldots, n$, which have taken place up to, and including, time t. Equations (20) and (21) describe the same underlying system and one equation is called the dual of the other. Note that χ and x_i are integer-valued. The functions $x_i(t)$ and $\tau_i(\chi)$ are each others inverse in a way.

If in the original event graph there would have been a positive firing time, then the equations above do not exclude the possibility that a transition 'works' simultaneously on two or more tokens. If one wants to exclude this, a loop, including one place with one token, around the transition concerned should be added. The holding time of this new place is defined to be equal to the original firing time of this transition. This loop now takes care of the fact that in the equivalent event graph with only zero firing times, the transition cannot work on two or more tokens simultaneously anymore.

Just as in conventional system theory, the product and parallel composition can also be defined for Petri nets by connecting the proper inputs with the proper outputs; we will not do that explicitly here. Recently evolution equations of timed Petri nets more general than event graphs have been studied; the reader is referred to [1] and [8].

3.2 Axiomatic Foundations.

The operations $\oplus$ and $\otimes$ defined on the set R can also be defined with respect to a more general set of elements $\mathcal{D}$. One then speaks of a *dioid* (sometimes also referred to as a semiring).

Definition 3.6 (Dioid). *A* dioid *is a set $\mathcal{D}$ endowed with two operations denoted $\oplus$ and $\otimes$ (called 'sum' or 'addition', and 'product' or 'multiplication') obeying the following axioms:*

Axiom 1 (Associativity of addition).

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

Axiom 2 (Commutativity of addition).

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

Axiom 3 (Associativity of multiplication).

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

Axiom 4 (Distributivity).

$$\begin{aligned} \forall a, b, c \in \mathcal{D}, \quad & (a \oplus b) \otimes c = (a \otimes c) \oplus (b \otimes c) \ , \\ & c \otimes (a \oplus b) = c \otimes a \oplus c \otimes b \ . \end{aligned}$$

This is right, *respectively* left, *distributivity of multiplication with respect to addition. One statement does not follow from the other since multiplication is not assumed to be commutative.*

Axiom 5 (Existence of a zero element).

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

Axiom 6 (Absorbing zero element).

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

Axiom 7 (Existence of an identity element).

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

Axiom 8 (Idempotency of addition).

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

Definition 3.7 (Commutative dioid). *A dioid is* commutative *if multiplication is commutative.*

With the noticeable exception of Axiom 8, most the axioms of dioids are, required for rings too. Indeed, Axiom 8 is the most distinguishing feature of dioids. Because of this axiom, addition cannot be cancellative, that is, $a \oplus b = a \oplus c$ does not imply $b = c$ in general. Multiplication is not necessarily cancellative either (of course, because of Axiom 6, cancellation would anyway only apply to elements different from ε). For an example in which multiplication is not cancellative take $\mathcal{D} = R \cup \{-\infty\} \cup \{+\infty\}$ and define $\oplus$ as max and $\otimes$ as min.

It is easily shown that in dioids the distributivity with respect to matrices also holds, i.e. $A \otimes (B \otimes C) = (A \otimes B) \otimes C$, where these multiplications only make sense if the matrices have appropriate dimensions. For more information on this axiomatic approach, see Chapter 4 of [2].

3.3 Minimal Realizations.

In subsection 2.4 it was shown how to derive the transfer matrix of a system if the representation of the system in the 'event domain' is given. This event domain representation is characterized by the matrices A, B and C. Now one could pose the opposite question; how to obtain an event domain representation, or equivalently, how to find A, B and C if the transfer matrix is given. In the conventional linear system theory the corresponding theory is known as the realization theory and one speaks of a minimal relization if the sizes of A, B and C are as small as possible, see [11].

