Resource Optimization and (min,+) Spectral Theory

Stéphane Gaubert

Abstract—***We show that certain resource optimization problems relative to Timed Event Graphs reduce to linear programs. The auxiliary variables which allow this reduction can be interpreted in terms of eigenvectors in the (min,+) algebra.***

Keywords—***Resource Optimization, Timed Event Graphs, (max,+) algebra, spectral theory.***

I. Introduction

Timed Event Graphs (TEGs) are a subclass of timed Petri nets which can be used to model deterministic discrete event dynamic systems subject to saturation and synchronization phenomena: typically, flexible manufacturing systems, multiprocessor systems, transportation networks [5], [1], [3], [2], [16], [17]. The most remarkable result about TEGs [4], [3], [1] is certainly the following: *a TEG functioning at maximal speed reaches after a finite time a periodic regime*. More precisely, let x denote the *counter function* of a given transition of the graph. That is, $x(t)$ represents the number of firings of the transition up to time t, usually the number of parts of a certain type produced up to time t, the number of messages sent up to time t…Then, there exists a constant λ (the *periodic throughput*) and $c \in \mathbb{N}\setminus\{0\}, T \in \mathbb{N}$ such that

$$t \geq T \Rightarrow x(t+c) = c \times \lambda + x(t) \ . \tag{1}$$

The denomination of *periodic throughput* is justified, because we get from (1):

$$\lambda = \lim_{t\to\infty} \frac{x(t)}{t} = \text{ mean } \frac{\text{number of events}}{\text{time}} \ .$$

We shall consider the case where some markings are *unknown*: we assume that the initial markings (number of tokens) of some places are given by some indeterminates $q_1, \ldots, q_k \in \mathbb{N}$. Typically, the indeterminate q_i associated with place P_i represents an unknown quantity of resources (number of machines, pallets, processors, storage places, buffers) which corresponds to the (unknown) initial marking of this place P_i (see the example in §IV below). Then, the periodic throughput $\lambda = \lambda(q)$ becomes a (nondecreasing) function of the resource indeterminates $q_1, \ldots, q_k$. Given a linear cost

$$J(q) = p_1 q_1 + \cdots + p_k q_k \tag{2}$$

(p_i is the price of one unit of resource i) and a minimal required periodic throughput $\overline{\lambda}$, we consider the following *resource optimization* problem:

$$(\mathcal{RO}) \qquad \min\{J(q) |\ q \in \mathbb{N}^k,\ \lambda(q) \geq \overline{\lambda}\} \ ,$$

which consists in minimizing the cost of the resources needed to obtain (at least) the periodic throughput $\overline{\lambda}$. A slightly different resource optimization problem was first considered by Cohen, Dubois, Quadrat and Viot in [4] where an iterative algorithm was given in order to find a minimal allocation of resources saturating the bottleneck process. The particular problem $(\mathcal{RO})$ has been previously considered by Hillion and Proth [16], Laftit, Proth and Xie [19], [17] and by the author in [12]. In [16], it was noticed that $(\mathcal{RO})$ is an integer linear programming problem, with unfortunately as many constraints as elementary circuits in the graph. In [19], [17], the authors obtained a nice reduction to an auxiliary linear program —with real and integer variables— involving essentially as many constraints as edges in the graph, so that the exact solution can be obtained for much larger systems. However, this result was only given for a restricted class of cost functions and of TEGs. The purpose of this note is to extend the results of [19], [17] to general TEGs and general cost functions: the linear program that we give is exactly the same as in [19], [17], but without undesirable restrictions. As a by product, using the duality between holding times and initial markings in TEGs, we obtain an analogous reduction for an extended resource optimization problem (which involves the possibility of selecting a higher performance equipment instead of buying more machines with a given performance). The simple proof proposed here relies on an elementary key result of the (min,+) spectral theory: we show that the throughput constraint $\lambda(q) \geq \overline{\lambda}$ is equivalent to the existence of a finite "sub-eigenvector" of a particular matrix (sub-eigenvectors are analogous to *potentials* in scheduling theory [2] and to *excessive functions* in potential theory). Then, this potential inequality translates to a set of linear constraints. These results are taken from the thesis of the author, up to some subsidiary extensions. We also mention that the related problem of the symbolic computation of the periodic throughput $\lambda(q)$ has been dealt with in [12], [13].