The simplest formulation of the (minimal) realization problem in the max-plus algebra is probably as follows. Let G be a sequence of real numbers $\{g_j\}_{j=0}^{\infty}$ and let $A \in R^{n \times n}$, $x_0 \in R^{n \times 1}$, $C \in R^{1 \times n}$ be such that $g_j = C \otimes A^j \otimes x_0$, $j = 0, 1, \ldots$, then G is said to be reproduced by the discrete event system $x(k+1) = A \otimes x(k)$, $x(0) = x_0$ and $y(k) = C \otimes x(k)$. Given a sequence produced in this way, find its realization of the smallest dimension. This realization problem has attracted a lot of attention recently, but for the moment it remains unclear whether an exact algorithmic procedure of polynomial complexity can be found for the general case. This problem was originally formulated in [20]. For recent results the reader is referred to

[6] and [23]. In most approaches the theorem of Cayley-Hamilton, suitably adapted to the max-plus algebra, see [2], plays a crucial role.

3.4 Stochastic Discrete Event Systems.

The evolution equation studied in this part is

$$x(k+1) = A(k) \otimes x(k), \ k = 0, 1, 2, \ldots, \tag{22}$$

with some initial condition $x(0)$. Some (or all) entries of $A(k)$ are stochastic. We assume that

- the underlying distribution functions do not depend on k.
- the stochastic entries can assume only a finite number of different values. It will also be assumed that these values are finite, though the method to be described can be generalized to the case that $-\infty$ is also allowed as a value.
- $A(k)$ and $A(l)$ are independent stochastic matrices for $k \neq l$. (Extensions exist for problems where $A(k)$ and $A(k+1)$ are correlated.)
- no correlation between stochastic entries of $A(k)$ exists, though such correlations can be treated rather routinely.
- $\mathcal{G}(A(k))$ is strongly connected. (If this assumption is true for one k, it automatically is true for all k due to the second assumption above.)

The quantity of central interest is

$$\lim_{k \to \infty} E(x_i(k)/k), \tag{23}$$

for an arbitrary i, being the average cycle time for component i. This quantity is a kind of 'average cycle time'; it can been proved [2] that this average cycle time is independent of i. The method of calculation of the average cycle time will be shown by means of a simple example.

Example 3.1. Consider the case that $x \in R^2$ and that for each k the matrix A is one of the following two matrices

$$\begin{pmatrix} 3 & 7 \\ 2 & 4 \end{pmatrix}, \begin{pmatrix} 3 & 5 \\ 2 & 4 \end{pmatrix}.$$

Table 1: Transitions of stochastic states

initial state	$a_{12} = 7$	$a_{12} = 5$
$n_1 = (0,2)'$	$n_2 + 91$	$n_3 + 71$
$n_2 = (0,-3)'$	$n_4 + 41$	$n_3 + 31$
$n_3 = (0,-1)'$	$n_2 + 61$	$n_3 + 41$
$n_4 = (0,-2)'$	$n_2 + 51$	$n_3 + 31$

Both matrices occur with probability 1/2 and there is no correlation in time. Starting from an arbitrary $x(0)$-vector, say $x(0) = (0,2)'$, we will set up the reachability tree of all possible states x. This is indicated in Table 1, being a table of transitions. In order to get a concise notation, the different state vectors are indicated by n_i, $i = 1, \ldots$. The table has been obtained in the following way. The starting point is $n_1 := (0,2)'$. From there, two states can be reached in one step: $(9,6)'$ or $(7,6)'$, depending on which A-matrix occurs. The states will be normalized such that the first component equals zero. This results in $(0,-3)'$ and $(0,-1)'$. (Other normalizations are possible, and they will lead to the same results.) Both states are new and are therefore added to the list, as n_2 and n_3 respectively. Now we take n_2 as the starting point. Two states can be reached from there: $(4,2)'$ and $(3,2)'$, or, after normalization, $(0,-2)'$ and $(0,-1)'$. Only the first of these states is new and will be added to the list as the next state n_4. In this way we continue: from all states obtained sofar we construct the states which can be reached from there in one step. If a state is found which did not exist sofar, it is added to the list. For the current example it turns out that there exist four different states. The notation $n_i + j\mathbf{1}$ in the table refers to the state n_i of which all components are increased by the number j. One directly notices, by viewing the table, that the system never returns to n_1. Hence this node is a transient one. In the stationary situation a Markov chain results with the three states n_2, n_3 and n_4. Let us be slightly more explicit. The elements of this Markov chain, to be denoted by $z(k)$, are, by construction,