II. A Sub-Eigenvector Lemma

We first recall some (min,+) spectral theory. The traditional term "(min,+)-algebra" refers to the set $\mathbb{R}\cup\{+\infty\}$ equipped with $\min$ (denoted by $\oplus$) and addition (denoted by $\otimes$). The zero element is written $\varepsilon \stackrel{\text{def}}{=} +\infty$, and we set $e \stackrel{\text{def}}{=} 0$ for the unit. We denote by $\mathbb{R}_{\min}$ this algebraic structure. There is a natural order relation on $\mathbb{R}_{\min}$ given by

$$a \preceq b \iff a \oplus b = \min(a,b) = b \iff a \geq b$$

This is precisely the dual of the usual order (e.g. $2 \succeq 3$). The (min,+) notation extends to matrices in the obvious way. We shall write for instance

$$(AB)_{ij} = (A \otimes B)_{ij} = \bigoplus_k A_{ik} \otimes B_{kj} = \min_k (A_{ik} + B_{kj}) \ ,$$

and consequently $A^k = A \otimes \cdots \otimes A$ (k times). The *spectral radius* $\rho(C)$ [14], [1], [6], [9] of a $n \times n$ matrix C with entries in

S. Gaubert is with INRIA, Domaine de Voluceau, BP 105, 78153 Le Chesnay Cédex, France. E-mail: Stephane.Gaubert@inria.fr .

$\mathbb{R}_{\min}$ is defined by

$$\rho(C) = \min_{1\le k\le n} \min_{i_1,\dots,i_k} \frac{C_{i_1i_2}+C_{i_2i_3}+\cdots+C_{i_ki_1}}{k} \quad (3)$$

$$= \bigoplus_{k=1}^{n} \left(\mathrm{tr}(C^k)\right)^{\frac{1}{k}} \quad , \quad (4)$$

the latest expression being written in the (min,+) algebra (so that $x^{\frac{1}{k}}$ stands for $\frac{1}{k}\times x$ in the usual algebra). The key of our approach is the following *sub-eigenvector* lemma which is reminiscent of Wielandt's proof of the Perron-Frobenius theorem [10], [18], [21]. This lemma is essentially one half of the (min,+) spectral theorem [14], [1], [9]. It seems to have first appeared in [11], [12]. Recall that A is *irreducible* if $\forall ij, \exists k, \quad A^k_{ij} \neq \varepsilon$.

Lemma 1 (sub-eigenvector) Let $A \in (\mathbb{R}_{\min})^{n\times n}$ be irreducible, let $\lambda \in \mathbb{R}_{\min}$. The following assertions are equivalent:

(i) there exists $u \in \mathbb{R}^n_{\min}\backslash\{\varepsilon\}$ such that $Au \preceq \lambda u$,

(ii) there exists $u \in (\mathbb{R}_{\min}\backslash\{\varepsilon\})^n$ such that $Au \preceq \lambda u$,

(iii) $\rho(A) \preceq \lambda$.

Proof of Lemma 1: (i)⇒(ii). We have

$$\forall k, \quad A^k u \preceq \lambda^k u \quad . \quad (5)$$

Take i such that $u_i \neq \varepsilon$. Since A is irreducible, $\forall j, \exists k$ such that $A^k_{ji} \neq \varepsilon$. Thus, we get from (5):

$$\forall j, \exists k, \quad \varepsilon \prec A^k_{ji}u_i \preceq (A^k u)_j \preceq \lambda^k u_j \quad .$$

This implies that

$$\forall j, u_j \neq \varepsilon \quad . \quad (6)$$

(ii)⇒(iii). It follows from (5) that

$$\forall i,k \quad A^k_{ii}u_i \preceq \lambda^k u_i \quad .$$

We get after cancellation of $u_i \neq \varepsilon$:

$$\forall i,k \quad (A^k_{ii})^{\frac{1}{k}} \preceq \lambda \quad . \quad (7)$$

Summing all these inequalities:

$$\rho(A) = \bigoplus_{k=1}^{n} (\mathrm{tr}A^k)^{\frac{1}{k}} = \bigoplus_{i=1}^{n} (A^k_{ii})^{\frac{1}{k}} \preceq \lambda \quad . \quad (8)$$

(iii)⇒(i). Let $\tilde{A} = \lambda^{-1}A$. We introduce the *star* of $\tilde{A}$:

$$\tilde{A}^* \overset{\mathrm{def}}{=} \mathrm{Id} \oplus \tilde{A} \oplus \tilde{A}^2 \oplus \cdots$$

which is well defined because $\rho(\tilde{A}) = \lambda^{-1}\rho(A) \preceq e$ (see [15, Ch. 3, Th. 1] or [1, 3.17]). Since

$$\lambda^{-1}A\tilde{A}^* = \tilde{A}\tilde{A}^* = \tilde{A} \oplus \tilde{A}^2 \oplus \cdots \preceq \tilde{A}^* \quad ,$$

we have

$$A\tilde{A}^* \preceq \lambda\tilde{A}^* \quad ,$$

which means that any column of $\tilde{A}^*$ is a sub-eigenvector of A. ■

III. Application to the Resource Optimization Problem

We show that the throughput constraints $\lambda(q) \geq \overline{\lambda}$ can be formulated as a potential constraint (existence of a sub-eigenvector) in the (min,+) algebra. We shall make certain assumptions which do not restrain the generality but allow a simpler exposition. We consider a TEG such that:

- The holding times are put only on places (so that the firing of the transitions are instantaneous).
- The graph is strongly connected.
- For all transitions i, j, there is at most one place $j \to i$. This allows us to denote by (ij) the unique place $j \mapsto i$, when it exists. We may reduce an arbitrary TEG to this form by adding $k-1$ auxiliary transitions when there are $k \geq 2$ arcs $j \to i$.

We denote by $\{1,\dots,n\}$ the set of transitions. For $1 \leq i,j \leq n$, we denote by T_{ij} the holding time of the place (ij), and by N_{ij} the initial marking (number of tokens). If the place (ij) does not exist, by convention, $T_{ij} = -\infty, N_{ij} = +\infty$. When the initial marking of the place represents a resource l with unknown quantity q_l, we have $N_{ij} = q_l$ (otherwise, N_{ij} is a constant). We shall write $N(q)$ instead of N to emphasize this dependency. We are now in position to state the main result:

Theorem 1: The following assertions are equivalent:

(i) $\lambda(q) \geq \overline{\lambda}$

(ii) $\rho(C(q)) \preceq e$, where the matrix $C(q)$ is defined by $C_{ij}(q) = N_{ij}(q) - \overline{\lambda}T_{ij}$.

(iii) there exists a vector $u \in \mathbb{R}^n$ such that $C(q) \otimes u \preceq u$.

Condition (iii) rewrites as follows in the conventional algebra:

$$\forall i, \quad \min_j (C_{ij}(q) + u_j) \geq u_i \quad ,$$

i.e.

$$\forall i,j, \quad N_{ij}(q) - \overline{\lambda}T_{ij} + u_j \geq u_i \quad . \quad (9)$$

We observe that the constraints which appear in (9) are linear functions of u_i and q_l. Hence, we get from Theorem 1:

Corollary 1: For a strongly connected TEG, the resource optimization problem $(\mathcal{RO})$ is equivalent to the following linear programming problem with integer and real variables:

$$\min\left\{ J(q) \middle| \begin{array}{l} q \in \mathbb{N}^k, \quad u \in \mathbb{R}^n, \\ \forall i,j, \quad N_{ij}(q) - \overline{\lambda}T_{ij} + u_j \geq u_i \end{array} \right\} \quad .$$

Proof of Theorem 1: (i)⇔(ii). Given the matrix B, we define the *weight* $w_B(\alpha)$ of the circuit $\alpha = (i_1,\dots,i_k)$ as follows

$$w_B(\alpha) = B_{i_1i_k} + \cdots + B_{i_3i_2} + B_{i_2i_1} \quad .$$

As it is well known [4], [3], [1], the periodic throughput is given by

$$\lambda(q) = \min_{\alpha \text{ circuit}} \frac{w_N(\alpha)}{w_T(\alpha)} \quad . \quad (10)$$

The following assertions are clearly equivalent:

	$\lambda(q) \geq \overline{\lambda}$	
For all circuit α,	$\frac{w_N(\alpha)}{w_T(\alpha)} \geq \overline{\lambda}$	(by (10))
For all circuit α,	$w_N(\alpha) \geq \overline{\lambda}w_T(\alpha)$	
For all circuit α,	$w_N(\alpha) - \overline{\lambda}w_T(\alpha) \geq 0$	
For all circuit α,	$w_C(\alpha) \geq 0$	
	$\rho(C) \geq 0$.	(from (4)).