$$z(k) = \begin{pmatrix} 0 \\ x_2(k) - x_1(k) \end{pmatrix}.$$

It is easily shown that

$$z(k+1) = \begin{pmatrix} 0 \\ (Ax(k))_2 - (Ax(k))_1 \end{pmatrix} = \begin{pmatrix} 0 \\ (Az(k))_2 - (Az(k))_1 \end{pmatrix},$$

and hence the process $\{z(k)\}$ is indeed Markovian. The transition matrix of the Markov chain is

$$\begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1/2 \\ 1/2 & 0 & 0 \end{pmatrix}.$$

The stationary distribution of this chain is easily calculated to be

$$\Pr(n_2) = 1/3,\ \Pr(n_3) = 1/2,\ \Pr(n_4) = 1/6.$$

The average cycle time becomes

$$\begin{aligned} &\Pr(n_2)(4\Pr(A_1) + 3\Pr(A_2)) + \Pr(n_3)(6\Pr(A_1) + 4\Pr(A_2)) \\ &\quad + \Pr(n_4)(5\Pr(A_1) + 3\Pr(A_2)) = 13/3, \end{aligned}$$

where the coefficients are the appropriate numbers out of the table above. The first term in this expression for instance, $\Pr(n_2)(4\Pr(A_1 + 3\Pr(A_2))$, is obtained as follows. If the state is in n_2, then this happens with (stationary) probability $\Pr(n_2)$. The next step either leads to n_4, with probability $\Pr(A_1)$ and obtained after 4 time units (see Table 1), or it leads to n_3, with probability $\Pr(A_2)$ and obtained after 3 time units. The other terms are obtained similarly. It is the quantity at the right-hand side,13/3, which equals the expression in (23).

□

This example described a method to calculate the average cycle time. The crucial feature in this method is that the number of different normalized state vectors is finite. See [22] and [2] for extensions.

3.5 Min-Max-Plus Systems and Nonexpansive Mappings.

Referring to the right-hand side of (8), one can define a max-plus expression as a (finite) set of $x_i + a_{ij}$ terms, connected by the max operator. Similarly, one defines a min-max-plus expression as a (finite) set of $x_i + a_{ij}$ terms,

connected by both the max and min operators. An example of such an expression is

$$\max(x_1 + 7, \min(x_2 - 4, x_3 + 1, \max(x_1, x_4 + 2))).$$

With respect to the max operator, we introduced the neutral element $-\infty$. Since we now also deal with the min operator, it is convenient to introduce its neutral element $+\infty$ also. Exactly as one can define a max-plus system by means of max-plus expressions, as (8) can be viewed, one can define a min-max-plus system. Such systems are nonlinear in the max-plus algebra (because of the presence of the min operator) and also nonlinear in the min-plus algebra (because of the presence of the max operator). Notwithstanding this higher complexity, various results about min-max-plus systems are known; the reader is referred to [17], [19] and [9]. Specifically necessary and sufficient conditions are known under which the evolution of (subclasses of) min-max-plus systems show a regular pattern as the evolution of max-plus systems does.

Min-max-plus systems are quite naturally imbedded in the class of so-called nonexpansive mappings, which also are known to have the possibility of periodic behaviour. For more information on such mappings, the reader is referred to [16, 13, 10]. We end this subsection by the definition of nonexpansive mappings.

Definition 3.8. *A mapping f, which maps R^n into R^n is called nonexpansive if*

$$\| f(z) - f(\bar{z}) \| \leq \| z - \bar{z} \|, \tag{24}$$

for arbitrary $z, \bar{z} \in R^n$, and where the norm is an arbitrary $\| \cdot \|_p$ norm, with $1 \leq p \leq \infty$.

3.6 Numerical Procedures.

In this subsection we will confine ourselves to numerical procedures which yield the eigenvalue and eigenvector of a matrix A as expressed by (15).