We have shown (i)$\Leftrightarrow$(ii). The equivalence of (ii) and (iii) is an immediate consequence of the sub-eigenvector Lemma 1 applied to the matrix C. ■

We remark that the reduction of Corollary 1 also works for some generalized resource optimization problems. It is well known that there is a duality between holding times and initial markings, e.g. the counter/dater duality exhibited in [1, Chap. 5]. In the same spirit, let us assume that the holding times $q'_1, \ldots, q'_l$ of certain places are also unknown (contrarily to the resource quantity q_i, the holding time q'_i need not be an integer, we just require that $q'_i \geq 0$). We thus write $T_{ij} = T_{ij}(q')$. This allows the modeling of certain optimization problems in which we may select the processing times, for instance if we have the choice between several machines with different speeds. Then, λ depends both on the resources and on the processing times, and we write $\lambda = \lambda(q, q')$. We consider a generalized cost function of the form

$$J(q, q') = \sum_{i=1}^{k} p_i q_i - \sum_{i=1}^{l} p'_i q'_i \ ,$$

where $p_i \geq 0$ is defined as in (2) and $p'_i \geq 0$ measures the "price" of the processing time q'_i. This particular structure of cost is needed here for obvious duality reasons. It becomes meaningful for instance if the task with duration q'_i can be performed by a machine whose cost C_i is a decreasing affine function of the processing time, i.e. $C_i = C_i^0 - p'_i q'_i$ for some constant C_i^0. This leads us to consider the extended resource optimization problem:

$$(\mathcal{ERO}) \qquad \min \{ J(q, q') | \ \lambda(q, q') \geq \overline{\lambda} \} \ .$$

We have the immediate extension of Corollary 1:

Corollary 2: The extended resource optimization problem $(\mathcal{ERO})$ is equivalent to the following Linear Program

$$\min \left\{ J(q, q') \middle| \begin{array}{l} q \in \mathbb{N}^k, \quad q' \in (\mathbb{R}^+)^l, \quad u \in \mathbb{R}^n, \\ \forall i, j, \quad N_{ij}(q) - \overline{\lambda} T_{ij}(q') + u_j \geq u_i \end{array} \right\} \ .$$

The effective resolution of linear problems of this type is out of the scope of this paper. Some additional reductions can be found in [19], [17].

IV. Example

In order to illustrate this reduction, we consider a very simple example: a Kanban production line with two cells. See [7], [8] for a more complete presentation of the Kanban policy. The first cell is composed of n_1 machines working independently in parallel on parts of the same type with a processing time of t_1. The description of the second cell is similar. Moreover, we assume that K_i kanbans are allocated to cell $i = 1, 2$ (the number of kanbans limit the total number of jobs in process in a given cell [7]). The number of machines n_i and the number of kanbans K_i are seen as unknown resources. We only consider the autonomous regime (when the system is not delayed by a shortage of raw materials or a lack of demand). Then, we obtain the TEG of Figure 1. For instance, x_1 represents the entrance of a new part in the first cell, x'_1 the beginning of processing of this part by one of the n_1 machines, x''_1, the end of processing, x_2 the entrance of a part coming from cell 1 into cell2,…The matrix N of initial markings is given by:

$$N(n, K) = \begin{array}{c} \\ x_1 \\ x'_1 \\ x''_1 \\ x_2 \\ x'_2 \\ x''_2 \\ x_3 \end{array} \begin{array}{c} \begin{array}{ccccccc} x_1 & x'_1 & x''_1 & x_2 & x'_2 & x''_2 & x_3 \end{array} \\ \begin{pmatrix} & & & K_1 & & & \\ 0 & & n_1 & & & & \\ & 0 & & & & & \\ & & 0 & & & & K_2 \\ & & & 0 & & n_2 & \\ & & & & 0 & & \\ & & & & & 0 & \end{pmatrix} \end{array}$$

where the absence of value at ij stands for $N_{ij} = +\infty$. Similarly,

$$T(t) = \begin{array}{c} \\ x_1 \\ x'_1 \\ x''_1 \\ x_2 \\ x'_2 \\ x''_2 \\ x_3 \end{array} \begin{array}{c} \begin{array}{ccccccc} x_1 & x'_1 & x''_1 & x_2 & x'_2 & x''_2 & x_3 \end{array} \\ \begin{pmatrix} & & & 0 & & & \\ 0 & & 0 & & & & \\ & t_1 & & & & & \\ & & 0 & & & & 0 \\ & & & 0 & & 0 & \\ & & & & t_2 & & \\ & & & & & 0 & \end{pmatrix} \end{array}$$

where dually the absence of value at ij stands for $T_{ij} = -\infty$. In this particular case, the symbolic expression $\lambda(K, n, t)$ can be immediately obtained from Formula (10) by a simple enumeration of the elementary circuits:

$$\lambda(n, K, t) = \min_{i=1,2} \min(\frac{n_i}{t_i}, \frac{K_i}{t_i}) \ . \tag{11}$$