Theorem 3.3. *Given is an $n \times n$ matrix A with corresponding precedence graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The maximum cycle mean is given by*

$$\lambda = \max_{i=1,\ldots,n} \min_{k=0,\ldots,n-1} \frac{(A^n)_{ij} - (A^k)_{ij}}{n - k}, \quad \forall j. \tag{25}$$

In this equation, A^n and A^k are to be evaluated in the max-plus algebra; the other operations (subtraction and division) are conventional ones.

This theorem is known as Karp's theorem after [12]. This theorem yields the eigenvalue but does not give information about the eigenvectors. For that purpose construct the matrix B by subtracting λ, obtained by Karp's theorem, from all elements of A. The maximum circuit weight of $\mathcal{G}(B)$ equals 0. Hence $B^* = I \oplus B \oplus B^2 \oplus \cdots$ and $B^+ := BB^*$ exist. Matrix B^+ has some columns with diagonal elements equal to zero. To prove this, pick a node k of a circuit χ such that $\chi \in \arg\max_\zeta |\zeta|_w/|\zeta|_l$. The maximum weight of paths from node k to k is 0. Therefore $B^+_{kk} = 0$. Let $B_{\cdot k}$ denote the k-th column of B. Then, since $B^+ = BB^*$ and $B^* = I \oplus B^+$ (I is the identity matrix), for that k,

$$B^+_{\cdot k} = B^*_{\cdot k} \Rightarrow BB^*_{\cdot k} = B^+_{\cdot k} = B^*_{\cdot k} \Rightarrow AB^*_{\cdot k} = \lambda B^*_{\cdot k}.$$

Hence $v = B^+_{\cdot k} = B^*_{\cdot k}$ is an eigenvector of A corresponding to the eigenvalue λ.

A few other numerical approaches exist which calculate the eigenvalue and/or eigenvector:

- Study of the zero(s) of the characteristic equation in the max-plus algebra yields the eigenvalue. For the definition of this equation see [2].
- By means of Linear Programming techniques, see [14].
- Consider (7). Calculate $x(k)$, $k = 0, 1, \ldots$, starting from an arbitrary initial condition, until a state becomes linearly dependent on a state already calulated ($x(7) = 8 \otimes x(4)$). Now $8/(7-4)$ equals the eigenvalue and $(x(4)+x(5)+x(6))/(7-4) = (15\frac{2}{3}\ 13\frac{1}{3}\ 13\ 10\frac{2}{3})'$ is the eigenvector. See [4] for a full account of this method.

3.7 'Continuous' Discrete Event Systems and the Fenchel Transform.

The central equations of this part are (21) and (20). It is assumed now that in addition to t and τ_i, also χ and x_i are real-valued, and so are the quantities b_{ij}. The interpretation of these equations is still a (strongly connected) network with n transitions (also called nodes now). These nodes can now fire continuously. The intensity of this firing is indicated by $v_i(t)$. Quantity $x_i(t)$ denotes again the total amount produced by node i up to (and including) time t. As initial condition it is assumed that $x_i(0) = 0$. The production of a continuously firing transition is sent with unit speed along the outgoing

arcs to the downstream transitions. Thus along an arc there is a continuous flow. The intensity of this flow is $\varphi_i(t,l)$, where l is the parameter indicating the exact location along the arc; $l = 0$ coincides with the beginning of the arc, $l = a_{ji}$ coincides with the end of the arc, where it is assumed that the downstream transition is q_j. As long as the parameters lie in appropriate intervals, we have $\varphi_i(t,l) = \varphi_i(t+s, l+s)$. Moreover, $\varphi_i(t,l) = \varphi_i(t-l,0) = v_i(t-l)$.