Corollary 2 shows that the constraint $\lambda(n, K, t) \geq \overline{\lambda}$ is equivalent to the existence of a vector

$$u = [x_1, x'_1, x''_1, x_2, x'_2, x''_2, x_3]^T \in \mathbb{R}^7$$

such that

$$\begin{array}{lcllcl} x_1 & \geq & x'_1 & \qquad x_2 & \geq & x'_2 \\ -\overline{\lambda} t_1 + x'_1 & \geq & x''_1 & \qquad -\overline{\lambda} t_2 + x'_2 & \geq & x''_2 \\ x''_1 & \geq & x_2 & \qquad n_2 + x''_2 & \geq & x'_2 \\ n_1 + x''_1 & \geq & x'_1 & \qquad x''_2 & \geq & x_3 \\ K_1 + x_2 & \geq & x_1 & \qquad K_2 + x_3 & \geq & x_2 \ . \end{array} \tag{12}$$

In this particular case, the naive enumeration of circuits (11) is simpler than writing the auxiliary linear program (12). However, for large graphs, such an enumeration becomes practically impossible (for a complete graph with n vertices, there is $O((n-1)!)$ elementary circuits, hence, a priori $O((n-1)!)$ terms in (11)) while the auxiliary program of Corollary 2 which contains at most n^2 inequalities can always be written.

References

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

[2] J. Carlier and P. Chretienne. *Problèmes d'Ordonnancement: modélisation, complexité, algorithmes*. Masson, Paris, 1988.

[3] P. Chretienne. *Les Réseaux de Petri Temporisés*. Thèse Université Pierre et Marie Curie (Paris VI), Paris, 1983.

[4] G. Cohen, D. Dubois, J.P. Quadrat, and M. Viot. Analyse du comportement périodique des systèmes de production par la théorie des dioïdes. Rapport de recherche 191, INRIA, Le Chesnay, France, 1983.

[5] G. Cohen, P. Moller, J.P. Quadrat, and M. Viot. Algebraic tools for the performance evaluation of discrete event systems. *IEEE Proceedings: Special issue on Discrete Event Systems*, 77(1), Jan. 1989.

[6] R.A. Cuninghame-Green. *Minimax Algebra*. Number 166 in Lectures notes in Economics and Mathematical Systems. Springer, 1979.

[7] M. di Mascolo. *Modélisation et évaluation de performance de systèmes de production gérés en Kanban*. Thèse, Institut National Polytechnique de Grenoble, Grenoble, 1990.

[8] M. di Mascolo, Y. Frein, Y. Dallery, and R. David. A unified modeling of kanban systems using petri nets. *The International J. of Flexible Manufacturing Syst.*, 3:275–307, 1991.

[9] P. Dudnikov and S. Samborskiĭ. Endomorphisms of semimodules over semirings with an idempotent operation. *Math. in USSR, Izvestija*, 38(1), 1992. translation of Izv. Akad. Nauk SSSR Ser. Mat. 55, 1991.

[10] F.R. Gantmacher. *Theorie des Matrices, volume II*. Dunod, 1966.

[11] S. Gaubert. Exotic and ordinary spectral radii of nonnegative matrices. Unpublished, September 1991.

[12] S. Gaubert. *Théorie des systèmes linéaires dans les dioïdes*. Thèse, École des Mines de Paris, July 1992.

[13] S. Gaubert. Symbolic computation of periodic throughputs of timed event graphs. In *Proceedings of the Belgian-French-Netherlands Summer School on Discrete Event systems*, Spa, June 1993.

[14] M. Gondran and M. Minoux. Valeurs propres et vecteurs propres en théorie des graphes. In *Problèmes combinatoires et théorie des graphes*, number 260 in Colloques internationaux CNRS, Orsay, 1976.

[15] M. Gondran and M. Minoux. *Graphes et algorithmes*. Eyrolles, Paris, 1979. Engl. transl. *Graphs and Algorithms*, Wiley, 1984.

[16] H.P. Hillion and J.M. Proth. Performance evaluation of job-shop systems using timed event-graphs. *IEEE Trans. on Automatic Control*, 34(1):3–9, Jan 1989.

[17] S. Laftit, J.M. Proth, and X.L. Xie. Optimization of invariant criteria for event graphs. *IEEE Trans. on Automatic Control*, 37(6):547–555, 1992.

[18] H. Minc. *Nonnegative matrices*. Wiley, 1988.

[19] J.M. Proth and X.L. Xie. Les critères invariants dans un graphe d'événements déterministes. *CRAS, Série I*, 313:797–800, 1991.

[20] C.V. Ramamoorthy and G.S. Ho. Performance evaluation of asynchronous concurrent systems using petri nets. *IEEE Trans. on Software Engineering*, SE–6:440–449, 1980.

[21] E. Seneta. *Non-negative matrices and Markov chains*. Springer series in statistics. Springer, 1981.