At time t the total amount of material along the arc from q_i to q_j equals

$$\int_{l=0}^{l=a_{ji}} \varphi_i(t,s)\, ds. \tag{26}$$

The quantities b_{ij} satisfy $b_{ij} = \int_0^{a_{ij}} \varphi(0,l)\, dl$. The integrand and the integral in (26) must be considered with some care. It is quite well possible that the integrand will contain δ- functions. This will particularly happen at the end of an arc, when material must wait there to be processed by the downstream transition because the other incoming arcs to the same transition have brought in less material sofar. If q_k is a downstream transition to both q_i and q_j and if $x_i(t) < x_j(t)$, then φ_{kj} will start to build a δ-function at $l = a_{kj}$, from t onwards. Of course this δ-function can disappear again later on if $x_i(s) > x_j(s)$ for an s-value with $s > t$. The total amount of material along an arc, as expressed by (26), will in general be time dependent. Many standard results of subsection 2.3 on periodic behaviour remain valid for this continous version of flows on networks, see [18].

Theorem 3.4. *Along a circuit the total amount of material is constant. In formula, if the circuit ζ is characterized by the transitions $\{q_{i_1}, q_{i_2}, \ldots, q_{i_{k+1}} = q_{i_1}\}$, then*

$$\sum_{l=1}^{l=k} \int_0^{a_{i_{l+1},i_l}} \varphi_{i_l}(t,s)\, ds$$

is constant (it does not depend on time).

This is the 'continuous event' analogue of Theorem 3.2. Please note that the total amount of material in the network is not necessarily constant.

Definition 3.9 (Cycle mean). *Given a circuit $\zeta = \{q_{i_1}, q_{i_2}, \ldots, q_{i_{k+1}} = q_{i_1}\}$, its cycle mean is defined as $|\zeta|_w/|\zeta|_l$, where the weight $|\zeta|_w$ and the length $|\zeta|_l$ (assumed to be positive) are defined as*

$$|\zeta|_w = \sum_{l=1,\ldots,k} a_{i_{l+1},i_l}, \quad |\zeta|_l = \sum_{i=1,\ldots,k} b_{i_{l+1},i_l}.$$

Definition 3.10 (Critical circuit). *The circuits which have the maximum cycle mean are called critical. The corresponding cycle mean is indicated by* λ.

Theorem 3.5. *Equations (21) have a solution* $x_i(t) = \frac{1}{\lambda}t + d_i$, *with appropriately chosen constants* d_i.

Just as there are linear difference equations versus z-transforms, and similarly, linear differential equations versus Laplace transforms, we have in the max-plus algebra setting their counterparts. For the discrete event systems we had the max-plus algebra systems versus the γ transform. For the continuous flow variation we have the description above versus a slight variation of the Fenchel transform, called Fenchel* transform. This Fenchel* transform turns out to be the max-plus algebra variant of Laplace transforms, see [2]. The Fenchel transform of a function is also referred to as the conjugate function. We conclude with its definition.

Definition 3.11. *The Fenchel transform* $\mathcal{F}(f)$ *of a function* f *in a Hilbert space* H *is a function in the dual space* H^*;

$$\forall c \in H^* : \ [\mathcal{F}(f)](c) := \sup_{z \in H}(< c, z > - f(z)),$$

where $< \cdot, \cdot >$ *denotes the innerproduct. The Fenchel* transform* $\mathcal{F}^*(f)$ *is defined by*

$$\forall c \in H^* : \ [\mathcal{F}^*(f)](c) := \sup_{z \in H}(< c, z > + f(z)). \tag{27}$$

Usually one confines the definition to convex functions f. Note the resemblance between (16) and (27).

3.8 Synchronization of Traffic Light Processes

In this subsection a Traffic Route System (TRS) for a (part of a) city plan will be mentioned briefly. It concerns a recent application of the max-plus algebra, see [7]. Given a set of interacting traffic routes, the problem is to determine the Sinal Timing Plan (STP) that minimizes the time duration that cars have to wait for green lights while following their predetermined routes (along the route no diiferent choices can be made). Compared to the time table design problem for a railroad network, the railroad stations have become crossings in the city plan and trains have become platoons (i.e. a single sequence of cars that departs from a crossing in the same direction during a green phase).

References

[1] F. Baccelli, G. Cohen, and B. Gaujal. Recursive equations and basic properties of timed petri nets. *Journal of Discrete Event Dynamic Systems*, 2:415–439, 1992.

[2] F. Baccelli, G. Cohen, G.J. Olsder, and J.P. Quadrat. *Synchronization and Linearity.* Wiley, 1992.

[3] J.G. Braker. *Algorithms and applications in timed discrete event systems.* PhD thesis, Delft university of Technology, 1993.

[4] J.G. Braker and G.J. Olsder. The power algorithm in max-algebra. *Linear Algebra and its Applications*, 182:67–89, 1993.

[5] C.G. Cassandras and S. Lafortune and G.J. Olsder. Introduction to the Modelling, Control and Optimization of Discrete Event Systems. *pp. 217-291 in: Alberto Isidori (ed.), Trends in Control, a European Perspective* Springer Verlag, 1995

[6] R.A. Cuninghame-Green and P.Butkovič. Discrete-event dynamic systems: the convex case. Technical Report 93/12, School of Mathematics and Statistics, The University of Birminghame, 1993.

[7] R.-J. van Egmond and G.J. Olsder. The (max,+) algebra applied to synchronization of traffic light processes Submitted for publication, 1998.

[8] B. Gaujal. *Parallélisme et simulation des systèmes à évenements discrets.* PhD thesis, Université de Nice Sophia-Antipolis, 1994.

[9] J. Gunawardena. Min-max functions. *Journal of Discrete Event Dynamic Systems*, 4:377–407, 1994.

[10] J. Gunawardena. *Idempotency.* Cambridge University Press, 1998.

[11] T. Kailath. *Linear Systems.* Prentice-Hall, 1980.

[12] R. M. Karp. A characterization of the minimum cycle mean in a digraph. *Discrete Mathematics*, 23:309–311, 1978.

[13] V. N. Kolokoltsov and V. P. Maslov. *Idempotent analysis and its applications.* Kluwer, 1997.

[14] S. Morioka and T. Yamada. Performance evaluation of marked graphs by linear programming. *International Journal of Systems Science*, 22:1541–1552, 1991.

[15] T. Murata. Petri nets: Properties, analysis and applications. *Proceedings of the IEEE*, 77:541–580, 1989.

[16] R. D. Nussbaum. Lattice isomorphisms and iterates of nonexpansive maps. *Nonlinear Analysis, Theory, Methods and Applications*, 22:945–970, 1994.

[17] G.J. Olsder. Eigenvalues of dynamic min-max systems. *Journal of Discrete Event Dynamic Systems*, 1:177–207, 1991.

[18] G.J. Olsder. Synchronized continuous flow systems. In S. Balemi, P. Kozàk, and R. Smedinga, editors, *Discrete Event Systems; Modelling and Control*, pages 113–124. Birkhäuser, Basel, 1993.

[19] G.J. Olsder. On structural properties of min-max systems. In Guy Cohen and Jean-Pierre Quadrat, editors, *11th International Conference on Analysis and Optimization of Systems*, pages 237–246. Springer Verlag, 1994.

[20] G.J. Olsder and R.E. de Vries. On an analogy of minimal realizations in conventional and discrete event dynamic systems. In *Algèbres exotiques et systèmes à évenements discrets*, pages 191–213. CRNS/CNET/INRIA, CNET, Issy-les-Moulineaux, France, 1987.

[21] W. Reisig. *Petri nets.* Springer Verlag, 1985.

[22] J.A.C. Resing, R.E. de Vries, M.S. Keane, G. Hooghiemstra, and G.J. Olsder. Asymptotic behavior of random discrete event systems. *Stochastic Processes and their Applications*, 36:195–216, 1990.

[23] B. De Schutter and B. De Moor. The characteristic equation and minimal state space realization of SISO systems in the max algebra. In Guy Cohen and Jean-Pierre Quadrat, editors, *11th International Conference on Analysis and Optimization of Systems, Discrete Event Systems*, pages 273–282. Springer Verlag, 1994.

[24] R. Sengupta and S. Lafortune. A deterministic optimal control theory for discrete event systems. In *Proc. 32nd IEEE Conf. on Decision and Control*, pages 1182–1187, San Antonio, TX, December 1993.